



Contents lists available at **ETG**

Emerging Technologies and Governance

Journal homepage: www.etg.ihu.ac.ir



Conceptual Framework for the Development of Digital Humanities Based on Artificial Intelligence

Akram Fathian Dastgerdi¹ , Saeed Bakhtiari^{2*} 

1. Assistant Prof., Knowledge and Information Science, Islamic World Science & Technology Monitoring and Citation Institute (ISC), Shiraz, Iran

2. Assistant Prof., Knowledge and Information Science, Information Management Research Unit, Islamic World Science and Technology Monitoring and Citation Institute (ISC), Shiraz, Iran

Corresponding author: fathian@isc.ac

 <https://doi.org/10.47176/ETG.2026.1009>

ARTICLE INFO

Article history:

Received: 2 September 2025

Revised: 27 November 2025

Accepted: 11 December 2025

Keywords:

Digital Humanities;

Artificial Intelligence;

Conceptual Framework;

Deep Learning;

Natural Language Processing.

ABSTRACT

This study aims to develop a conceptual framework for advancing Digital Humanities (DH) through Artificial Intelligence (AI) technologies. A systematic literature review was conducted using the protocol proposed by Kitchenham et al., combined with thematic analysis and guided by the PRISMA framework. Articles were retrieved from the Scopus database (2015-2025). The initial search yielded 544 articles, which were reduced to 20 articles after applying inclusion and exclusion criteria and screening for direct relevance to AI-based digital humanities research. The findings indicate that research in AI-driven digital humanities is predominantly data-driven and computational, with quantitative methodologies accounting for approximately 60% of the analyzed studies. Machine learning models, particularly deep learning and transformer-based architectures such as BERT, T5, CNNs, and LSTMs, were widely applied to tasks including text and image classification, topic modeling, and automated cultural data analysis. These approaches were frequently integrated with Natural Language Processing (NLP) techniques and knowledge graph-based representations, enabling context-aware and semantically enriched analyses of large-scale, complex, and unstructured cultural and historical datasets. The results also suggest an emerging tendency toward multimodal and hybrid approaches. Based on the analysis, seven core components were identified as the foundation of the proposed conceptual framework: deep learning and neural network models; NLP and structural analysis; cultural and archival data analysis; challenges and limitations; opportunities and benefits; computational resources and data requirements; and community engagement and ethical considerations. Overall, the findings suggest that AI can effectively augment human-centered approaches, offering a structured foundation for advancing research in digital humanities.

How to cite this article: Fathian Dastgerdi, A., Bakhtiari, S. (2026). Conceptual Framework for the Development of Digital Humanities Based on Artificial Intelligence. *Emerging Technologies and Governance*, 1(1), 93–111. <https://doi.org/10.47176/ETG.2026.1009>

3115-9311/ © 2026 The Authors. Published by Imam Hossein University.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



1. Introduction

Digital Humanities (DH) is an interdisciplinary field in which digital technologies are employed to formulate new research questions, analyze complex datasets, and offer innovative modes of interpretation within the humanities. Among its most significant contributions is the transformation of computational and data-driven capacities into tools for generating new forms of inquiry, methodology, and analysis that were previously unattainable in non-digital environments ([Schreibman et al., 2016](#)). These approaches enable researchers to study human phenomena on a larger scale and with greater precision, thereby yielding novel insights. Accordingly, the Digital Humanities create a space where interpretive analysis is integrated with computational analysis, opening new horizons for interdisciplinary inquiry.

The growing importance of DH in education and research stems from its ability to perform scalable analyses, uncover hidden patterns in large textual, historical, or visual datasets, and provide new modes of interaction with cultural heritage ([Berry & Fagerjord, 2017](#)).

In recent years, the emergence of advanced artificial intelligence technologies, especially machine learning, deep learning, natural language processing (NLP), and computer vision, has profoundly transformed the capabilities of DH. Advanced language models facilitate the automated reading of ancient texts, retrieval and recognition of manuscripts, extraction of latent themes, stylistometric comparison, historical analysis, and even narrative reconstruction ([Boden, 2018](#)).

In literary and linguistic studies, large language models (LLMs) have demonstrated unprecedented precision in identifying writing patterns, stylistic markers, textual similarities and semantic structures. In the domains of codicology and cultural heritage, computer vision methods support the automated analysis of manuscript images, detection of damage, and estimation of provenance. These capacities illustrate that AI functions not only as an auxiliary tool but also as a transformative force within the DH.

1.1. Artificial Intelligence in Digital Humanities

Recent and comprehensive research indicates that AI in the DH extends far beyond digitization or data archiving; it encompasses data analysis, knowledge extraction from qualitative sources, cultural visualization, and hybrid human–computer methodological integration. Consequently, AI has played an instrumental role in expanding the scope and accuracy of DH scholarship ([Chapinal-Heras & Díaz-Sánchez, 2023](#)).

In this regard, Frontoni et al. (2024) emphasized that AI should be understood not only as a supportive tool but also as a methodological reframing factor within DH. This paradigmatic shift encompasses cultural heritage preservation, text analysis, data classification, multimodal analysis and innovative representation of cultural information.

Similarly, [Russo et al. \(2025\)](#), reporting on the “MAGIC” project, provide a concrete example of AI application in DH: in this project, damaged pages of manuscripts and old printed books (13th–16th centuries) are reconstructed through AI-based OCR/HTR techniques and image restoration.

This case vividly demonstrates the pivotal role that AI can play in safeguarding tangible and intangible cultural heritage.

To understand the diffusion and scholarly reception of AI within DH literature, [Shang et al. \(2025\)](#) conducted a bibliometric analysis of 2,488 abstracts published in leading DH journal. Their findings show that discussions of AI in DH predate the recent rise of generative models but have increased significantly in recent years. The analyses encompass both technical domains (data processing and machine learning) and humanistic or critical dimensions (methodology, theory, and ethics).

[Caruso and Spadaro \(2024\)](#) underscore that although AI creates significant opportunities for DH, theoretical, ethical, and cultural reflections surrounding these transformations remain insufficient; that is, many studies are predominantly application-driven and pay limited attention to the structural, philosophical, and human implications of AI in the humanities.

In the domain of cultural heritage preservation and innovation, [Fu et al. \(2025\)](#) indicate that AI and machine learning can contribute effectively to safeguarding intangible cultural heritage, including customs, languages, techniques, and local cultural practices, while highlighting the need to incorporate ethical, cultural, and legal considerations (such as intellectual property, community participation, and cultural sensitivity) into system design.

Overall, AI provides numerous opportunities for DH, including large-scale text and data analysis, restoration of physical heritage, cultural visualization, interdisciplinary research expansion, and enhanced access to resources. However, AI adoption in DH is accompanied by serious challenges, including tensions between computational rigor and human interpretive depth, ethical, legal, and intellectual property concerns, and insufficient theoretical and structural reflection in many existing projects.

Despite the rapid growth of AI applications in DH, most efforts remain project-based, limited in scope, and lack a coherent conceptual framework. Although several global projects have employed AI for automated historical text analysis, cultural knowledge graph construction, literary character network analysis, and historical scenario simulation, there is still no comprehensive theoretical model elucidating how AI technologies should be integrated with the DH ecosystem or what principles should govern this integration. This gap illustrates the necessity of developing an integrated framework.

Therefore, it is essential to design a coherent conceptual framework for AI-driven Digital Humanities development, one that accounts for varying levels of application (from digitization and preservation to analysis and content generation), the mechanisms of human-machine collaboration (hybrid intelligence), ethical and cultural considerations, and the infrastructural needs related to data, standardization, metadata, intellectual property, and community engagement. Such a framework can guide researchers in employing AI responsibly, critically, and humanely in the DH and in ensuring the quality and scholarly integrity of digital humanities research.

In this context, this study aims to design and propose a conceptual framework for AI-based Digital Humanities development. “To achieve this goal, the present study seeks to answer the following

research question: What are the core components and the overall structure of a conceptual framework for AI-driven Digital Humanities development?"

2. Methods

This study was conducted using a qualitative approach, employing systematic literature review and thematic analysis methods. A systematic review is a specific and reproducible method for searching, identifying, selecting, appraising, and summarizing all studies relevant to a particular question ([Al-Khabori & Rasool, 2022](#); [Shekari, 2023](#)). A systematic review is a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research and collect and analyze data from the studies included in the review ([Moher et al., 2010](#)).

Systematic reviews answer pre-defined research questions using explicit, reproducible methods to identify, critically appraise, and combine the results of primary research studies. Key stages in the production of systematic reviews include clarifying the aims and methods in a protocol, finding relevant research, collecting data, assessing study quality, synthesizing evidence, and interpreting findings ([Pollock & Berge, 2017](#)).

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework was employed to ensure transparency and rigor in reporting. The PRISMA Statement consists of a 27-item checklist and a three-phase flow diagram. The PRISMA Statement aims to help authors improve the reporting of systematic reviews and meta-analyses. PRISMA may also be useful for the critical appraisal of published systematic reviews ([Moher et al., 2010](#)). PRISMA provides authors with guidance and examples of how to completely report why a systematic review was done, what methods were used, and what results were found. Thus, in this study, PRISMA served as a practical tool to structure the review process, enhance methodological clarity, and facilitate the critical appraisal of the literature included.

Following the PRISMA guidelines, a detailed search strategy was developed, specifying the inclusion and exclusion criteria, databases searched, keywords, publication years, document types, and language restrictions. Table 1 presents the full search strategy used in this study.

Table 1.
Search Strategy.

Criteria	Description
Keywords	TITLE-ABS-KEY (("digital humanities" OR "computational humanities") AND ("artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "natural language processing" OR NLP OR "large language model*" OR LLM*) AND (framework OR "conceptual framework" OR "conceptual model" OR "development model" OR "theoretical framework" OR taxonomy OR typology OR "model*" OR "architecture"))
Search field	Title, Abstract, Keywords
Year	2015 - 2025
Inclusion criteria	Document type: Article Source type: Journal

Criteria	Description
Exclusion criteria	Language: English
	Subject area: Computer Science, Social Sciences, Arts and Humanities
	Document type: Non-article documents (e.g., books, book chapters, conference reviews, data papers, critiques, etc.)
	Language: Non-English articles
	Source type: Sources other than journals (e.g., books, conference proceedings, edited volumes, etc.)
	Subject area: Articles outside of Computer Science, Social Sciences, Arts and Humanities
	Year: Articles published before 2015
	Full text availability: Articles without full-text access
	Relevance to research objectives: Articles not related to research objectives based on abstract and full-text review
	Subject relevance: Articles not related to Digital Humanities based on abstract and full-text review

In this study, Scopus was employed based on the quality of journal papers, as it is a database and an indexing system used for bibliometric analysis ([Mongeon & Paul-Hus, 2016](#)). In addition, Scopus contains reliable and high-quality literature with a breadth of coverage that has been rigorously reviewed and selected.

For the present study, articles related to the conceptual framework for the development of digital humanities based on artificial intelligence were examined in the Scopus database from 2015 to 2025. Using the query presented in Table 1, an initial total of 544 articles was identified. After applying various filters, including the English language, document type (articles), source type (journals), subject areas related to Artificial Intelligence, Digital Humanities, Arts, and Architecture, and a publication time frame of 2015 to 2025, the number of articles was reduced to 185.

For the screening process, the titles and abstracts of all 185 articles were reviewed. Articles that were clearly unrelated to the research topic, which focused on the development of conceptual frameworks in the digital humanities based on artificial intelligence, were excluded. Following the initial screening, the remaining articles were further assessed for eligibility based on their full text.

Ultimately, 40 articles were selected for full-text review. However, nine of these articles could not be retrieved. The full texts of the remaining 31 articles were assessed for their eligibility. During this phase, articles were excluded for the following reasons: six articles were excluded due to a lack of relevance to the conceptual framework, four articles were excluded due to a lack of alignment with the research topic, and one article was excluded due to a language mismatch (it was not written in English). After this final assessment, 20 articles were selected for detailed analysis as they best aligned with the research focus and comprehensively covered the topic.

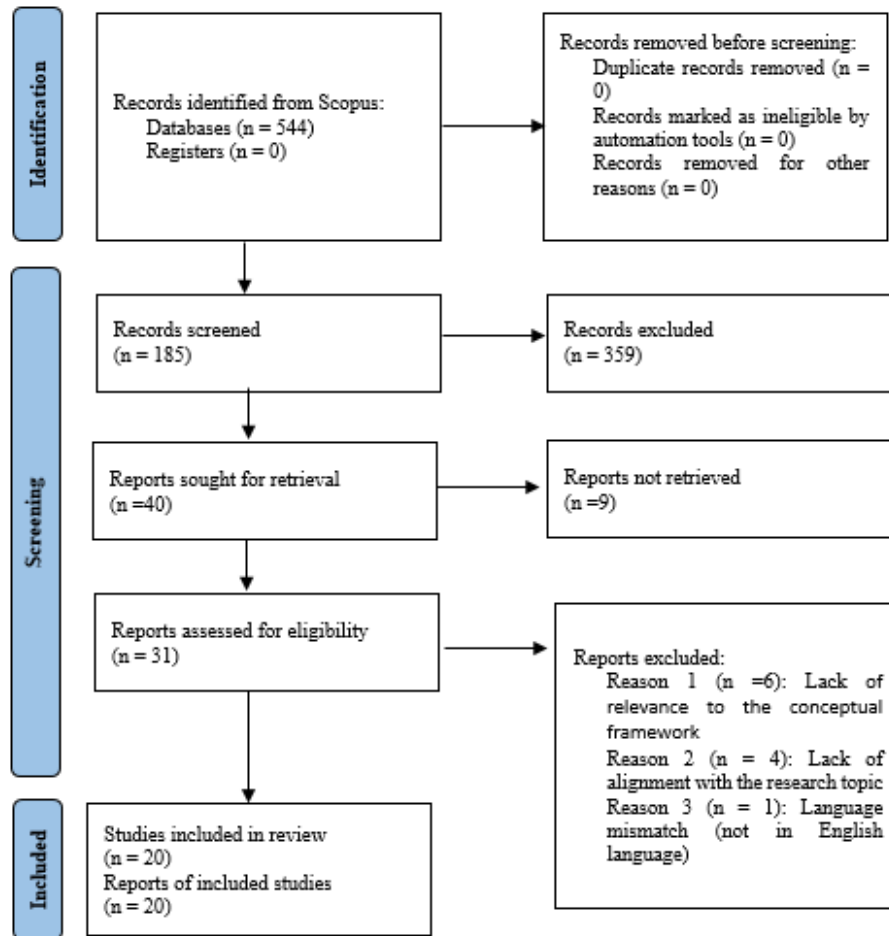


Fig. 1. PRISMA flow diagram.

In this study, because the systematic review was conducted using only the Scopus database, the PRISMA 2020 flow diagram template for new systematic reviews based on database and register searches only was employed.

This study followed the protocol outlined by [Kitchenham et al. \(2007\)](#), encompassing the following key stages: formulation of research questions, development of a search strategy, specification of inclusion and exclusion criteria, primary study selection process, quality assessment, data collection, and data analysis. These steps were conducted concurrently with the PRISMA framework to ensure accuracy and transparency in the review process.

3. Results

The findings of this study were derived from an in-depth analysis of 20 articles selected through a systematic literature review process based on the PRISMA methodology. The research findings showed that most of the analyzed studies were conducted in 2025 (55% of the total). These articles have been published in journals such as the “Journal on Computing and Cultural Heritage”, “Urban

Planning”, “Journal of Cultural Analytics”, “Knowledge-Based Systems”, “EPJ Data Science”, “Frontiers in Computer Science”.

Based on the 20 articles analyzed, it is clear that the research methodologies, tools, and techniques varied across studies, reflecting the diverse approaches in the fields of digital humanities and artificial intelligence. In terms of research methodologies, most studies employed quantitative approaches, with 12 of 20 articles using numerical methods such as machine learning, deep learning, and statistical analysis. These studies primarily focused on tasks such as data classification, regression, and automated analysis. Quantitative methods were used in 60% of the studies analyzed. The second most common methodology was qualitative, used in seven articles (35%). These studies focused on interpretive methods, including thematic analysis, literature reviews, and case studies, often in the realms of cultural heritage, language analysis, and interpretive data analysis. A smaller group of studies (1 out of 20 [5 %]) adopted a mixed or combined approach. These studies integrated both quantitative and qualitative methods to address complex research questions that require numerical analysis alongside human-centered insights.

3.1. Tools and Techniques

In terms of software and tools, the studies predominantly used deep learning frameworks and machine learning models, which were central to many quantitative studies. Tools such as PyTorch, TensorFlow, ResNet, Word2Vec, and Graph Attention Networks (GAT) were frequently mentioned. These tools have been used for a wide range of tasks, including image recognition, text classification, and knowledge graph integration.

Interestingly, Python and its associated libraries, such as Pandas, NumPy, and OpenCV, are widely used for data manipulation, cleaning, and model training. Additionally, Neo4j is often cited as a key tool for graph-based data storage and retrieval, particularly in studies utilizing Knowledge Graphs.

3.2. Models and Frameworks

Many of the studies analyzed employed transformer models such as BERT ([Chun, 2024](#); [Do et al., 2024](#)), ViT (Vision Transformer) ([Castellano et al., 2022](#)), and T5 ([Bulín et al., 2025](#)) for a variety of tasks, including semantic understanding, question generation, and text classification. These models have proven particularly effective in enhancing the ability to interpret complex textual data, allowing for better context-aware processing. In addition to transformer models, other methods such as LDA ([Kozłowski et al., 2023](#); [Schofield et al., 2025](#)) and NMF ([Schofield et al., 2025](#)) have been used for topic modeling and theme extraction, helping to identify underlying themes and patterns within large text corpora.

Another prominent feature across the studies was the use of Knowledge Graphs ([Suissa et al., 2023](#)), which played a crucial role in providing contextual understanding of the data. For instance, models such as ArtGraph ([Castellano et al., 2022](#)), along with other domain-specific knowledge graphs, have been integrated with deep learning models to tackle tasks such as artwork classification and the preservation of intangible cultural heritage. These knowledge graphs allow

for richer and more informed analyses by embedding contextual and historical data into computational models.

Several studies have also introduced hybrid models ([Al-Anazi et al., 2025](#)), which combine neural networks with symbolic reasoning or statistical techniques. This combination aims to improve the overall performance of the models, especially in tasks that require both computational power and human-like reasoning. By merging the strengths of both neural networks and traditional reasoning methods, these hybrid models can capture a broader range of nuances and deliver more accurate results.

3.3. Methodological Trends and Commonalities

A few common trends can be observed in these methodologies.

Deep Learning and AI Integration: Most studies have employed AI-driven methodologies, particularly deep learning and neural networks, to tackle complex data classification and recognition tasks. For example, CNNs ([Al-Anazi et al., 2025](#); [Li et al., 2025](#); [Rath et al., 2024](#)) and LSTMs ([Al-Anazi et al., 2025](#); [Nashir et al., 2025](#); [Suisa et al., 2022](#)) have been used for image classification, whereas transformers such as BERT and T5 have been used for text processing.

1. **Use of Knowledge Graphs:** Several studies have focused on enhancing data analysis through the integration of Knowledge Graphs (KGs), such as ArtGraph and DBpedia ([Castellano et al., 2022](#)), in combination with deep learning methods. This integration facilitates richer context-aware analyses, particularly in the domains of art history, language processing, and cultural heritage.

2. **Multimodal Approaches:** There was a significant emphasis on multimodal learning, where visual and textual data were combined for a more holistic analysis. Studies such as those on fine art classification utilize both visual features (from images) and contextual data (from KGs) to improve performance.

3. **Interdisciplinary Approaches:** Many articles highlighted the integration of humanities and computational methods, emphasizing the importance of interdisciplinary collaboration in overcoming the challenges posed by complex data in the humanities domain.

In conclusion, quantitative methods have been the dominant research approach in the articles analyzed, comprising more than half (55%) of the studies. These studies predominantly utilize deep learning and AI-based methodologies. The use of Knowledge Graphs, multimodal approaches, and deep learning frameworks was prevalent, particularly in studies related to art, language, and cultural heritage. The findings suggest a clear trend toward computational methods, with AI and deep learning tools being the most commonly used in the digital humanities.

3.4. The Conceptual Framework for AI-based Digital Humanities

The articles were thoroughly examined to identify the key components, themes, and challenges related to the development of a conceptual framework for digital humanities using artificial intelligence. This analysis led to the creation of a conceptual model that outlines the key elements,

codes, main themes, subthemes, proposed components, structure of the proposed conceptual model, and references from which these components were extracted.

After analyzing the texts, we identified seven key components, which are further analyzed in the following sections.

3.5. Deep Learning & Neural Networks Models

In this study, after reviewing 20 selected texts in the field of AI-based digital humanities, a deep learning component was observed in 17 sources.

Deep learning and neural networks have become fundamental aspects of AI-based analysis in the digital humanities. These models are particularly effective in analyzing complex textual data, which are often unstructured and vast in scope. AI technologies such as transformer models (e.g., BERT, GPT) (Isac & Das, 2025; Palanivelu et al., 2025; Suissa et al., 2022) and other neural networks enable the analysis of large datasets, helping scholars in the humanities to process and make sense of historical, cultural, and philosophical texts. By employing these models, researchers can achieve more accurate results in tasks such as text classification, sentiment analysis, and even semantic understanding of complex narratives. The integration of deep learning enhances the performance of these models by improving their ability to interpret the context, structure, and meaning of data.

AI-based models are particularly crucial in humanities research because they allow the automation of tasks that would traditionally require human labor, such as annotating texts or identifying hidden patterns in vast archives of cultural data. This deep learning-based conceptual framework ultimately aims to augment human capabilities by providing more efficient and insightful methods for analyzing human and cultural data.

Table 2 presents explanations of the “Deep Learning & Neural Network Models” component as one of the main elements in the conceptual framework for AI-based digital humanities development and the sources providing evidence for this component.

Table 2.

Deep learning and neural network models (component).

Proposed Component	References (N=17)
<u>Deep Learning & Neural Networks Models</u>	
<p style="text-align: center;"><i>Description:</i></p> <ul style="list-style-type: none"> - Using neural networks and deep learning to analyze complex and textual data and to improve accuracy and performance in data analysis - Developing AI models for human data analysis - Use of machine learning and transformer models for analyzing textual and unstructured data - AI-based conceptual model for analyzing human and cultural data 	<p>(Al-Anazi et al., 2025; Castellano et al., 2022; Chun, 2024; Do et al., 2024; Gallagher et al., 2021; Isac & Das, 2025; Kuhn, 2019; Li et al., 2025; Nashir et al., 2025; Olaniyan et al., 2025; Palanivelu et al., 2025; Rath et al., 2024; Schofield et al., 2025; Suissa et al., 2022; Zhang et al., 2025)</p>

3.6. NLP & Structural Analysis

The “NLP & Structural Analysis” component, as one of the key elements in the conceptual framework for AI-based digital humanities development, was identified in 18 of the reviewed studies. Natural Language Processing (NLP) techniques are pivotal for analyzing and extracting meaning from texts, particularly in the study of classic texts in the humanities. By using algorithms such as semantic modeling and text analysis, NLP allows researchers to process and understand large volumes of written material, including historical, philosophical, and cultural documents. These techniques can identify patterns, extract significant themes, and provide interpretations of texts that align with the original semantic intent of the author. In particular, semantic modeling techniques help comprehend the deeper meanings behind words and phrases, which is essential for accurate textual analysis.

NLP’s role in structural analysis extends beyond simply reading text; it involves processing syntactic and semantic structures to provide meaningful insights. This allows for the study of complex, dense, and varied forms of cultural expression, such as classical literature, oral histories, and archival documents. Ultimately, the application of NLP enables a more nuanced understanding of historical and cultural texts, revealing connections that might otherwise remain obscured. Table 2 outlines the “NLP & Structural Analysis” component and lists the sources that provide evidence.

Table 3.
NLP and Structural Analysis (component).

Proposed Component	References (N= 18)
<p><u>NLP and Structural Analysis</u></p> <p><i>Description:</i></p> <ul style="list-style-type: none"> - Using NLP techniques to analyze and extract meaning from text, especially classic texts - NLP-based analysis for complex historical and cultural texts - Processing and analyzing textual data using NLP algorithms and semantic modeling - Use of NLP models for analyzing and semantically modeling historical and philosophical texts 	<p>(Al-Anazi et al., 2025; Castellano et al., 2022; Chun, 2024; Do et al., 2024; Foroughi et al., 2025; Gallagher et al., 2021; Isac & Das, 2025; Kozłowski et al., 2023; Kuhn, 2019; Li et al., 2025; Nashir et al., 2025; Olanayan et al., 2025; Palanivelu et al., 2025; Rath et al., 2024; Schofield et al., 2025; Suisa et al., 2022; Suisa et al., 2023; Zhang et al., 2025)</p>

3.7. Cultural and archival data Analysis

In this study, the component “Cultural and Archival Data Analysis” was identified in 18 of the reviewed texts, underscoring its significance within the conceptual framework for AI-based digital humanities development (Table 4).

Cultural and archival data are among the richest sources of information in digital humanities. This component focuses on leveraging AI technologies to analyze, preserve, and retrieve data from historical and cultural archives. AI-driven tools, such as advanced information retrieval models, allow scholars to delve deeper into these archives, uncovering insights from multimedia data, including text, images, and videos. Archival information retrieval enhances the ability to search, organize, and interpret cultural data, enabling researchers to better understand the evolution of cultural practices, societal norms and historical events. By employing AI technologies to analyze

cultural and archival data, researchers can improve the efficiency of data organization and retrieval, thus supporting the preservation of valuable cultural heritage. These techniques are particularly relevant when dealing with vast amounts of unstructured or fragmented data, such as oral histories, ancient manuscripts, and cultural artifacts.

Table 4.

Cultural and Archival Data Analysis (Component).

Proposed Component	References (N= 18)
<p><u>Cultural and Archival Data Analysis</u></p> <p><i>Description:</i></p> <ul style="list-style-type: none"> - Analyzing and retrieving information from cultural and archival sources, including historical texts, images, videos, and other multimedia materials, using AI-driven tools and advanced information retrieval models - Enhancing the organization, searchability, and interpretation of archival and cultural data - Supporting the preservation and management of large and unstructured datasets, such as oral histories, ancient manuscripts, and cultural artifacts 	<p>(Al-Anazi et al., 2025; Bulín et al., 2025; Castellano et al., 2022; Do et al., 2024; Foroughi et al., 2025; Gallagher et al., 2021; Isac & Das, 2025; Kozłowski et al., 2023; Li et al., 2025; Nashir et al., 2025; Olaniyan et al., 2025; Palanivelu et al., 2025; Rath et al., 2024; Rollo, 2025; Schofield et al., 2025; Suissa et al., 2022; Suissa et al., 2023; Zhang et al., 2025)</p>

3.8. Challenges and Limitations of AI-based Digital Humanities

The component “Challenges and Limitations of AI-based Digital Humanities” was identified in all 20 of the analyzed texts, highlighting its consistent presence and significance across the literature reviewed (Table 5). AI-based digital humanities face a range of technical, structural, and cultural challenges. One of the primary obstacles is related to data: many textual, visual, or audio datasets lack accurate and structured annotations, and the sheer volume of data demands substantial computational and human resources to process. For example, in the analysis of classical Arabic poetry, nearly 75% of poems lack thematic labels, and their poetic themes are complex and multilayered, which single-label models cannot capture. Similar limitations are observed in processing oral data, historical images, or intangible cultural heritage, where data structures are disorganized and highly dependent on temporal and spatial contexts.

In addition to data-related challenges, methodological and ethical issues are prominent. AI models may exhibit algorithmic biases, and their decision-making processes are often opaque to users. Furthermore, the analysis of cultural texts and data requires domain-specific expertise, and automated tools alone cannot replicate the human-level understanding of context and meaning. Practical challenges, such as limited computational resources, difficulties in optimizing deep models, and the risk of overfitting, further complicate research efforts.

Another significant challenge in AI-based digital humanities is domain adaptation, which involves applying models trained in one context (e.g., modern language texts) to another (e.g., classical or historical texts). Differences in historical, linguistic, and cultural contexts create domain shifts, making model adaptation a critical issue. The scarcity of annotated data in specialized fields,

particularly ancient languages and historical records, also limits the applicability of AI models. Efforts to overcome these challenges include improving data quality, enhancing model transparency, and addressing algorithmic bias through better dataset curations and fairness measures. Finally, the scale of human annotation required and the limitations of the available tools impose additional constraints on project scalability, timing, and execution.

Table 5.

Challenges and Limitations of AI-based Digital Humanities (Components).

Proposed Component	References (N= 20)
<p data-bbox="224 562 779 632"><u>Challenges and Limitations of AI-based Digital Humanities</u></p> <p data-bbox="423 678 574 709"><i>Description:</i></p> <ul data-bbox="224 716 779 1136" style="list-style-type: none"> - Addressing data-related challenges, including unstructured, sparse, or complex datasets, and ensuring sufficient computational resources for AI-based analysis - Tackling domain adaptation issues across different historical, linguistic, and cultural contexts - Ensuring algorithmic transparency, mitigating biases, and improving data quality to enhance model performance in human and cultural data analysis 	<p data-bbox="829 688 1414 963">(Al-Anazi et al., 2025; Bulín et al., 2025; Castellano et al., 2022; Chun, 2024; Do et al., 2024; Foroughi et al., 2025; Gallagher et al., 2021; Isac & Das, 2025; Kozłowski et al., 2023; Kuhn, 2019; Li et al., 2025; Nashir et al., 2025; Olaniyan et al., 2025; Palanivelu et al., 2025; Rath et al., 2024; Rollo, 2025; Schofield et al., 2025; Suissa et al., 2022; Suissa et al., 2023; Zhang et al., 2025)</p>

3.9. Opportunities and Benefits of AI-based Digital Humanities

Despite these challenges, the application of AI in the digital humanities presents significant opportunities. This component was identified and observed in all 20 reviewed texts, highlighting its consistent presence across the examined literature (Table 6). The development of large and multi-labeled datasets, such as Arabic poetry archives, genealogical records, and cultural heritage data, enables data-driven research and rapid and precise analysis. Advanced deep learning and transfer learning models can identify complex themes and extract latent knowledge from textual and multimodal data without replacing human analysis. Moreover, AI technologies can enhance researchers' interactions with data by providing interactive analytical platforms and enriched digital environments. Researchers can leverage AI as an "Augmented Social Scientist" to analyze unstructured data more efficiently, accurately, and interpretably, which not only increases productivity but also fosters educational opportunities and interdisciplinary skill development.

However, the effective integration of AI in the digital humanities relies heavily on collaboration among AI experts, humanities scholars, and tool developers. Such interdisciplinary cooperation ensures that AI systems are both technically sophisticated and deeply informed by the cultural, historical, and social contexts that they aim to study. Through this collaborative approach, AI tools can handle multimodal data, including text, images, audio, and other formats, allowing researchers

to capture the richness of human culture and history more fully. Advanced models and tools improve data quality, enable interactive and cross-media analysis, and facilitate access to knowledge and cultural heritage, overcoming the limitations of traditional manual or human-centered methods. Consequently, AI not only supports data analysis but also aids in preserving heritage, promoting educational equity and enabling interdisciplinary research.

Table 6.
Opportunities and Benefits of AI-based Digital Humanities (Component).

Proposed Component	References (N= 20)
<p><u>Opportunities and Benefits of AI-based Digital Humanities</u></p> <p><i>Description:</i></p> <ul style="list-style-type: none"> - Supporting interdisciplinary collaboration among AI experts, humanities scholars, and tool developers - Leveraging AI to analyze and interpret multimodal cultural data (e.g., text, images, audio), enabling new research methods and insights in digital humanities - Utilizing AI-driven models and interactive analytical platforms to enhance research efficiency, preserve cultural heritage, and promote educational and interdisciplinary skill development 	<p>(Al-Anazi et al., 2025; Bulín et al., 2025; Castellano et al., 2022; Chun, 2024; Do et al., 2024; Foroughi et al., 2025; Gallagher et al., 2021; Isac & Das, 2025; Kozłowski et al., 2023; Kuhn, 2019; Li et al., 2025; Nashir et al., 2025; Olaniyan et al., 2025; Palanivelu et al., 2025; Rath et al., 2024; Rollo, 2025; Schofield et al., 2025; Suissa et al., 2022; Suissa et al., 2023; Zhang et al., 2025)</p>

3. 10. Computational Resources & Data Requirements

Another component of the conceptual framework for AI-based digital humanities development, “Computational Resources & Data Requirements,” was identified in 18 of the 20 reviewed articles. The successful implementation of AI in digital humanities research requires robust computational resources and reliable data. AI models, particularly deep learning-based approaches, require substantial computing power; therefore, institutions engaged in digital humanities research must invest in appropriate high-performance computing infrastructure. In addition to hardware resources, access to large, diverse, and high-quality datasets is essential for training accurate and effective AI models. Without reliable and well-curated data, AI systems cannot generate meaningful insights or produce reliable analytical outcomes.

This component emphasizes the importance of establishing sustainable data pipelines, ensuring the availability of curated and interoperable datasets, and supporting standardized data management practices in the humanities domain. Furthermore, the underlying infrastructure must be scalable and adaptable to accommodate the increasing complexity, volume, and multimodal nature of AI-driven digital humanities research. Table 7 presents an overview of the “Computational Resources & Data Requirements” component, along with the sources that provide evidence supporting this element.

Table 7.
Computational Resources and Data Requirements (Component).

Proposed Component	References (N= 18)
--------------------	--------------------

Computational Resources & Data Requirements

Description:

- Ensuring access to robust computational infrastructure to support AI-based digital humanities research
 - Providing large, diverse, and high-quality datasets for accurate training and analysis of AI models
 - Establishing scalable, interoperable, and standardized data pipelines to handle complex, multimodal data in digital humanities projects
-

([Al-Anazi et al., 2025](#); [Bulín et al., 2025](#); [Castellano et al., 2022](#); [Chun, 2024](#); [Do et al., 2024](#); [Gallagher et al., 2021](#); [Isac & Das, 2025](#); [Kozłowski et al., 2023](#); [Kuhn, 2019](#); [Li et al., 2025](#); [Nashir et al., 2025](#); [Olaniyan et al., 2025](#); [Palanivelu et al., 2025](#); [Rath et al., 2024](#); [Schofield et al., 2025](#); [Suissa et al., 2022](#); [Suissa et al., 2023](#); [Zhang et al., 2025](#))

3. 11. Community Engagement & Ethical Considerations

AI systems in the digital humanities involve not only technology but also community engagement and ethical considerations across AI-based research projects. Evidence from the studies reviewed in this research (seven out of 20 studies, e.g., Chun, 2024; Rollo, 2025; Olaniyan et al., 2025) supports this claim (Table 8).

This component emphasizes the need for the active participation of diverse communities, ensuring that ethical AI practices guide the development and implementation of digital humanities initiatives. Community involvement promotes inclusivity, transparency, and cultural sensitivity, particularly when addressing the needs of historically marginalized or underrepresented groups.

Moreover, ethical considerations in AI-based digital humanities research include mitigating algorithmic bias, ensuring transparency in AI-driven decision-making, and fostering collaborative models in which multiple stakeholders, including academics, technologists, and community members, contribute to project design and execution. These ethical practices are essential to ensure that AI systems are not only technically advanced but also socially responsible and aligned with the values and expectations of the communities whose data and cultural heritage are being studied and preserved.

Table 8.

Community Engagement & Ethical Considerations (Component)

Proposed Component	References (N= 7)
<p><u>Community Engagement & Ethical Considerations</u></p> <p><i>Description:</i></p> <ul style="list-style-type: none"> - Promoting community engagement and participation in AI-based digital humanities projects - Ensuring ethical AI practices, including mitigating algorithmic bias and fostering transparency in decision-making - Developing collaborative models where academics, technologists, and community members contribute to 	<p>(Chun, 2024; Do et al., 2024; Foroughi et al., 2025; Olaniyan et al., 2025; Palanivelu et al., 2025; Rollo, 2025; Zhang et al., 2025)</p>

project design, implementation, and preservation of
cultural heritage

3. 12. Validation of the Conceptual Framework

Expert feedback was incorporated into the validation process to validate the proposed conceptual framework for the development of digital humanities based on artificial intelligence. After the conceptual framework was developed, it was reviewed by two experts in the field of digital human sciences. Their insights and feedback were used to refine and ensure the relevance and effectiveness of the framework in addressing key aspects of the intersection between artificial intelligence and the digital humanities.

The experts evaluated the framework based on its alignment with current trends in the digital humanities, practical applicability, and capacity to integrate artificial intelligence technologies effectively. Based on their feedback, several adjustments were made to the framework to ensure that it was comprehensive and aligned with the needs of the field.

This expert validation process confirmed the robustness of the conceptual framework, highlighting its relevance and potential for guiding future research and applications in the digital humanities. In various studies, such as [Seifi and Kazemi \(2018\)](#), [Sadein et al. \(2025\)](#), and [Kahdouei et al. \(2025\)](#), expert feedback has also been employed to validate the credibility of the research in systematic review studies.

4. Discussion

The present study provides a comprehensive overview of the current landscape of AI-based digital humanities, drawing on 20 peer-reviewed articles published predominantly in 2025. The analysis highlights a clear trend toward integrating computational methods, particularly deep learning and transformer-based models, into the study of historical, cultural, and philosophical texts. The widespread adoption of these models underscores their effectiveness in processing large, unstructured, and complex datasets, enabling scholars to perform tasks such as text classification, semantic analysis, and multimodal interpretation with improved accuracy and efficiency.

The findings reveal that natural language processing (NLP) and structural analysis techniques are pivotal for extracting meaning from classical texts. These approaches facilitate the identification of semantic patterns, themes, and syntactic structures, allowing researchers to uncover nuanced insights that traditional human-centered methods may overlook. The integration of NLP with deep learning frameworks and knowledge graphs provides a context-aware approach that enhances interpretability and enriches cultural and historical analyses.

Cultural and archival data analysis has emerged as another critical component, with AI-driven tools enabling the retrieval, organization, and interpretation of diverse materials, including texts, images, and multimedia artifacts. The application of these methods contributes not only to more efficient data management but also to the preservation and safeguarding of cultural heritage, particularly for unstructured or fragmented datasets, such as oral histories and ancient manuscripts.

Despite these advancements, this study identified significant challenges, including data scarcity, domain adaptation, algorithmic bias, and the need for specialized expertise. AI systems often struggle with historical, linguistic, and cultural variations, highlighting the importance of domain-specific adaptations and high-quality curated datasets. Moreover, ethical considerations and community engagement are essential to ensure that AI applications in the digital humanities are socially responsible, inclusive, and culturally sensitive. Collaborative approaches involving humanities scholars, AI experts, and community stakeholders are critical for aligning technological solutions with ethical norms and cultural values.

The opportunities afforded by AI in the digital humanities are substantial. The use of AI enables interdisciplinary collaboration, enhances the analysis of multimodal datasets, and provides interactive platforms for research and education. By functioning as an “augmented social scientist,” AI can support human analysis without replacing it, facilitating more efficient and insightful research processes, while promoting skill development and educational equity. These opportunities demonstrate that when AI technologies are integrated thoughtfully, AI technologies can complement traditional methodologies and expand the analytical capabilities of researchers in the humanities.

Finally, computational resources and data infrastructure are foundational for the successful application of AI in this domain. High-performance computing, scalable data pipelines, and standardized data management practices are crucial for ensuring the accuracy and reliability. The findings suggest that investments in computational infrastructure, coupled with curated and interoperable datasets, are essential for sustaining the growth of AI-based digital humanities research.

5. Conclusions

In conclusion, AI offers transformative potential in the digital humanities by enhancing analytical capabilities, enabling the preservation of cultural heritage, and fostering interdisciplinary collaboration. Seven key components—deep learning and neural network models, NLP and structural analysis, cultural and archival data analysis, challenges and limitations, opportunities and benefits, computational resources and data requirements, and community engagement and ethical considerations—provide a comprehensive foundation for understanding how AI can be effectively applied in historical, cultural, and philosophical research.

The analysis demonstrates that deep learning and transformer-based models, combined with NLP techniques and knowledge graphs, enable accurate, context-aware, and multimodal analyses, thereby augmenting human capabilities in cultural and historical research fields. Moreover, AI applications create opportunities for interdisciplinary collaboration, educational development, and cultural heritage preservation.

However, AI-based digital humanities also face significant challenges, including data scarcity, algorithmic bias, domain adaptation and ethical considerations. Addressing these issues through community engagement, collaborative models, and careful attention to ethical standards is essential to ensure that AI systems are reliable, transparent, and socially responsible.

Future research should focus on several directions to advance the field. These include the development of domain-specific AI models, expansion of high-quality and well-curated datasets, and establishment of robust ethical frameworks and community-based approaches. Additionally, studies should explore other components of digital humanities, such as digital preservation, the role of big data analytics, and emerging computational tools that can enhance AI applications. The practical implementation of the proposed framework in real-world contexts, such as academic research projects, cultural heritage preservation, and interactive digital archives, should also be investigated. Given the growing influence of AI in cultural research, future studies must carefully examine the ethical implications of AI technologies to ensure their responsible and culturally sensitive use.

Funding Statement

This study received no external funding.

Conflicts of interest

The authors declare no conflicts of interest.

Authors contribution statement

Akram Fathian Dastgerdi: Conceptualization; Data curation; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Saeed Bakhtiari: Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Both authors approved the final manuscript. Akram Fathian Dastgerdi is the guarantor for the integrity of the work.

Data Availability Statement

The data supporting the findings of this study are available upon reasonable request from the corresponding author of this paper. Interested researchers may contact the corresponding author (Akram Fathian Dastgerdi) to access the data.

References

- Al-Anazi, R. G., Nashir, W. A., Al-Onazi, B. B., Alhamad, A. A., & Al-Shargabi, A. A. (2025). Maqasid: A Hybrid CNN-BiLSTM Framework for Nuanced Thematic Classification of Arabic Poetry. *IEEE Access*.
- Al-Khabori, M., & Rasool, W. (2022). Introduction to Systematic Reviews and Meta-analyses of Therapeutic Studies. *Oman Medical Journal*, 37(5), e428. <https://doi.org/10.5001/omj.2022.42>

- Berry, D. M., & Fagerjord, A. (2017). *Digital Humanities: Knowledge and Critique in a Digital Age*. Polity Press.
- Boden, M. A. (2018). *Artificial Intelligence: Its Nature and the Future*. Oxford University Press.
- Bulín, M., Švec, J., Ircing, P., Frémund, A., & Polák, F. (2025). Exploring Oral History Archives Using State-of-the-Art Artificial Intelligence Methods. *Acta Informatica Pragensia*, 14(2), 207–214.
- Caruso, M., & Spadaro, A. (2024). Digital Humanities and Artificial Intelligence: An Accelerationist Perspective of the Future.
- Castellano, G., Digeno, V., Sansaro, G., & Vessio, G. (2022). Leveraging knowledge graphs and deep learning for automatic art analysis. *Knowledge-Based Systems*, 248, 108859.
- Chapinal-Heras, D., & Díaz-Sánchez, C. (2023). A review of AI applications in Human Sciences research. *Digital Applications in Archaeology and Cultural Heritage*, 30, e00288.
- Chun, J. (2024). MultiSentimentArcs: A novel method for measuring coherence in multimodal sentiment analysis of long-form narratives in film. *Frontiers in Computer Science*, 6, 1444549.
- Do, S., Ollion, É., & Shen, R. (2024). The augmented social scientist: Using sequential transfer learning to annotate millions of texts with human-level accuracy. *Sociological Methods & Research*, 53(3), 1167–1200.
- Foroughi, M., Wang, T., & Pereira Roders, A. (2025). In praise of diversity in participatory heritage planning empowered by artificial intelligence: windcatchers in Yazd. *Urban Planning*, 10.
- Fu, Y., Shi, K., & Xi, L. (2025). Artificial intelligence and machine learning in the preservation and innovation of intangible cultural heritage: Ethical considerations and design frameworks. *Digital Scholarship in the Humanities*, 40(2), 487–508. <https://doi.org/10.1093/llc/fqaf034>
- Gallagher, R. J., Frank, M. R., Mitchell, L., Schwartz, A. J., Reagan, A. J., Danforth, C. M., & Dodds, P. S. (2021). Generalized word-shift graphs: a method for visualizing and explaining pairwise comparisons between texts. *EPJ Data Science*, 10(1), 4.
- Isac, N. B., & Das, H. (2025). SLIP: A Sanskrit Linguistic Intelligence Pipeline for Enhanced Neural Machine Translation of Classical Texts. *IEEE Access*.
- Kahdouei, E., Farzan, F., & Abedi Samakosh, M. (2025). Sports Tourism in the Desert: A Systematic Review. *Sport and Health Sciences Reviews*, 1(1), 1–23. <https://doi.org/10.22054/jshsr.2025.19636>
- Kitchenham, B., Brereton, P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2007). *Protocol for a Tertiary Study of Systematic Literature Reviews and Evidence-Based Guidelines in IT and Software Engineering*. <https://madeyski.e-informatyka.pl/download/Kitchenham07Protocol.pdf>
- Kozłowski, D., Lozano, G., Felcher, C. M., González, F., & Altszyler, E. (2023). Large-scale computational content analysis of magazines targeting men and women: the case of Argentina 2008–2018. *Feminist Media Studies*, 23(5), 2235–2253.
- Kuhn, J. (2019). Computational text analysis within the Humanities: How to combine working practices from contributing fields? *Language Resources and Evaluation*, 53(4), 565–602.
- Li, S., Nguyen, L. T., Chansanam, W., Iam-On, N., & Boongoen, T. (2025). Enhancing Dongba Pictograph Recognition Using Convolutional Neural Networks and Data Augmentation Techniques. *Information*, 16(5), 362.

- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, P. (2010). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *International Journal of Surgery*, 8(5), 336–341.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Nashir, W. A., Mohsen, A. M., Al-Shargabi, A. A., Nour, M. K., & Al-Onazi, B. B. (2025). Noor: A Pipeline Framework for Classical Arabic Parsing. *IEEE Access*.
- Olaniyan, J., Verkijika, S. F., & Obagbuwa, I. C. (2025). NLP-Based Restoration of Damaged Student Essay Archives for Educational Preservation and Fair Reassessment. *Electronics*, 14(16), 3189.
- Palanivelu, R., Alabdeli, H. M., Srujan Raju, K., Kishore, D., Balamurugan, R., & Abdikadrovich, S. S. (2025). Evolution of Vector-Based Retrieval in Digital Humanities Archives. *Indian Journal of Information Sources and Services*, 15(3), 301–311. <https://doi.org/10.51983/ijiss-2025.IJISS.15.3.34>
- Pollock, A., & Berge, E. (2017). How to conduct a systematic review. *International Journal of Stroke*, 13(2), 138–156. <https://doi.org/10.1177/1747493017743796>
- Rath, S., Priyadarshini, S. B. B., Kumar, P., Sahu, M. P. D. K. P., Jagadev, N., Sahoo, S., & Patra, N. (2024). Enhancing Visual Content Quality in Education and Religion: A Fusion Approach of MIRNet-V2 and Autoencoders for Image Preprocessing and Noise Reduction. *Journal of Theoretical and Applied Information Technology*, 102(15).
- Rollo, M. F. (2025). Interconnected Nature and People: Biosphere Reserves and the Power of Memory and Oral Histories as Biocultural Heritage for a Sustainable Future. *Sustainability*, 17(9), 4030.
- Russo, G., Momtaz, Y., & Maqsood, S. (2025). Artificial Intelligence in Digital Humanities. *Journal of Artificial Intelligence Research and Innovation*, 1(1), 042–045. <https://doi.org/10.29328/journal.jairi.1001006>
- Sadein, S., Abbaspour, J., Hemati Alamdarloo, G., & Alborzi, M. (2025). Identification of Features of E-Book Reading Applications for Children with Dyslexia: A Systematic Review. *Academic Librarianship and Information Research*, 58(3), 1–32. <https://doi.org/10.22059/jlib.2025.387582.1766>
- Schofield, A., Wu, S., Bayard de Volo, T., Kuze, T., Gomez, A., & Sultana, S. (2025). "My Very Subjective Human Interpretation": Domain Expert Perspectives on Navigating the Text Analysis Loop for Topic Models. *Proceedings of the ACM on Human-Computer Interaction*, 9(1), 1–30.
- Schreibman, S., Siemens, R., & Unsworth, J. (2016). *A New Companion to the Digital Humanities*. Wiley-Blackwell.
- Seifi, L., & Kazemi, R. (2018). The use of marketing concepts in public library services: a systematic review. *Library and Information Sciences*, 21(3), 91–124. <https://doi.org/10.30481/lis.2018.68910>
- Shang, W., Ma, R., & Moulaison Sandy, H. (2025). How does digital humanities research discuss AI? A bibliometric analysis. *Information Research*, 30(iConf), 635–645. <https://doi.org/10.47989/ir30iConf47242>
- Shekari, F. (2023). Proposing a Multi-faceted Conceptual Framework for Religious Tourist Experience: A Systematic Review. *Tourism Management Studies*, 18(63), 105–148. <https://doi.org/10.22054/tms.2023.74621.2848>

- Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2022). Text analysis using deep neural networks in the digital humanities and information science. *Journal of the Association for Information Science and Technology*, 73(2), 268–287.
- Suissa, O., Zhitomirsky-Geffet, M., & Elmalech, A. (2023). Around the GLOBE: numerical aggregation question-answering on heterogeneous genealogical knowledge graphs with deep neural networks. *ACM Journal on Computing and Cultural Heritage*, 16(3), 1–24. <https://doi.org/10.1145/3586081>
- Zhang, L., Wang, Z., Yang, R., & Yi, Q. (2025). Digital presentation and interactive learning for intangible cultural heritage preservation using artificial intelligence. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3588520>