Development Compilation

Bangladesh Institute of Professional Studies
https://bips.bd/journal
ISSN: 2072-3334 (Print), ISSN: 2663-1547 (Online)

Volume 19 | Issue 2 | October 20, 2026 | Page: 255-265 | Article |

Artificial Intelligence for the Preservation of Persian Literary Heritage: A Digital Humanities Approach

Md. Sazzad Hossain[1], Md. Mahbubul Alam[2], Md. Mahedi Hasan Dipu[3], Meherun Nesa Mukta[4], Md. Ismail Hossain[5], Nur Ahad Mashrafi[6], MD Mehedi Hasan[7], Jubaer Hossain[8]

1. Department of Persian Language and Literature, University of Rajshahi, Rajshahi, Bangladesh
   Email: s2410812120@ru.ac.bd
2. B.S.S. Department of Political Science, University of Rajshahi, Rajshahi, Bangladesh
   Email: marufsheikh792@gmail.com
3. Department of Political Science, University of Rajshahi, Rajshahi, Bangladesh
   Email: mahedihasan.dipu19@gmail.com
4. Department of psychology, Rajshahi collage, Rajshahi, Bangladesh
   Email: mehrunnesa270@gmail.com
5. Department of Persian Language and Literature, University of Rajshahi, Rajshahi, Bangladesh
   Email: s2411112113@ru.ac.bd
6. Department of Political Science, University of Rajshahi, Rajshahi, Bangladesh
   Email: mdmashrafi858@gmail.com
7. Department of Microbiology, University of Rajshahi, Rajshahi, Bangladesh.
   Email: mahedihasanm2005@gmail.com
8. Department of Philosophy, University of Rajshahi, Rajshahi, Bangladesh
   Email: zubair.ahmed.64777@gmail.com

HOW TO CITE

Hossain, M. S., Alam, M. M., Dipu, M. M. H., Mukta, M. N., Hossain, M. I., Mashrafi, N. A., Hasan, M. M., & Hossain, J. (2026). Artificial intelligence for the preservation of Persian literary heritage: A digital humanities approach. *Development Compilation*, *19*(2), 255–265. https://bips.bd/journal

ARTICLE HISTORY

Received: 30 July 2026
Revised: 22 August 2026
Accepted: 25 September 2026
Available online: 20 October 2026

© 2026 The Author(s). Published by **Development Compilation**, Bangladesh Institute of Professional Studies. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. https://creativecommons.org/licenses/by/4.0/

CC BY 4.0

Abstract

The heritage of Persian literature is threatened by the deterioration of manuscripts, lack of digitization and poor digital archives. The present study is qualitative and explorative, aiming at analyzing the potential of supporting its preservation with the application of artificial intelligence (AI) in a digital humanities perspective. Semi-structured key informant interviews and secondary literature were used to collect data which was thematically analyzed. The results demonstrate that AI has the potential to support digitization and OCR/HTR, as well as metadata creation and information retrieval and computational analysis. But, this accuracy is lowered when the scripts are complex, the manuscripts are damaged or when the text has symbolic or metaphorical meanings. Overall, the study concludes that there is a need for a Human-Centered AI Framework (HCAI-PH) that focuses on expert verification, ethical responsibilities, and collaboration among different fields.

Keywords: Artificial Intelligence, Persian Literary Heritage, Digital Humanities, Handwritten Text Recognition (HTR/OCR), Digital Preservation, Human-Centered Framework

1. Introduction

Persian literary heritage is a treasure trove in the world of literature, which includes poetry, prose, philosophy, historical narratives and Sufi literature. The use of 'artificial intelligence' (AI) as an emerging digital technology is providing new opportunities for preserving, analyzing and enhancing access to this heritage. Some of the main challenges to the Persian literary heritage are manuscript fragility and limited digitization, the historical scriptures, poor digital archives, and academic inaccessibility. AI can help in preservation, but there are significant issues with the accuracy of the technology to recognize historical text and to understand and interpret literary and cultural meaning. In the intersection of digital technologies and literary/cultural studies, AI can be called in when digitizing and recognizing texts, when organizing texts in metadata, and when analyzing texts. In the past work has explored AI applications in cultural heritage (Harmanda et al., 2026), and handwritten text recognition (Nockels et al., 2022) and computational analysis of Persian poetry (Rasti Meymandi et al., 2024). Harmanda et al. (2026) emphasizes the role of AI in cultural heritage preservation, but also mentions some of the technical and ethical challenges. As for the automated handwriting recognition in humanities research, Nockels et al. (2022) and responsible use of AI and human oversight emphasised by UNESCO 2022, there have been studies on the relevance of computational methods to Persian literary analysis by Rasti Meymandi et al. (2024). This research examines the potential of AI for preserving Persian literature and the various technical, ethical and interpretive challenges it faces. Being there it wants to propose a human centered model of digitization and retrieval of written text, scientific verification and viable digital archiving. The main objective is to examine the potential of AI in the preservation of Persian literature in the context of Digital Humanities. The goal is to uncover preservation challenges, discuss the potential and constraints of AI in preservation, and offer a roadmap for a more robust and authentic preservation of texts, accessible and sustainable in the long term.

2. Methodology

This study has adopted the qualitative and exploratory research design by using the perspective of Digital Humanities in the study of the role of artificial intelligence (AI) in preserving the Persian literary heritage. Interviewees were selected using the purposive sampling method and should have relevant knowledge, experience or expertise in one of the following categories: Persian language and literature, preservation of manuscripts, Digital humanities, Artificial Intelligence. The primary data was gathered using semi structured interview on challenges in preservation, digitization with the help of AI, optical character recognition (OCR), handwritten text recognition (HTR), computational literary analysis and ethical aspects. The Secondary data has been collected from academic books, peer-reviewed journal articles, conference proceedings and various reports and files of recognized institutions from digital archives. Collected data were subjected to thematic analysis to look for the patterns of information, emerging problems, possible uses, and gaps in the literature. The secondary literature was compared with the results from the interviews to enrich the understanding of the results as areas of agreement and disagreement were discovered. Based on these findings, the researchers propose a Human-Centered AI Framework for Digital Preservation of Persian Literary Heritage, emphasizing Textual Authenticity, Expert Validation, Interdisciplinary Cooperation, Ethical Stewardship, and Sustainable Future of Digital Archives. Ethical considerations (informed consent, confidentiality (where applicable), correct referencing/acknowledgement and careful data handling) were taken into consideration in the research. The proposed framework is an exploratory study and as such is more of a conceptual contribution and demands additional empirical evaluation to make sure that its effectiveness is proven.

3. Literature review

Meymandi et al. (2024) have analyzed the patterns of poems in terms of thematic, stylistic, and linguistic similarity by using artificial intelligence and evaluated the role of computers in Persian Digital Humanities. Rasool and Pashaei (2026) introduced a model approach called "Smart literature", which showed that the use of AI-based and data-driven tools in combination with traditional methods can enhance the Persian literary analysis with an average accuracy of 72% and encourage interdisciplinary studies, but facing the challenge of defining ethical guidelines and digital infrastructure. Literature translations into Persian with the assistance of

Artificial Intelligence mostly resort to literal, semantically equivalent translations, which leave out cultural, stylistic and emotional aspects of the translation, rendering the art of translation to a lower level (Ostadi & Heidari Tabrizi, 2025). Zargar et al. (2026) compared the performance of different models and discovered that GPT-4o achieved the most accurate performance in comparison with human evaluation, and that the score of sentiment of Rumi poems with different poetic meters were more diverse than Parvin E'tesami's poems. Thus, Berger and Shahghasemi (2026) found that discussions on AI futures in Persian on X were mostly neutral or positive, with more intense negative reactions to existential risks and misinformation, and that there was a relationship between the themes of competition among models and broader social and geopolitical themes. Evaluation of retrieval from historical texts in Persian shows the effectiveness of using Qwen2.5 7B (0.90 BLEU score) compared to a non-adapted version (0.48 BLEU score), and also the sentence similarity improved (Haghdadi et al., 2026). Berger and Shahghasemi (2026) reported the presence of three themes – every day; epistemic; institutional – in Persian language discussions about AI, as well as the prevalence of risk aspects that occurred simultaneously across the various themes and also alongside positive aspects within each theme. Rasti-Meymandi et al., (2024) present an application of transformer-based language models for the unsupervised clustering and computational analysis of Forough Farrokhzad's poems as a way of studying patterns in Persian poetry such as thematic, stylistic and linguistic patterns. However, Mehrpouyan (2026) stated that customized prompts and cybernetical guidelines proved to be effective for considering the interpretation of iham and other rhetorical features; the accuracy of this decision-making process and synchronization with the perspective of human evaluators require culturally-informed rhetorical frameworks and human oversight. Rezaei and Eslami (2024) found ChatGPT and Bing struggled to comprehend the meaning and interpretation of Persian poetry in four literary styles, possibly because their knowledge of literary works lacks the hermeneutic context. Varzandeh (2026) emphasizes during his lecture the necessity of creating a solid Persian LLM as a national mission to safeguard cultural heritage and elevate Iran's AI capabilities and digital sovereignty and its economy. Hosseini (n.d) presents the applications of Artificial Intelligence in the field of Persian language and literature teaching and notes that AI can facilitate and simplify classical texts, play an adaptive role in teaching and learning and emphasizes ethical problems and limitations in teaching higher order language skills. Hussain et al. (2024) studied the Persian social media reviews, used ParsBERT with Persian datasets and obtained 98% accuracy in the sentiment classification process. In addition, they used LIME and SHAP techniques for prediction results explanation and for consumers' preference identification. Amini et al. (2026) have created a Persian model for tourism question answering by fine-tuning ToKA-BERT on around 20,000 generated question–answer pairs with better performance on tourism queries in comparison to general-purpose Persian QA models. Persian-language discussions of AI-enabled misconduct on X present a mix of viewpoints on its implications across various areas like academic dishonesty, fraud, deepfakes, and employment, with many posts including concrete warnings and verification tips. Based on the ethical concerns, the researchers suggest that there are four steps to take when applying AI to the preservation of lullabies in the Persian language: cultural analysis, the design of the technology, ethical assessment, and social feedback. Sadr et al. (2020) tried a couple of machine learning algorithms to detect fake news in Persian Twitter data and their results showed that the randomized decision tree and decision tree algorithms achieved the best result with 90.25% and 90.20% accuracy, respectively. The analysis of Persian language posts from Twitter and Instagram to identify posts that suggest the probability of domestic violence against women was done by Salehi et al. (2023), which the Naive Bayes model had the highest accuracy of 86.77%. In another study conducted by Mirdar Rezaei and Eslami (2024), ChatGPT and Bing were found to lack the hermeneutic scope to work with poetry in four literary styles, noting that they "miss the literary context, and their limited hermeneutic horizons hinder meaningful engagement with literary texts.

4. Literature Gap

Based on the reviewed literature, there is an increase in the number of works which use AI in cultural heritage and AI in Persian language and literature. The study of Harmanda et al. (2026) involved a systematic review of using AI in the cultural heritage sector, which does not focus on Persian manuscripts but is applicable to heritage in general. Without testing Transkribus on historical Persian handwritten text, Nockels et al., (2022) demonstrated that handwritten text recognition can be helpful to them. If there are studies dealing with Persian

literature, they only focus on the computational analysis such as studies conducted by Rasti Meymandi et al. (2024), Zargar et al. (2026) and Rasool and Pashaei (2026). Analysis of the thematic clustering, sentiment and stylistic classification instead of preservation. Other researches (Rezaei & Eslami, 2024; Mehrpouyan, 2026 and Ostadi & Heidari Tabrizi, 2025) have demonstrated that the Persian texts have the problem of the hermeneutic depth, rhetorical features and cultural nuances in their comprehension by AI. Oral and folk heritage is discussed in ethic code of preserving the Persian lullabies in the article by Arian et al. (2026); however, the article is about oral and folk heritage which does not relate to the manuscript-based or literary heritage.

First, there is no study which presents digitization, OCR/HTR, metadata managing, computation, expert verifying and sustainability in a single framework for the preservation of Persian literary heritage. The existing studies can be classified into two groups: The study of preservation technologies, and the study of literature.

Second, there is not so much evidence of the performance of AI on PMs. Gandomkar et al. (2025) provide an OCR dataset of a manuscript of Shahnameh. Only a few examples of a variety of historical handwriting and calligraphic styles, damaged pages, or orthographic variation are examined.

Third, there is little engagement in existing studies with practitioners, such as scholars of Persian literature, manuscript specialists, and digital humanities specialists. Most are based on technical experiments or on secondary sources and lack a feel for the types of problems that are faced by those working in the field and a feel for the institutional problems and a need for human supervision.

Fourthly, there is a general discussion of ethics, law and culture (among them UNESCO 2022). They are presented without a great deal of research to transform them into direct suggestions on copyright, referencing, usage of AI-generated text, and cultural knowledge of Sufi symbols and poetic metaphors to Persian literary archives.

Fifth, many of the Persian AI literature are on other areas such as discourse in social media, Sentiment classification, Tourism, Health and fake news. Does not consider literary preservation. This study aims to fill these gaps. It is based on key informant interviews, secondary literature and proposes a Human-Centered AI Framework (HCAI-PH) for digital preservation of the Persian literary heritage. The framework connects AI-powered digitization experiences with subject-matter expert verification, ethical and legal tips, and sustainable archiving.

5. Proposed Framework

The research introduces an innovative **Human-Centered AI Framework for the Digital Preservation of Persian Literary Heritage (HCAI-PH)** that can combine artificial intelligence (AI), digital humanities, and expert knowledge for the betterment of the preservation, accessibility, and analysis of Persian literary materials. The framework consists of seven interdependent stages to process heritage resources: heritage resource collection, digitization and image processing, AI-based text recognition with optical character recognition (OCR) and handwritten text recognition (HTR), metadata management and digital archiving, computational literary analysis, human verification and scholarly validation, and long-term digital access and sustainability. It focuses upon the preservation of original manuscript images, preservation of accurate transcriptions, documentation of text variants and recording of historical and bibliographic information. In addition, ethical responsibility, transparency, copyright protection, cultural sensitivity, quality assurance, and interdisciplinary cooperation of Persian literature specialists, computer scientists, researchers in the field of digital humanities, librarians and archival institutions are all part of the framework. The role of human verification is still vital in guaranteeing the textual authenticity and avoiding the wrong understanding of historical language, metaphors and symbolism in the sufi texts. Transcription accuracy, completeness of metadata, effectiveness of information retrieval, expert evaluation and user satisfaction can be used to assess the effectiveness of the proposed framework. In sum, the framework sets AI as a supporting tool instead of a substitute to the scholarly judgment and provides a conceptual way of seeing how AI can be used to preserve and explore Persian literature sustainably. It needs to be further empirically tested on representative Persian literary texts in order to prove its practicality.

6. Result and discussion

6.1. AI and digital preservation of cultural heritageAI has been studied and investigated in its ability to help preserve the culture and document its existence in a way that it can be managed and be accessible to the public. Harmanda et al. (2026) conducted a systematic review on 64 papers in the period between 2020 and 2024, which show five areas of application: restoration and preservation, documentation and digitisation, management and decision support, generation and creative reconstruction, and education and experience. The barriers to implementation that are also identified include data quality, technical limitations, ethical concerns, trust and interdisciplinary collaboration. While this study is on cultural heritage in general, it offers a valuable perspective on Persia's literary heritage, albeit not specifically addressing Persian manuscripts (Harmanda et al., 2026). In the field of literary heritage conservation, the use of AI is suggested not only for manuscript digitization but also for improving the cataloguing and retrieval of text, as well as for providing deeper insights and interpretations into literary heritage materials. But if any general conclusions are to be made about cultural heritage, these should not be drawn as evidence of the accuracy of use of a specific AI system for Persian manuscripts.

6.2. Optical character recognition and handwritten text recognition

Optical character recognition (OCR) makes text in pictures into machine-readable text, and handwritten text recognition (HTR) is developed to recognize handwritten text. It is pertinent to making historical documents searchable and appropriate for computational analysis. In a systematic review, Nockels, et al. (2022) analysed 381 articles that were published between 2015 and 2020 and contain the word Transkribus. What they discovered was that the ability of handwritten text recognition was being utilized in archival science, library science, history, computer science and even in other humanities disciplines. Their review highlights the potential for automated transcription to help gain access to historical resources and research on a large scale. But the authors also mention issues to deal with such as complex scripts, document layouts, and human involvement. The results of these findings are applicable to the preservation of Persian manuscripts, as historical manuscripts might contain handwriting variations, complicated calligraphy, damaged pages and different orthographic conventions. The Transkribus review, however, does not directly assess the recognition of Persian manuscripts. Thus, its results are meant to complement the overall HTR methodology, in conjunction with specific technical studies done for the Persian language (Nockels et al., 2022).

6.3. AI-assisted computational analysis of Persian literature

This paper examines the capabilities of AI and natural language processing (NLP) in the context of Persian digital humanities and how they can be applied to analyze the poetry of Forough Farrokhzad. They have studied transformer-based language models and unsupervised clustering for exploring the thematic, stylistic and linguistic patterns in Persian poetry. This study is unique compared to other studies as it directly links AI to the field of Persian literary analysis rather than cultural heritage in general, which is more relevant to the current study. Such forms of computation can be used to look for patterns, contrast the styles of various poets, and discover affinities in large data sets of poems. These strategies can be used as part of close reading, or to offer alternative opportunities for analysing patterns across texts. However, it is possible to say that computational clustering, even if it cannot determine the historical significance of a poem, or its literary significance, can provide a pattern in the data available. When dealing with this poetry, meaning analysis of Persian metaphors, cultural references, symbolism and philosophical aspects of it still plays a significant role in human interpretation (Rasti Meymandi et al., 2024).

6.4. Ethical responsibility and human oversight

UNESCO Recommendation on the Ethics of Artificial Intelligence emphasizes the need for human control and responsibility, transparency and the ethical use of AI. Furthermore, it offers inspiration to use AI appropriately to protect and preserve access to cultural heritage – in the broadest sense of endangered languages and endangered cultural knowledge (UNESCO, 2022) – in the future.

These principles are of utmost significance when it comes to literary preservation in Persia. Mechanical transcription could have incorrect words, AI translation could lack context or symbolic interpretation of a poem. Where an incorrect output is put into the digital record in the absence of documentation and/or review it may impact on future research.

6.5. Research gaps and the need for a Persian-focused framework

The sources reviewed indicate two lines of research that are complementary. First, transdisciplinary cultural heritage research has general issues of data quality, readiness of institutions, ethics, trust and cooperation (Harmanda et al., 2026). Second, the research work in the field of Persian digital humanities shows the potential of computational methods in analyzing Persian poetries (Rasti Meymandi et al., 2024).

In combination these studies offer the foundation for exploring the potential for bringing together technologies of preservation and computational analysis of literary works. Yet, no one of the above-mentioned sources has provided evidences that an integrated framework has been validated for the Persian literary heritage.

7. Primary Data (KII)

From the interview results, it can be seen that artificial intelligence (AI) has a great potential in preserving and maintaining the legacy of Persian literature with the application of digital humanities. The answers cite that deterioration of the manuscripts, lack of financial resources, lack of digital archives and lack of qualified specialists are significant problems for preservation. One informant emphasizes the need not just for scanning manuscripts, but for cataloguing, for the creation of metadata, and for long-term digital access as well: “Preservation doesn't just involve scanning manuscript materials.” The results indicate that AI can be utilized to assist with Optical Character Recognition (OCR), automated transcription, creation of metadata, comparison of texts and computational literary analysis. But the historical Persian script, misspelled words, damaged manuscripts, and the intricate beauty of poetry can cause inaccuracies in AI-generated text. So the words human verification is essential” highlight the importance of expert opinion and technical evaluation. These also show that for successful preservation, textual content, manuscript images, annotations, textual variants and historical context should be retained. As the responses to the interviews demonstrate, ‘text + image + metadata + historical context' gives a rounded and all-inclusive way to retain literary authenticity. Moreover, AI-enhanced research tools can enhance research accessibility by providing full-text search capabilities, thematic analysis, stylistic comparisons, and multilingual support. However, AI might not accurately interpret metaphors, cultural references, and Sufi symbolism, especially in the poetry of Rumi, Hafez, and others. The results also reveal ethical and legal issues such as copyright and ownership, source attribution, cultural sensitivity, and transparency of the results produced by AI. Lastly, the data suggest a multidisciplinary, humanistic approach to preservation, encapsulated in the phrase: “AI should be used to support, not supplant, scholarly judgment” and stress collaboration, long-term funding, independent evaluation, and sustainable digital infrastructure.

The results indicate that there are a number of problems in the preservation of Persian literary heritage namely; the deterioration of manuscripts, lack of financial support, lack of technological facilities and lack of competent staff. In the case at hand, the respondent's definition of successful preservation included such considerations as maintaining the original text, historical authenticity and cultural significance of literary works. Artificial intelligence (AI) can play an important role in manuscript digitization, recognition, metadata creation, and providing access to Persian literary resources. But damaged manuscripts, complex calligraphy, unclear handwriting, and spelling variations from the past can decrease the accuracy of AI-generated transcriptions. The respondent stated that “expert review is needed to ensure accuracy”, which shows that there is a need for human oversight when using AI for preservation. Digital collections need to be thoroughly documented as to corrections, and original manuscript reference, and a large amount of metadata in order to ensure its authenticity and reliability. The use of computational techniques like word-frequency analysis, thematic analysis, authorship attribution and textual comparison can facilitate an understanding of the literary styles and historical linguistic progressions characteristic of Persian literature. However, the interpretation of metaphors

and Sufi philosophies, symbolism, and cultural context in the works of Rumi, Hafez, and Saadi may pose challenges for AI. The respondent shows that mechanical translation of works of Persian literature cannot be based on literal translation, because "Relying solely on literal interpretations may distort the deeper meanings". This framework for the preservation of Persian literature has to be an interdisciplinary effort, that involves the participation of both Persian literature scholars, computer scientists, digital humanists, libraries and archival institutions. It can be assessed by transcription accuracy, experts' feedback, error-rate, effectiveness of information retrieval, and user satisfaction and ethical and legal aspects. In sum, the results indicated that while there are technological benefits to the use of AI preservation solutions, "human supervision must be ensured alongside the technological advantages of AI-based preservation systems" and that "Collaboration between Persian language and literature experts, computer scientists, digital humanities researchers, libraries, and institutions" is necessary in order to ensure sustainable preservation, accessibility, and scholarly use of Persian literary heritage.

8. Overall Findings of the Study

The research reveals that AI can be used to help preserve Persian literary heritage in various aspects, including digitization, text recognition, cataloguing, information retrieval, and computational analysis. But technological power is not enough to ensure text authenticity, historical accuracy or preservation of the literary meaning. AI-powered preservation hinges on quality digital resources, suitable technologies, expert validation, ethical considerations, and sustainable institutional partnerships. The main conclusion is thus that the use of AI should be complementary to the expertise of the specialists of the Persian language and literature. Human centered and interdisciplinary and digital sustainable approach is a possible way to better preserve, access and research the Persian literary heritage.

Table 1 Secondary and primary findings

Study	Evidence that can extract	Relevance
Harmanda et al. (2026)	Review of 64 studies; application domains and implementation challenges	General cultural heritage framework
Nockels et al. (2022)	Review of 381 publications on handwritten text recognition	Historical document transcription
Rasti Meymandi et al. (2024)	AI and transformer-based clustering of Persian poetry	Direct Persian digital humanities evidence
Gandomkar et al. (2025)	CHAMROSH dataset for OCR and character detection in a handwritten Shahnameh manuscript	Direct Persian manuscript evidence
UNESCO (2022)	Human oversight, accountability, transparency, and responsible AI	Ethical framework

9. Conclusion

Based on the findings of this study, it can be concluded that artificial intelligence can be a valuable supporting technology to preserve the literary heritage of the Persians. It can be used to enhance digitisation, text recognition, cataloguing, creation of metadata, information retrieval and computational literary analysis. The interviews, and the reviewed literature, demonstrate that technology does not necessarily ensure authenticity of the text, historical accuracy, or maintenance of the meaning of the literature. AI-generated transcriptions are less accurate with damaged manuscripts and complicated calligraphy, as well as due to historic spelling variations. AI can also misinterpret the metaphors and cultural references, including Sufi symbolisms, especially in poets like Rumi, Hafez and Saadi. It is therefore important that every form of preservation also involves preserving the text along with the manuscript image, metadata, and historical context and not a mere scan of documents. The proposed Human-Centered AI Framework (HCAI-PH) views AI tools as auxiliary rather than a substitute for scholars' judgment. It prioritizes expert verification, ethical responsibility, interdisciplinary collaboration and sustainable digital infrastructure in the Centre. This human approach and collaboration prove to be a way forward for accessible, authentic and sustainable preservation of the Persian literary heritage. The framework still needs to be further empirically validated to demonstrate its effectiveness.

10. Recommendations

The results and findings suggest that there are a number of recommendations which can be put forward for preserving Persian literary heritage through the use of artificial intelligence. Firstly, human verification should remain the core of all AI supported processes. Any machine-generated transcriptions and/or interpretations should be read through by a Persian language or literature expert before being placed in any digital archive. The original manuscript images, separate machine-generated text from original text, should also be preserved in archives, as well as all corrections should be recorded transparently. Secondly, there is a need to invest on Persian specific resources, e.g. annotated datasets, OCR/HTR models for historical scripts. These would assist in solving the issues of difficult calligraphy, defaced leaves and variations in spelling. Thirdly, the preservation projects should promote interaction between Persian literature scholars, computer scientists, digital humanities scholars, librarians and archival institutions. Skill-based specialists should be trained and long-term funding and sustainable digital infrastructure should be provided to policymakers and heritage organizations to support this work. Fourth, projects need to comply with ethical and legal standards, such as by ensuring transparency, accountability and human oversight in accordance with the UNESCO (2022) principles. Additionally, they should be aware of respecting copyright and source attribution, as well as cultural sensitivity. Lastly, AI-powered tools should be assessed on their own, with a focus on transcription accuracy, error rates and metadata completeness, retrieval effectiveness, expert feedback, and user satisfaction. The **Human-Centered AI Framework (HCAI-PH)** should be applied on representative Persian manuscripts before starting to be widely adopted.

11. Limitation of the study

There are a number of limitations to the study which should be taken into account when interpreting the findings. It adopted a descriptive or non-experimental type of research design and qualitative and exploratory design so that the findings of the study cannot be statistically generalized to all the contexts of Persian literature. Participants of the interview were selected using purposive sampling. The sample may thus be small, and may not reflect all views in the fields of Persian studies, digital humanities and artificial intelligence. Interviews are also "interpretive" and may give voice to the opinions of the interviewees and/or the institutions where the interviews took place. Besides, the proposed HCAI-PH framework is a conceptual one, and has not been tested on actual Persian manuscripts, and therefore has not been proven to be effective in practice. Since a significant portion of the secondary literature focuses on issues of cultural heritage and handwriting recognition only in general, the amount of information pertinent to Persian manuscripts is limited. Since there was no actual benchmarking of any OCR, HTR or language model with Persian text in this study, accuracy was not measured. Lastly, AI technologies are evolving quickly, and some of the findings could be out of date.

Several of the sources cited are preprints or articles in press that have not yet been completely peer reviewed. Larger and more diverse samples, empirical testing and Persian specific technical evaluation are suggested for future research to fill the gaps.

Conflict of authors

The authors have declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

No funding is available.

12. Reference

Amini, S., Afshar Kazemi, M., Alam Tabriz, A., & Khatami Firouzabadi, S. M. (2026). Proposing an artificial intelligence model for extracting Persian question–answer in the tourism domain. *Journal of Business Intelligence Management Studies, 15*(55), 375–396. https://doi.org/10.22054/ims.2026.88566.2683

Arian, H., Bazargani, M., & Afrasiabpour, A. (2026). Ethical principles for applying artificial intelligence in preserving Persian lullabies. *[Journal title and publication details not provided]*. https://doi.org/10.48310/rpllp.2026.20360.1258

Berger, A. A., & Shahghasemi, E. (2026). Fear frames: Mapping AI anxiety in Persian social media discourse. *AI & Society*. Advance online publication. https://doi.org/10.1007/s00146-026-03214-3

Berger, A. A., & Shahghasemi, E. (2026). Vernacular futurism: How Persian language users imagine AI. *AI & Innovation*. Advance online publication. https://doi.org/10.1002/aiv2.70010

Farazmand, M., Asgari, M., Bouraghi, H., Khodaveisi, T., Mohammadpour, A., & Saeedi, S. (2025). Applications and role of artificial intelligence in the management of cardiac diseases: A systematic review of Persian research studies. *Journal of Payavard Salamat, 19*(3), 202–226. Article link

Gandomkar, M., Khoramipour, S., & Gandomkar, M. S. (2025). CHAMROSH: A Persian OCR dataset for Shahnameh manuscripts (Version 1.0) [Dataset]. GitHub. https://github.com/MojtabaGK/CHAMROSH-A-Persian-OCR-Dataset

Haghdadi, A., Pourmohammadbagher, L., & Safa, M. (2026). *Reviewing generative AI models for answering Persian historical questions*. *International Journal of Digital Content Management*. https://doi.org/10.22054/dcm.2026.90071.1293

Harmanda, T. T., Hidayanto, A. N., Aji, R. F., & Ilin, I. (2026). Artificial intelligence adoption in the cultural heritage sector: A systematic review of trends, application domains, and challenges. Computers in Human Behavior Reports, 23, Article 101147. https://doi.org/10.1016/j.chbr.2026.101147

Hosseini, S. A. (n.d.). *A comprehensive literature review of artificial intelligence in Persian literature education: Theories, applications, and ethical challenges* [Original research article in press]. *Articles in Press*. https://aikr.irandoc.ac.ir/?_action=press&_issue=-1&_is=Articles%20in%20Press

Hussain, B., Muradi, M., Bhuiyan, S. A., & Rasel, A. A. (2024). Unveiling Persian market dynamics: A comprehensive analysis of consumer demand using NLP techniques with explainable artificial intelligence. In *Proceedings of the 3rd International Conference on Computing Advancements* (pp. 452–459). Association for Computing Machinery. https://doi.org/10.1145/3723178.3723238

Mehrpouyan, A. (2026). Digital rhetoric and cyberethics in contemporary Iran: The application of artificial intelligence in the critical analysis and representation of literary and cultural texts. *Rhetoric and Grammar Studies, 16*(1), 155–172. https://doi.org/10.22091/jls.2026.15332.1785

Meymandi, A. R., Hosseini, Z., Davari, S., Moshiri, A., Rahimi-Golkhandan, S., Namdar, K., Feizi, N., Tavakoli-Targhi, M., & Khalvati, F. (2024). Opportunities for Persian digital humanities research with artificial intelligence language models: Case study: Forough Farrokhzad. arXiv. https://doi.org/10.48550/arXiv.2405.06760

Mirdar Rezaei, M., & Eslami, M. (2024). The negative encounter of artificial intelligence to the hermeneutic understanding of Persian poetry (case study: Verses from four styles of Persian poetry). *Literary Researches, 21*(81). https://doi.org/10.22080/lpr.2024.26907.1038

Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: A systematic review of Transkribus in published research. Archival Science, 22, 367–392. https://doi.org/10.1007/s10502-022-09397-0

Ostadi, S., & Heidari Tabrizi, H. (2025). AI-powered literary translation: A case study of Persian translations of *Charles*. *Translation and Interpreting Research, 2*(8), 43–53. https://doi.org/10.22054/tir.2026.21084

Rasool, B., & Pashaei, M. R. (2026). Stylistic analysis of Persian poetry and prose in the age of artificial intelligence: Explanation and feasibility of the intelligent literature framework. Journal of the Stylistic of Persian Poem and Prose, 18(12), 111–124. https://www.sid.ir/paper/1906391/en

Rasti Meymandi, A., Hosseini, Z., Davari, S., Moshiri, A., Rahimi-Golkhandan, S., Namdar, K., Feizi, N., Tavakoli-Targhi, M., & Khalvati, F. (2024). *Opportunities for Persian digital humanities research with artificial intelligence language models: Case study: Forough Farrokhzad*. arXiv. https://arxiv.org/abs/2405.06760

Rasti Meymandi, A., Hosseini, Z., Davari, S., Moshiri, A., Rahimi-Golkhandan, S., Namdar, K., Feizi, N., Tavakoli-Targhi, M., & Khalvati, F. (2024). Opportunities for Persian digital humanities research with artificial intelligence language models: Case study: Forough Farrokhzad [Preprint]. arXiv. https://arxiv.org/abs/2405.06760

Rezaei, M. M., & Eslami, M. (2024). The negative encounter of artificial intelligence to the hermeneutic understanding of Persian poetry (Case study: Verses from four styles of Persian poetry). *Journal of Literary–Philosophical Researches, 2*(1), 83–106. https://doi.org/10.22080/lpr.2024.26907.1038

Sadr, M. M., Mousavi Chelak, A., Ziaei, S., & Tanha, J. (2020). A predictive model based on machine learning methods to recognize fake Persian news on Twitter. *International Journal of Nonlinear Analysis and Applications, 11*(1), 1–?. https://doi.org/10.22075/IJNAA.2020.4563

Salehi, M., Ghahari, S., Hosseinzadeh, M., & Ghalichi, L. (2023). Domestic violence risk prediction in Iran using a machine learning approach by analyzing Persian textual content in social media. *Heliyon, 9*(5), Article e15667. https://doi.org/10.1016/j.heliyon.2023.e15667

Totaro, M. W., Gheisi, L., & Shahghasemi, E. (2026). Tool or deception: Persian discourse on AI-enabled misconduct. *Code, Cognition and Society, 1*(3), 1–24. https://doi.org/10.22034/ccs.2026.574018.1033

UNESCO. (2022). Recommendation on the ethics of artificial intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Varzandeh, M. (2026). *Building strong Persian language models is a strategic step to ensure Iran's active role in shaping the future of artificial intelligence*. SSRN. https://doi.org/10.2139/ssrn.7408258

Zargar, A., Moshiri, A., Shafaei, M., Rahimi-Golkhandan, S., Tavakoli-Targhi, M., & Khalvati, F. (2026). *Artificial intelligence for sentiment analysis of Persian poetry*. arXiv. https://arxiv.org/abs/2603.11254