

Integration of Artificial Intelligence in the Preservation of Regional Arts and Languages: A Digital Humanities Approach in Indonesia

Malathy Batumalay¹, Haryono^{2*}, Masrin³

¹Faculty of IT, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia

²Faculty of Engineering University of Bumigora, Mataram, Indonesia

³Postgraduate Faculty of Languages, University of Indraprasta PGRI, Jakarta, Indonesia

Corresponding Author's e-mail: haryono@universitasbumigora.ac.id



e-ISSN: 2964-2981

ARMADA : Jurnal Penelitian Multidisiplin

<https://ejournal.45mataram.ac.id/index.php/armada>

Vol. 03, No. 09, September, 2025

Page: 303-308

DOI:

<https://doi.org/10.55681/armada.v2i6.1756>

Article History:

Received: Agustus 20, 2025

Revised: September 10, 2025

Accepted: September 16, 2025

Abstract : The development of artificial intelligence (AI) technology has opened up new opportunities for the preservation of cultural heritage, including endangered regional arts and languages. This study aims to analyze how the application of AI in the context of digital humanities can support efforts to document, revitalize, and disseminate local culture in Indonesia. The method used is descriptive-qualitative with a literature study approach and policy analysis of digital culture preservation projects that have been running, such as Google AI for Indigenous Languages and the Digital Culture program of the Ministry of Education, Culture, Research, and Technology. The results show that the integration of AI through Natural Language Processing (NLP), machine translation, speech recognition, and digital archiving is able to accelerate the process of documentation and learning of regional languages and encourage the participation of the creative community in the preservation of traditional arts. However, the main challenges lie in the limitations of linguistic data, algorithmic bias, and technological gaps between regions. This study concludes that the successful implementation of AI-based digital humanities depends on multidisciplinary collaboration between governments, academics, cultural communities, and technology developers.

Keywords : Artificial intelligence, digital humanities, cultural preservation

Abstrak : Perkembangan teknologi kecerdasan buatan (AI) telah membuka peluang baru bagi pelestarian warisan budaya, termasuk seni dan bahasa daerah yang terancam punah. Penelitian ini bertujuan untuk menganalisis bagaimana penerapan AI dalam konteks humaniora digital dapat mendukung upaya pendokumentasian, revitalisasi, dan diseminasi budaya lokal di Indonesia. Metode yang digunakan adalah deskriptif-kualitatif dengan pendekatan studi pustaka dan analisis kebijakan terhadap proyek-proyek pelestarian budaya digital yang telah berjalan, seperti Google AI untuk Bahasa Daerah dan program Budaya Digital Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Hasil penelitian menunjukkan bahwa integrasi AI melalui Pemrosesan Bahasa Alami (NLP), penerjemahan mesin, pengenalan suara, dan pengarsipan digital mampu mempercepat proses dokumentasi dan pembelajaran bahasa daerah serta mendorong partisipasi komunitas kreatif dalam pelestarian seni tradisional. Namun, tantangan utamanya

terletak pada keterbatasan data linguistik, bias algoritmik, dan kesenjangan teknologi antardaerah. Penelitian ini menyimpulkan bahwa keberhasilan implementasi humaniora digital berbasis AI bergantung pada kolaborasi multidisiplin antara pemerintah, akademisi, komunitas budaya, dan pengembang teknologi.

Kata Kunci : Kecerdasan buatan, humaniora digital, pelestarian budaya

INTRODUCTION

Indonesia, as a country with great linguistic and cultural diversity, faces two challenges: (1) the threat of extinction of a number of local languages and art traditions due to modernization and globalization, and (2) limited technical capacity and resources to systematically document, archive, and revitalize culture. This condition is driving the emergence of new approaches that combine traditional humanities methods with digital technologies often called digital humanities or digital humanities where artificial intelligence (AI) presents as a potentially transformative tool to accelerate and expand the scale of conservation efforts (Münster *et al.*, 2024; Foka & Griffin, 2024).

AI offers capabilities for transcription automation, speech recognition, machine translation, image classification, metadata extraction, and the creation of 3D representations and immersive experiences (VR/AR) that can enrich the way people access and learn about cultural heritage (Münster *et al.*, 2024; Romero *et al.*, 2024). On the linguistic side, the advancement of learning techniques is represented in self-supervised architectures such as wav2vec 2.0 which allows the acquisition of speech representations from unlabeled data, thereby improving speech recognition performance for languages with limited data sources, an important development for regional languages that are classified as low-resource (Baevski *et al.*, 2020; Romero *et al.*, 2024).

Empirical studies in the context of indigenous languages show that ASR (automatic speech recognition) combined with data augmentation techniques and the use of pre-trained models can result in significant performance improvements even though the corpus is small (Romero, Gómez-Canaval, & Torre, 2024; Zhao & Zhang, 2022). In the realm of traditional performing arts and dance, motion capture technology and machine learning techniques allow for highly detailed motion documentation, motion templateing, and digital reconstruction that can be used for education and revitalization (Mohd Herrow & Azraai, 2021; Chang-E dataset, 2024).

In the aspect of images and material objects, methods based on convolutional neural networks (CNN) have been shown to be able to classify artifacts, detect damage, and assist with image analysis-based restoration, although it still requires high-quality annotation data and attention to model interpretability issues (Vadineanu, 2024; Linardatos, Papastefanopoulos, & Kotsiantis, 2020). However, behind this technical potential, the critical literature shows serious risks if the application of AI is not designed with humanistic and ethical principles in mind: algorithmic bias can reinforce dominant narratives or diminish local variation; Distributed, heterogeneous, and often generated cultural data without standard metadata raises research reconnectivity and reproducibility issues; and questions of ownership, cultural rights, and accessibility become increasingly urgent when digital artifacts and records are easily replicated and disseminated (Foka & Griffin, 2024; Focus, 2025; Tracing the bias loop, 2025). Therefore, the concept of digital humanism and digital ethical principles that emphasize the involvement of local communities, respect for cultural authority, transparency of technological processes, and the protection of rights to cultural data must be integrated from the design stage of AI projects for culture (Prem, 2024; Münster *et al.*, 2024).

In addition to ethical issues, operational challenges are also real: the availability of digital infrastructure in remote areas, the capacity of local human resources to operate and maintain systems, and the availability of a language corpus and sufficient audiovisual collections to train AI models are still practical obstacles (Bromham *et al.*, 2022; Münster *et al.*, 2024). The experience of projects in different countries shows that participatory approaches in which native speakers, traditional artists, local curators, and computer scientists collaborate provide the most sustainable outcomes, as models trained with co-curated data tend to be more contextually relevant and more accepted by communities

(Romero *et al.*, 2024; Foka, 2025). On the policy side, a growing body of studies is calling for a balanced regulatory framework: supporting open access for educational and research purposes while establishing mechanisms for the protection of collective rights to traditional knowledge from commercial exploitation without community consent (Westenberger, 2025; Münster *et al.*, 2024).

Ultimately, the integration of AI into the preservation of art and regional languages should be seen not merely as a substitution of ethnographic practices but as a complement that expands the reach of documentation, enables large-scale analysis, and creates new forms of presentation that are interactive and engaging to young generations (Münster *et al.*, 2024; Foka & Griffin, 2024). But the success of this transition requires a clear R&D roadmap, investment in the development of a quality corpus, local competency training, and bias-monitoring and human-in-the-loop mechanisms to ensure technical decisions are always validated by humanities expertise (Linardatos *et al.*, 2020; Tracing the bias loop, 2025).

In the Indonesian context, this opens up opportunities for collaboration between institutions: universities with NLP expertise and technological vision, cultural institutions that maintain collections and community networks, and policymakers who can formulate metadata standards, ethics, and long-term financing models of a multi-stakeholder ecosystem that are key for AI to truly function as a tool of cultural empowerment, not just as a repressive technology or simply an object of modernization (Münster *et al.*, 2024; Bromham *et al.*, 2022; Foka, 2025).

Drawing on a diverse literature from technical studies on ASR and self-supervised representation to ethical and policy studies, this article will outline the potential, limitations, and practical recommendations for integrating AI in arts and regional language preservation efforts in Indonesia, with a particular emphasis on participatory strategies, bias mitigation, and equitable data governance (Baevski *et al.*, 2020; Romero *et al.*, 2024; Prem, 2024; Foka & Griffin, 2024).

METHOD

This research uses a descriptive-qualitative approach with the main objective of understanding in depth the dynamics of the application of artificial intelligence (AI) in the context of preserving art and regional languages through the perspective of digital humanities. This approach was chosen because it is able to explore the meaning, context, and social and cultural interactions that underlie the phenomenon of the use of technology in the cultural realm, rather than just measuring its technical or quantitative aspects. The methods used include literature studies and analysis of policy documents, both of which serve to identify patterns, strategies, and trends in the application of AI in cultural preservation programs in Indonesia during the period 2018–2024.

Research data was collected from various secondary sources, including scientific articles published in reputable international and national journals, reports from cultural institutions such as UNESCO and the Ministry of Education and Culture, government policies related to cultural digitalization, and reports on AI-based digital projects and initiatives run by local communities and higher education institutions. All of the data were purposively selected based on relevance to the topic of digital culture preservation and its suitability with the digital humanities framework.

The analysis stage was carried out by tracing the pattern of AI use in three main sectors, namely: (1) digitization and revitalization of regional languages through Natural Language Processing (NLP) technology and automatic speech recognition; (2) archiving and preservation of traditional art using image recognition, motion capture, and digital archiving systems; and (3) participatory innovation based on cultural communities, where AI is used to encourage collaboration between local communities and technology developers in the process of documenting and disseminating artworks.

Each sector is analyzed based on three evaluative dimensions: the effectiveness of the technology, which includes the success of AI in accelerating the documentation and learning process; socio-cultural impacts, i.e. the extent to which the application of this technology strengthens or even changes local traditional practices; as well as conformity with the principles of digital humanities, which emphasize the values of openness, cross-disciplinary collaboration, and long-term sustainability. This analytical approach allows researchers to identify both the potential

and challenges of the integration of AI in cultural preservation, as well as provide strategic recommendations that can be applied in national policies and community initiatives in the future.

RESULTS AND DISCUSSION

The results of the study show that the integration of artificial intelligence (AI) in the preservation of art and regional languages in Indonesia is still in its early stages, but has shown significant development in the realms of technology, social, and cultural policies. Based on the results of the literature analysis and policy documents, three main sectors that have made progress through the application of AI can be identified, namely digitization of regional languages, archiving of traditional arts, and community-based innovation. In the first sector, namely the digitization of regional languages, AI technology plays a major role in efforts to document, analyze linguistics, and revitalize minority languages in Indonesia which are classified as low-resource languages. Through the application of Natural Language Processing (NLP) and Automatic Speech Recognition (ASR), several research projects at universities and government institutions have succeeded in developing a deep learning-based language model that is able to recognize, transcribe and translate regional languages such as Sundanese, Javanese, and Sasak. The wav2vec 2.0 model developed by Baevski *et al.* (2020) is one of the main technical references due to its ability to extract phonetic features from voice data without the need for large labels, making it suitable for application to languages that have a limited corpus. A similar approach was used by Romero, Gómez-Canaval, and Torre (2024) who developed an ASR system for indigenous languages in Latin America, resulting in an accuracy of up to 87% despite limited training data. The adaptation of this method in the Indonesian context has been proven to increase the efficiency of digitization of regional languages that previously relied on manual documentation. However, problems still arise in the aspect of audio data quality, limitations of linguistic annotators, and model bias due to the phonetic dominance of Indonesian or English in the training data. This requires rigorous data curation and active collaboration between native speakers, linguists, and AI developers so that the resulting system truly represents the uniqueness of the local language in an authentic manner.

In the second sector, namely the archiving and preservation of traditional art, AI is used to digitize and visually document works of art, dance, music, and cultural artifacts with a high level of precision. Computer vision technology, especially based on convolutional neural networks (CNN), is able to recognize complex visual patterns of artifacts such as woven fabric motifs, wood carvings, or traditional dance movement forms (Vadineanu, 2024). Several institutions such as the National Gallery of Indonesia and the National Archives have begun to implement image recognition algorithms to catalog cultural collections, simplify classification, and accelerate public access to digital archives. In an international context, Mohd Herrow and Azraai (2021) show how motion capture technology is used to document the traditional dance movements of Joget in Malaysia, resulting in a motion database that can be accessed digitally and used in art education. This approach can be adapted to trace Indonesian regional dance movements such as Saman dance or Reog Ponorogo, where the characteristics of movement and rhythm can be stored as digital data for the preservation of future generations. In addition, AI also plays a role in digital restoration, for example to improve the quality of old photos or videos as well as in the creation of virtual museums that allow users to explore cultural heritage through virtual reality. The study of Münster *et al.* (2024) confirms that the application of AI in cultural heritage management in Europe has been able to increase the efficiency of documentation by up to 40% compared to manual methods, especially in terms of metadata management and visual pattern recognition. In the Indonesian context, this achievement can be optimized by expanding collaboration between cultural institutions, academics, and the digital creative industry to accelerate the digitization of traditional arts which have been hampered by limited human resources and infrastructure.

Meanwhile, the third sector that is the focus of the analysis is participatory innovation based on cultural communities, which places local communities as the main actors in the process of digitization and the use of AI. The participatory approach of digital humanities encourages the involvement of native speakers, artists, and indigenous communities in the process of data collection, content validation, and management of digitization results (Foka & Griffin, 2024). Thus, cultural preservation is not only a top-down technological project, but also a social transformation

that strengthens local identity through technology. In some regions, for example in West Nusa Tenggara and South Sulawesi, young communities have begun to develop NLP-based chatbots that are able to answer questions in regional languages for educational purposes and cultural promotion. Collaborative models like this show how AI can function as a means of empowerment, not just a passive documentation tool. However, the literature also suggests that community involvement without clear ethical guidance can lead to conflicts of data ownership and cultural representation (Prem, 2024). Therefore, the development of AI projects for cultural preservation needs to be based on the principles of digital humanism that emphasizes social justice, transparency, and the right to cultural data as a form of collective intellectual heritage (Prem, 2024; Foka & Griffin, 2024).

In addition to the technological benefits, this study also highlights a number of ethical challenges and risks in the integration of AI in the cultural realm. One of the main problems is algorithmic bias that can cause certain cultural representations to become dominant while others are marginalized. Foka (2025) emphasizes that these biases often arise due to data inequality and system design that do not take into account socio-cultural variations. In the context of regional languages, for example, AI trained using data from dominant languages such as Javanese or Indonesian can produce less accurate models for minority languages. Another risk is the issue of data ownership and cultural intellectual property, where the digitization of cultural artifacts has the potential to be used commercially without permission from the original owner. Therefore, a legal protection policy is needed that is in line with the principle of Free, Prior, and Informed Consent (FPIC) as applied in community-based cultural preservation projects in other countries (Westenberger, 2025).

In general, the results of this study reinforce the view that the success of AI integration in the preservation of art and regional languages does not depend solely on the sophistication of the technology, but also on the extent to which the technology is adopted with human-centered principles that respect the social context and human values. AI can be a means to expand the reach of cultural preservation enabling large-scale documentation, automated translation, and cross-lingual semantic analysis but still must be balanced with participatory approaches and ethical awareness. By combining technological innovation, progressive public policy, and local community participation, Indonesia has a great opportunity to become an example of the application of AI-based digital humanities in Southeast Asia. Intersectoral collaboration between academics, cultural institutions, technology developers, and indigenous peoples needs to be strengthened to build an inclusive, adaptive, and sustainable digital ecosystem. In this context, AI is not just a documentation tool, but part of a new cultural ecosystem that connects the heritage of the past with the nation's digital future.

CONCLUSION

The integration of artificial intelligence (AI) in the preservation of arts and regional languages opens up great opportunities for the transformation of digital humanities in Indonesia, which places technology as a catalyst in maintaining and developing national cultural wealth. The use of technologies such as Natural Language Processing (NLP), machine learning, and digital archiving systems not only speeds up the documentation process, but also allows for a wider and interactive dissemination of local culture. NLP, for example, can be used to recognize, analyze, and translate regional languages that were previously vulnerable to extinction due to a lack of active speakers. This technology is able to convert speech into text, standardize spelling, and create a digital dictionary that can be accessed by the public.

On the other hand, machine learning plays an important role in the grouping of art motifs, pattern recognition in traditional music, and visual reconstruction of damaged artifacts. This technology allows cultural heritage to be revived in a digital form that is close to the original, while opening up learning spaces across generations and regions. However, the successful application of this technology depends not only on the sophistication of the algorithm, but also on how humanitarian, ethical values, and local wisdom are integrated into the process. Without a deep understanding of the social and cultural context, digitalization risks reducing cultural significance to mere technical data. Therefore, participatory principles should be the foundation in any cultural

AI initiative, where indigenous peoples, artists, and native speakers play an active role in data collection, content validation, and interpretation of digitization results.

Close collaboration is needed between governments, academics, technology developers, and cultural communities to build an inclusive, adaptive, and sustainable digital humanities ecosystem. With this approach, AI does not function as a substitute for humans, but rather as a tool that expands reach, enriches interactions, and deepens understanding of the nation's cultural heritage, making it relevant in the digital age without losing its authentic values.

REFERENCES

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). *wav2vec 2.0: A framework for self-supervised learning of speech representations*. In NeurIPS 2020 Proceedings. <https://doi.org/10.5555/3495724.3496768>

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2022). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, 6(2), 163–173. <https://doi.org/10.1038/s41559-021-01604-y>

Casillo, M. (2025). Artificial intelligence in archaeological site conservation. *Journal of Computer Applications in Archaeology*. <https://doi.org/10.5334/jcaa.207>

Chang-E dataset authors. (2024). *Chang-E: A high-quality motion capture dataset of Dunhuang dance*. ACM. <https://doi.org/10.1145/3709000>

Foka, A. (2025). Tracing the bias loop: AI, cultural heritage and bias-mitigating in practice. *AI & Society*. <https://doi.org/10.1007/s00146-025-02349-z>

Foka, A., & Griffin, G. (2024). AI, cultural heritage, and bias: Some key queries that arise from the use of GenAI. *Heritage*, 7(11), 6125–6136. <https://doi.org/10.3390/heritage7110287>

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of machine learning interpretability methods. *Entropy*, 23(1), 18. <https://doi.org/10.3390/e23010018>

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). *Explainable AI: A review of machine learning interpretability methods*. *Entropy*. <https://doi.org/10.3390/e23010018>

Mohd Herrow, M. F., & Azraai, N. Z. (2021). Digital preservation of intangible cultural heritage of Joget dance movement using motion capture technology. *International Journal of Heritage, Art and Multimedia*, 4(15), 01–13. <https://doi.org/10.35631/IJHAM.415001>

Münster, S., Lehmann, C., Lazariv, T., Maiwald, F., Karsten, S., et al. (2024). Artificial intelligence for digital heritage innovation: Setting up a R&D agenda for Europe. *Heritage*, 7(2), 794–816. <https://doi.org/10.3390/heritage7020038>

Pagnin, L. (2023). Protecting street art from outdoor environmental threats: Review and perspectives. *Coatings*, 13(12), 2044. <https://doi.org/10.3390/coatings13122044>

Prem, E. (2024). Principles of digital humanism: A critical post-humanist view. *Journal of Responsible Technology*, 17(C), 100075. <https://doi.org/10.1016/j.jrt.2024.100075>

Romero, M., Gómez-Canaval, S., & Torre, I. G. (2024). Automatic speech recognition advancements for indigenous languages of the Americas. *Applied Sciences*, 14(15), 6497. <https://doi.org/10.3390/app14156497>

Vadineanu, S. (2024). Convolutional neural networks and their activations: An exploratory case study on mounded settlements. *Journal of Computer Applications in Archaeology*, 163. <https://doi.org/10.5334/jcaa.163>

Westenberger, P. (2025). Artificial intelligence for cultural heritage research: Copyright, policy and practice. *European Journal of Cultural Management and Policy*. <https://doi.org/10.3389/ejcmp.2025.14009>