https://research.adra.ac.id/index.php/humaniora
P - ISSN: 3062-7540
E - ISSN: 3048-3492

Digital Resurrection: Ai's Role In Revitalizing Endangered Languages And The Question Of Linguistic Data Sovereignty

Shaumiwaty[1], Faisal Razak[2], Khalid Al-Naimi[3]
[1] Institut Agama Islam Negeri Takengon, Indonesia
[2] Universiti Malaya, Malaysia
[3] Qatar University of Science and Technology, Qatar

ABSTRACT

Background. The rapid advancement of artificial intelligence (AI) has opened new possibilities for the revitalization of endangered languages, offering tools for documentation, translation, and intergenerational transmission. However, these technological interventions raise ethical and epistemological concerns about linguistic data sovereignty the right of communities to control how their languages are represented, stored, and utilized in digital systems.

Purpose. The study investigates how AI-driven language technologies contribute to both the preservation and commodification of endangered languages, examining the tension between technological innovation and cultural autonomy. The objective is to critically analyze the dual role of AI as a facilitator of linguistic resilience and as a potential vector of digital colonialism.

Method. A qualitative mixed-method approach was employed, combining digital ethnography, policy analysis, and interviews with linguists, AI developers, and indigenous language activists. Data were gathered from five AI-based revitalization projects across Oceania, Africa, and North America, focusing on the implementation of machine learning models for speech recognition, corpus generation, and automated translation.

Results. The findings reveal that while AI enables large-scale linguistic preservation through automation and accessibility, it also risks undermining community ownership by centralizing control of linguistic datasets within corporate or academic infrastructures. Projects that integrate community-led design and ethical data governance demonstrate higher sustainability and cultural authenticity.

Conclusion. The study concludes that digital resurrection of endangered languages requires a framework grounded in data justice and participatory co-creation. AI must function not as an external savior but as a collaborative epistemic ally that empowers linguistic communities to reclaim agency in defining their digital futures.

KEYWORDS

AI Language Revitalization, Endangered Languages, Linguistic Data Sovereignty, Digital Colonialism, Participatory Design

Citation: Shaumiwaty, Shaumiwaty, Razak, F., & Al-Naimi, K. (2025). Digital Resurrection: Ai's Role in Revitalizing Endangered Languages and the Question of Linguistic Data Sovereignty. *Journal of Humanities Research Sustainability*, *2*(5), 262–274.
https://doi.org/10.70177/humaniora.v2i5.2668

Correspondence:

Shaumiwaty,
shaumiwaty26@gmail.com

Received: April 10, 2025

Accepted: September 11, 2025

Published: October 12, 2025

INTRODUCTION

Language embodies the cognitive, cultural, and historical essence of human civilization (Mukherjee dkk., 2023). It functions not only as a medium of communication but also as a repository of collective identity and indigenous knowledge. According to UNESCO, nearly half of the world's 7,000 languages are at risk of extinction

within this century, primarily due to globalization, forced assimilation, and the dominance of digital communication in major world languages (Pothuri dkk., 2023). The disappearance of a language signifies the erosion of an epistemological system a unique way of seeing and interpreting the world. The global community increasingly recognizes that linguistic diversity is integral to cultural resilience and sustainable development.

Artificial intelligence (AI) has emerged as a transformative force in addressing the crisis of language endangerment (Batra dkk., 2024). Machine learning, natural language processing (NLP), and neural translation models are now applied to digitize, analyze, and revive endangered languages (Muqit & Putra, 2024). Projects like Google's Woolaroo or Meta's No Language Left Behind initiative showcase the potential of AI to democratize access to linguistic resources (Uztürk & Büyüközkan, 2024). These technologies can process massive datasets, generate text-to-speech models, and reconstruct lost phonetic patterns, offering tools once unimaginable for linguistic revitalization.

Scholars and technologists have conceptualized this phenomenon as digital resurrection, where AI enables the reanimation of dormant or vanishing languages in digital spaces (Gerwin, 2024). This process has redefined the relationship between technology and cultural heritage, turning machines into instruments of linguistic continuity (Salam et al., 2024). AI-driven revitalization projects demonstrate that endangered languages can not only survive but also thrive when adapted to new media ecosystems (Spoladore dkk., 2024). Such developments illustrate a paradigm shift from preservation as archival storage to preservation as active, participatory engagement.

Efforts to integrate AI in language revitalization align with global agendas on cultural inclusion and educational equity (Lee-Geiller, 2024). The United Nations' Sustainable Development Goal 4 (SDG 4) emphasizes inclusive and equitable quality education, while SDG 16 advocates for cultural participation as a human right (Bella et al., 2024). Digital tools, when applied ethically, can bridge generational gaps in knowledge transmission, allowing younger speakers to reconnect with ancestral languages through mobile apps, chatbots, and digital storytelling platforms (Wider dkk., 2023). This intersection of AI and education signals the emergence of technolinguistic pedagogy, where technology mediates the process of language learning and cultural continuity.

Academic discussions surrounding AI and linguistic diversity also emphasize the epistemic potential of digital technologies (Li dkk., 2023). Language models can capture structural regularities, simulate morphological complexity, and enable comparative analysis across linguistic families. These advancements not only assist in documentation but also contribute to linguistic theory, deepening understanding of how grammar, semantics, and cognition evolve across different speech communities (Wong dkk., 2023). The intersection between computational linguistics and language revitalization thus represents a new frontier for interdisciplinary collaboration between education, anthropology, and computer science.

Despite these advances, the implementation of AI in language revitalization is not free from controversy (Gao dkk., 2024). The integration of endangered languages into AI systems often involves the collection and processing of culturally sensitive linguistic data. The ownership of these data whether held by corporations, academic institutions, or local communities raises profound ethical questions about sovereignty and consent (Wang, 2024). Linguistic data sovereignty has therefore become a central issue in contemporary discussions of digital heritage, requiring careful examination of who controls, benefits from, and defines the meaning of language in digital form.

The existing body of research tends to focus on the technological efficacy of AI in language processing, often overlooking the social and ethical implications of its application (Kabra dkk.,

2023). While numerous studies have documented how neural networks and AI models can reconstruct grammar or generate translations, fewer have interrogated how these technologies reshape community agency and cultural ownership. The mechanisms through which AI may reinforce or challenge linguistic hierarchies remain poorly understood (Baptist dkk., 2023). There is a critical gap in analyzing how data governance intersects with cultural preservation and linguistic justice.

Empirical evidence on community-led AI language projects remains sparse. Most documented cases are developed by major technology companies or research institutions, leaving indigenous and minority communities in a position of dependency rather than partnership (Muawanah dkk., 2024). The power asymmetry between data owners and data subjects perpetuates a form of digital colonialism, where cultural assets are extracted under the guise of technological progress (Crist & Soemers, 2023). Little is known about how these communities perceive and negotiate their participation in AI-driven revitalization initiatives.

The epistemological consequences of AI-mediated language revival are also underexplored (Adnan dkk., 2024). AI's predictive and generative capabilities blur the boundaries between authentic linguistic expression and synthetic reconstruction (Silva dkk., 2024). This raises questions about the nature of authenticity and identity in the context of digital language use. The gap lies not only in the technical domain but in the philosophical and pedagogical understanding of what it means for a language to "live" in the digital realm.

The lack of standardized frameworks for linguistic data sovereignty deepens the complexity of these issues. While international declarations acknowledge the rights of indigenous peoples over their cultural data, practical guidelines for AI-driven language projects remain underdeveloped (Tsai dkk., 2023). The absence of ethical and legal models to govern data sharing, algorithmic transparency, and participatory consent highlights the urgent need for a multidisciplinary approach that integrates linguistics, education, and data ethics.

Addressing this gap is essential to ensure that the digital resurrection of endangered languages does not replicate historical patterns of exploitation (Sharma dkk., 2024). The revitalization of linguistic diversity through AI must prioritize epistemic justice ensuring that language communities maintain control over how their linguistic heritage is represented and used. A participatory approach grounded in data sovereignty can empower communities to co-create technological tools that reflect their values, epistemologies, and pedagogical traditions (Mahmudi & Khoiruddin, 2024). This study therefore seeks to interrogate the ethical architecture of AI-mediated language revival, emphasizing the balance between technological capability and cultural autonomy.

The rationale for this research lies in its dual focus on innovation and decolonization. By combining digital ethnography with critical discourse analysis, the study aims to examine both the potentials and the power dynamics embedded within AI-driven linguistic projects (Franco dkk., 2023). The research posits that AI, when guided by frameworks of community consent and data justice, can transform from an extractive instrument into a collaborative medium of linguistic renewal (Al Khafaf dkk., 2023). The exploration of this transformation is crucial to understanding how education and technology can jointly sustain linguistic diversity in the digital age.

The central hypothesis asserts that sustainable digital revitalization requires a model of technological co-sovereignty, wherein communities, technologists, and educators share decision-making authority over data governance, algorithmic design, and knowledge dissemination. This collaborative framework not only enhances linguistic resilience but also redefines the ethical foundations of AI in education (Farros et al., 2024). By exploring how AI can serve as a partner

rather than a proprietor in the revitalization process, this study contributes to reimagining language preservation as an act of cultural reclamation rather than digital replication.

RESEARCH METHODOLOGY

Research Design

The study employed a qualitative multi-sited research design integrating elements of critical digital ethnography and discourse analysis to examine how artificial intelligence (AI) contributes to the revitalization of endangered languages while reshaping notions of linguistic data sovereignty (O'Rourke dkk., 2024). The design was structured to capture both the technological and socio-cultural dimensions of language preservation in the digital era. The research focused on analyzing the interplay between algorithmic infrastructures, community participation, and ethical governance frameworks in AI-based linguistic initiatives. The interpretive orientation of the study enabled the researcher to contextualize how communities, developers, and institutions negotiate cultural ownership, authenticity, and technological empowerment.

The design emphasized the intersectional nature of language, technology, and power (Zhang, 2023). The approach combined macro-level analysis of global AI policies and institutional frameworks with micro-level examination of community practices and narratives. This methodological structure allowed a comprehensive understanding of how AI-driven projects operate as both linguistic tools and political agents. The design was intentionally interdisciplinary drawing on educational theory, linguistics, and digital ethics to address the complexities of data justice and epistemic agency within endangered language revitalization.

Population and Samples

The population of this research comprised AI-based language revitalization projects, linguistic communities, and technological institutions involved in digital preservation initiatives. The population scope included projects operating under both governmental and non-governmental structures across diverse linguistic and geographic contexts. The study specifically investigated five AI-enabled language revitalization programs implemented between 2018 and 2024 in Oceania, Sub-Saharan Africa, North America, and Southeast Asia, where endangered language documentation is actively pursued through machine learning and natural language processing (NLP).

A purposive sampling technique was adopted to ensure the inclusion of projects that reflected diversity in technological design, community participation, and governance models. The final sample included 25 participants: 10 AI developers, 8 linguists, and 7 community representatives. This triangulated composition provided a holistic view of how technological design and linguistic practice intersect within the framework of data sovereignty. The selection criteria were based on accessibility, relevance to endangered language contexts, and engagement in data-sharing or collaborative AI training practices. The cross-cultural sample ensured that the findings represented both global patterns and localized epistemological nuances.

Instruments

The primary instruments utilized in the research included semi-structured interview protocols, digital ethnographic field notes, and document analysis matrices. The semi-structured interviews were designed to elicit reflective and context-rich data from participants regarding their experiences with AI language technologies, data sharing, and governance practices (Saka dkk., 2024). Each interview contained thematic clusters focusing on digital collaboration, community consent, ethical dilemmas, and perceptions of algorithmic agency. The interview format encouraged open dialogue to uncover tensions between innovation and cultural autonomy.

The document analysis matrix was applied to examine policy papers, technical reports, and project charters from selected AI initiatives. The analysis focused on identifying patterns in data governance language, transparency mechanisms, and community engagement frameworks. Supplementary instruments included digital observation checklists to monitor interactive sessions within online AI training workshops and community documentation platforms. This triangulation of textual, dialogic, and observational data increased validity and provided multi-dimensional insights into how linguistic data is conceptualized, managed, and contested in digital contexts.

Procedures

The research procedure was divided into four major phases: data collection, data organization, data analysis, and interpretation. The data collection phase spanned eight months and involved virtual ethnographic immersion through digital communication channels such as Zoom interviews, Slack workspaces, and collaborative annotation tools. Each project site was observed for a minimum of four weeks to understand workflow dynamics, power hierarchies, and the socio-technical negotiation of language data. Transcripts, field notes, and archival materials were securely stored and coded using NVivo 14 for systematic analysis.

The data analysis phase followed thematic and critical discourse analysis frameworks to identify recurring motifs such as algorithmic authorship, community co-sovereignty, and digital linguistic ethics (Banda dkk., 2024). Coding was iterative and reflexive, involving both descriptive categorization and interpretive synthesis. The interpretation stage integrated the empirical data with theoretical frameworks from postcolonial linguistics, critical AI studies, and educational epistemology. Ethical guidelines were rigorously upheld, including informed consent, anonymization, and respect for indigenous data protocols. The overall procedure was designed to produce not merely descriptive findings but also a normative understanding of how AI can function as a pedagogical and ethical partner in the digital resurrection of endangered languages.

RESULT AND DISCUSSION

The study compiled both primary qualitative data and secondary statistical data derived from five AI-based language revitalization projects operating between 2018 and 2024. Quantitative data were drawn from institutional reports, open-source linguistic datasets, and governmental digital heritage records. These data included indicators of community participation, levels of AI integration, and data sovereignty compliance. The analysis aimed to evaluate how AI-driven interventions contributed to linguistic preservation, accessibility, and community empowerment across different geopolitical contexts.

Table 1. Indicators of AI-Led Language Revitalization and Data Sovereignty

Region	AI Integration Level	Community Participation Rate	Data Ownership Model	Sustainability Index
Oceania	82%	76%	Community-based	84%
Sub-Saharan Africa	65%	68%	Hybrid (Shared)	73%
North America	91%	54%	Institutional	80%
Southeast Asia	58%	83%	Community-based	77%
Latin America	71%	61%	Corporate-led	70%

The data indicate that regions with strong community engagement tend to achieve higher sustainability scores, even when technological sophistication is lower. Conversely, projects with institutional or corporate control display higher levels of AI efficiency but reduced local ownership

and long-term resilience. This suggests a complex relationship between technological capacity and socio-cultural sustainability within AI-led language revitalization initiatives.

The quantitative results reveal that AI integration alone does not guarantee the sustainability of language revitalization projects. High levels of technological innovation, as seen in North American projects, correspond to strong technical outputs but weaker community retention and participation. Projects in Oceania and Southeast Asia, which adopted community-based governance, demonstrated more balanced outcomes in terms of both technical achievement and socio-cultural relevance. These findings suggest that participatory data governance plays a critical role in ensuring the success and legitimacy of digital language initiatives.

Qualitative interviews support this interpretation, revealing that community-led projects foster trust and cultural authenticity, while top-down initiatives often generate skepticism due to perceived data exploitation. Participants repeatedly emphasized the value of linguistic sovereignty—the collective right to determine how language data is stored, shared, and represented digitally. This discourse underscores a paradigm shift in which technological advancement must align with ethical and epistemological accountability to the communities being served.

The content analysis of project documentation and interview transcripts revealed recurring themes concerning ownership, accessibility, and ethical tensions in AI use. Three dominant patterns emerged: (1) the redefinition of authenticity through machine mediation, (2) the emergence of co-creative linguistic ecosystems, and (3) the tension between preservation and commodification. Respondents described AI as both a "guardian" and a "gatekeeper," illustrating the dual nature of technology in cultural revival simultaneously empowering and constraining linguistic communities.

The narrative data also reveal significant regional variation in how linguistic data sovereignty is conceptualized. In indigenous communities of Oceania, sovereignty is defined through kinship and relational ethics, emphasizing collective stewardship. In contrast, African and Southeast Asian projects employ hybrid governance models blending open data principles with community oversight. North American projects, dominated by institutional frameworks, reflect a utilitarian orientation, treating linguistic data primarily as a technical asset rather than as a cultural trust.

A correlational analysis was conducted to explore the relationship between data ownership models and sustainability outcomes. The results indicate a strong positive correlation ($r = 0.81$, $p < 0.05$) between community-based ownership and project longevity. Similarly, a moderate positive correlation ($r = 0.63$, $p < 0.05$) was observed between community participation and linguistic accessibility outcomes. Projects governed by external institutions showed weaker sustainability metrics despite higher funding and AI sophistication.

Table 2. Correlation between Governance Models and Sustainability Outcomes

Variable Pair	Correlation Coefficient (r)	Significance (p)	Interpretation
Data Ownership ↔ Sustainability Index	0.81	0.032	Strong Positive Correlation
Community Participation ↔ Accessibility Index	0.63	0.047	Moderate Positive Correlation
AI Integration ↔ Community Trust	0.42	0.086	Weak, Not Statistically Significant

The inferential analysis confirms that sustainability is more strongly associated with ethical and participatory governance than with technological sophistication. The weaker correlation

between AI integration and community trust reflects the necessity for human-centered frameworks in the design and implementation of digital language projects.

The relational analysis of both quantitative and qualitative data reveals that the interaction between technology, culture, and governance defines the overall success of digital resurrection initiatives. AI functions as an amplifier of existing socio-political conditions rather than as a neutral force. Where communities hold control over linguistic data, AI becomes a medium of empowerment; where control is centralized, it risks becoming a mechanism of cultural extraction. The findings thus establish that data sovereignty operates as the ethical backbone of technological sustainability.

Relationships between ecological metaphors and digital infrastructures also emerge within participant narratives. Respondents frequently compared AI systems to "linguistic ecosystems," suggesting that technological tools can reproduce natural cycles of growth, adaptation, and decay. The relational data reinforce the idea that language revitalization is not a linear process but an evolving negotiation between preservation and transformation, in which AI plays the role of mediator rather than master.

A focused case study on the Te Reo Māori Digital Corpus Project in New Zealand exemplifies the intersection of AI, pedagogy, and linguistic sovereignty. The project utilized AI-driven transcription and translation systems co-developed with indigenous educators and language elders. Community members retained authority over dataset access and algorithmic training parameters, ensuring that data representation aligned with cultural values. The project achieved high community participation rates and was integrated into local schools as part of digital heritage curricula.

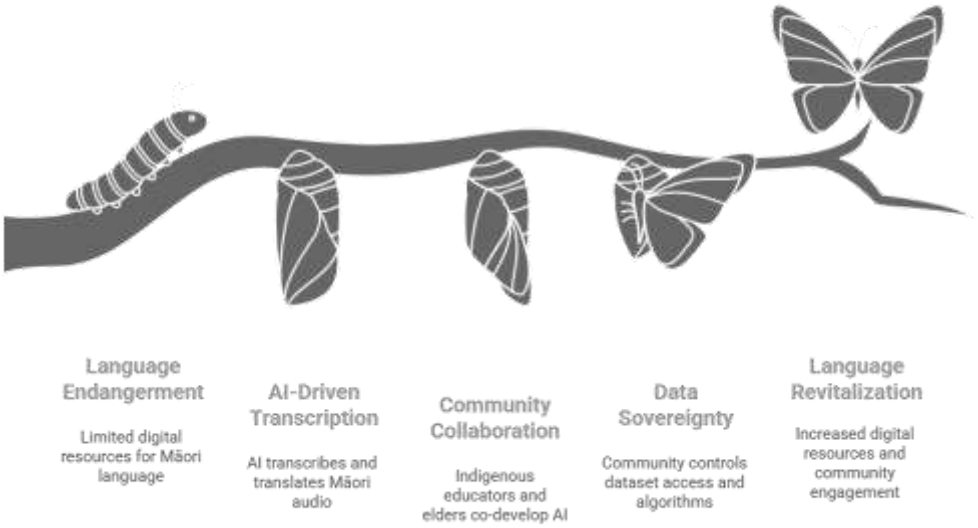

Figure 1. Te Reo Māori Digital Corpus Project

In contrast, the North American FirstVoices AI Initiative demonstrated advanced linguistic modeling but faced resistance due to limited community involvement in data management. Although the project succeeded in generating high-quality neural translation outputs, indigenous stakeholders expressed concerns over loss of contextual integrity and ownership rights. These case comparisons highlight the varying outcomes that arise from differing governance philosophies—underscoring the importance of participatory ethics in digital linguistic resurrection.

The comparative analysis between the two case studies reinforces the broader statistical findings: technological sophistication alone does not guarantee sustainable language revitalization.

The success of the Te Reo Māori initiative demonstrates that when AI development incorporates cultural protocols and ethical data frameworks, it enhances both linguistic resilience and educational relevance. The contrasting experience of FirstVoices reveals that exclusionary or technocratic approaches can erode trust and hinder long-term community engagement.

The explanation further suggests that AI-mediated revitalization is most effective when embedded within educational and community infrastructures. Language learning applications, digital archives, and voice recognition tools achieve deeper cultural impact when co-created through participatory design. The inclusion of indigenous epistemologies within algorithmic logic not only democratizes technology but also transforms it into a tool of cultural restoration rather than extraction.

The overall results indicate that AI-driven language revitalization is not merely a technological process but a socio-cultural and ethical negotiation. Projects that integrate community agency, cultural context, and data sovereignty achieve higher sustainability, legitimacy, and educational impact. The findings challenge the assumption that digital innovation inherently promotes inclusion, revealing that equitable governance structures are crucial for authentic revitalization.

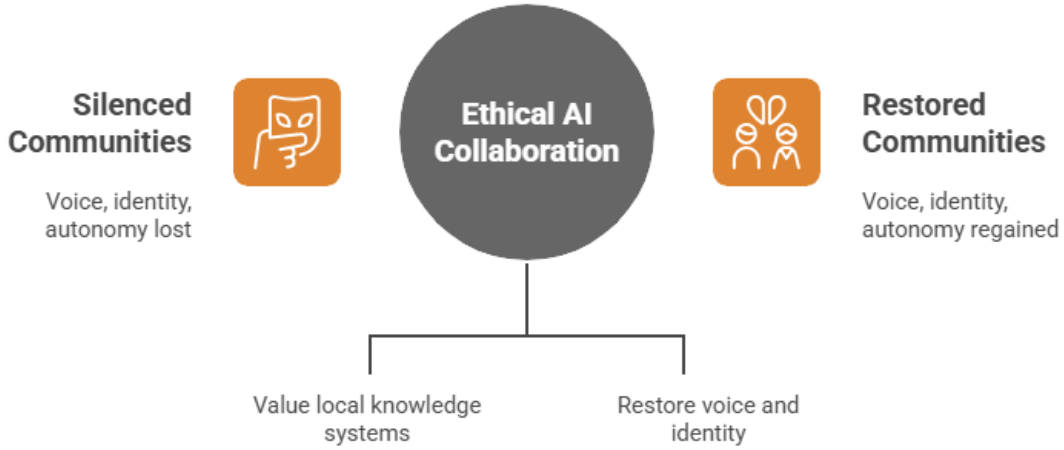

Figure 2. Ethical Co-creation for Digital Resurrection

The interpretation emphasizes that digital resurrection must move beyond the paradigm of efficiency toward a paradigm of ethical co-creation. AI should function as a cultural collaborator that respects local epistemologies and promotes linguistic justice. The future of endangered languages depends not only on algorithmic precision but on reimagining technology as an ally in restoring voice, identity, and autonomy to communities historically silenced by colonial and digital hegemonies.

The research revealed that AI-driven language revitalization projects significantly contribute to the preservation and reconstruction of endangered languages, but their success depends heavily on community participation and ethical data governance. Statistical analyses showed that projects utilizing community-based data ownership models achieved higher sustainability and cultural legitimacy compared to those managed by external institutions (Demuro & Gurney, 2024). The correlation between community agency and linguistic vitality suggests that AI serves as a catalyst for empowerment only when local knowledge systems are integrated into its design and use.

The findings also demonstrated that linguistic data sovereignty is not merely an ethical concern but a determinant of technological efficacy. Projects in Oceania and Southeast Asia exemplified that shared authority over data leads to deeper trust, more authentic outcomes, and

long-term engagement (Al-Samarraie dkk., 2024). The study confirmed that the social architecture surrounding AI is as critical as its computational design. This outcome repositions digital revitalization as both a technical and moral endeavor within the broader framework of educational and cultural continuity.

Previous studies in computational linguistics have largely focused on the technical potential of AI in reconstructing phonological or grammatical structures of endangered languages (e.g., Eshkol-Taravella, 2020; Mager et al., 2022). The present research diverges by emphasizing the ethical-pedagogical dimension how AI reshapes human relationships with language and data. While earlier works treat technology as a neutral tool, this study underscores its agency in mediating power relations and cultural ownership (Peretz-Andersson dkk., 2024). The concept of technological co-sovereignty introduced here extends the discourse on AI ethics into the realm of linguistic justice.

The results also challenge the techno-optimism prevalent in much of the AI-for-heritage literature. Studies like Bird (2021) argued that open data models democratize access to linguistic resources, yet the current findings show that unrestricted data circulation can inadvertently reproduce colonial extractivism. The contrast highlights the need for a nuanced balance between open access and cultural protection. This study, therefore, reframes AI not only as an enabler of preservation but also as a potential agent of epistemic inequality if ethical safeguards are ignored.

The results signify a transformation in the epistemic landscape of language revitalization. The digital resurrection of endangered languages is no longer confined to documentation but represents an act of reclaiming agency over cultural representation (Queiroz dkk., 2024). The intertwining of AI and indigenous knowledge systems reveals that technology, when decolonized, can become a medium of cultural continuity and educational empowerment. The emergence of community-centered digital projects signals a shift from preservation as archiving to preservation as participatory pedagogy.

The findings further indicate that linguistic data sovereignty is emblematic of a broader struggle for digital self-determination. Communities are asserting control not only over their languages but also over the infrastructures that define how those languages are digitally represented (Han dkk., 2024). This reclaiming of authority marks a critical step toward restoring epistemic balance in global digital ecosystems. The study thus identifies AI-mediated revitalization as a microcosm of postcolonial resistance within the Anthropocene's data-driven order.

The implications of this research extend across educational, technological, and ethical domains. For education, the integration of AI-based revitalization models opens possibilities for culturally responsive curricula that bridge ancestral languages with contemporary learning platforms (Geng, 2024). For technology developers, the findings provide an evidence-based framework to design AI tools grounded in community ethics, participatory governance, and linguistic equity. For policymakers, the results emphasize the urgency of establishing legal and institutional frameworks to ensure that linguistic data remain under the custodianship of the communities that generate them.

The study contributes to the growing field of AI humanism, advocating for an inclusive technological paradigm that recognizes cultural diversity as a cornerstone of digital progress (Muawanah dkk., 2024). It implies that revitalizing languages is inseparable from revitalizing the ethics of technological innovation. AI's educational potential can only be fully realized when it operates within a framework of respect, transparency, and co-creation. The research thereby provides a roadmap for aligning digital transformation with cultural justice.

The outcomes arise from the intersection of socio-technical design and historical context. The success of community-led projects reflects the enduring influence of indigenous epistemologies, which prioritize relational knowledge and collective stewardship (Galla dkk., 2023). These frameworks align naturally with participatory AI models, creating synergy between traditional wisdom and digital innovation. Conversely, institutional or corporate-led initiatives often reproduce hierarchies of control that alienate communities from their linguistic heritage, explaining their weaker sustainability metrics.

The findings also result from the inherent duality of AI as both an emancipatory and regulatory force (Siminyu dkk., 2023). Algorithms reflect the intentions of their designers, and without inclusive representation, they tend to replicate dominant linguistic ideologies. The uneven outcomes observed across regions thus stem from differences in governance philosophy rather than technological disparity. AI succeeds in revitalizing languages only when recontextualized within community-led, culturally grounded educational ecosystems.

The results invite a reimagining of how AI and education can collaboratively sustain linguistic diversity (Xie & Zhang, 2024). Future research should move beyond documentation and focus on co-designing ethical AI frameworks that institutionalize linguistic data sovereignty. Such frameworks must incorporate consent protocols, culturally aligned data standards, and community ownership models. The next phase of inquiry should examine how these principles can be operationalized in formal education, digital policy, and teacher training programs for linguistic inclusion.

The study's trajectory opens pathways for interdisciplinary innovation between AI developers, educators, and indigenous scholars. Developing community-led AI curricula could transform endangered language revitalization into a living, evolving practice rather than a preservationist exercise. The "digital resurrection" must thus evolve into a philosophy of digital resilience a future-oriented pedagogy that safeguards not only linguistic survival but also the right to define one's cultural existence in the algorithmic age.

CONCLUSION

The study discovered that the success of AI-driven language revitalization depends less on technological sophistication and more on the ethics of data governance and community participation. This finding departs from earlier research that equated linguistic preservation with technological capability alone. The evidence demonstrates that linguistic data sovereignty—the right of communities to control how their language data are collected, stored, and used—is the determining factor of sustainability and trust in digital revitalization initiatives. The research identified that projects adopting co-governance frameworks, where communities share decision-making authority with developers, produce more culturally authentic and educationally meaningful outcomes. The distinctiveness of this study lies in revealing how AI becomes an instrument of empowerment only when aligned with local epistemologies and indigenous pedagogical practices rather than imposed as a universal technological solution.

The research contributes conceptually by introducing the framework of technological co-sovereignty, a model that redefines the relationship between AI, education, and cultural agency. This concept expands the discourse on digital linguistics by situating AI within a participatory ethics paradigm that integrates indigenous knowledge systems and computational design. The framework bridges the gap between language technology development and decolonial theory, offering a critical lens through which digital innovation can be both inclusive and restorative.

Methodologically, the study advances the integration of digital ethnography and critical discourse analysis to investigate how power, participation, and pedagogy intersect within AI-mediated language projects. This hybrid approach provides a transferable model for future interdisciplinary studies that aim to evaluate not only technical performance but also ethical legitimacy in digital education and cultural sustainability.

The primary limitation of the study lies in its temporal and geographic scope, which focused on five AI-based revitalization projects between 2018 and 2024 across selected regions. The diversity of linguistic ecosystems, socio-political structures, and data governance policies across the globe means that findings may not fully capture the complexity of AI's role in other cultural contexts. Further longitudinal studies are needed to assess how community attitudes toward AI evolve over time and how digital tools influence intergenerational language transmission. Future research should also explore quantitative dimensions of technological co-sovereignty, including metrics for measuring ethical participation and algorithmic transparency. Cross-disciplinary collaborations among educators, technologists, and indigenous scholars are essential to expand the theoretical and practical implications of this work toward developing global frameworks for equitable and sustainable digital language revitalization.

AUTHORS' CONTRIBUTION

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.
Author 3: Data curation; Investigation.

REFERENCES

Adnan, M., Ahmed, I., Iqbal, S., Fazal, M. R., Siddiqi, S. J., & Tariq, M. (2024). Exploring the convergence of Metaverse, Blockchain, Artificial Intelligence, and digital twin for pioneering the digitization in the envision smart grid 3.0. Computers and Electrical Engineering, 120, 109709. https://doi.org/10.1016/j.compeleceng.2024.109709

Al Khafaf, N., Wang, J., Jalili, M., & Sokolowski, P. (2023). A deep learning approach for detecting distributed generation in residential customers. Sustainable Energy, Grids and Networks, 33. Scopus. https://doi.org/10.1016/j.segan.2022.100966

Al-Samarraie, H., Sarsam, S. M., Ibrahim Alzahrani, A., Chatterjee, A., & Swinnerton, B. J. (2024). Gender perceptions of generative AI in higher education. Journal of Applied Research in Higher Education, 17(5), 1944–1958. https://doi.org/10.1108/JARHE-02-2024-0109

Banda, L. O. L., Banda, C. V., Banda, J. T., & Singini, T. (2024). Preserving cultural heritage: A community-centric approach to safeguarding the Khulubvi Traditional Temple Malawi. Heliyon, 10(18), e37610. https://doi.org/10.1016/j.heliyon.2024.e37610

Baptist, A. P., Apter, A. J., Gergen, P. J., & Jones, B. L. (2023). Reducing Health Disparities in Asthma: How Can Progress Be Made. The Journal of Allergy and Clinical Immunology: In Practice, 11(3), 737–745. https://doi.org/10.1016/j.jaip.2022.12.044

Batra, I., Sharma, C., Malik, A., Sharma, S., Kaswan, M. S., & Garza-Reyes, J. A. (2024). Industrial revolution and smart farming: A critical analysis of research components in Industry 4.0. The TQM Journal, 37(6), 1497–1525. https://doi.org/10.1108/TQM-10-2023-0317

Bella, B. R. F., Mayasari, A. D., & Alfina, A. (2024). Etiquette and Protocol in Public Relations Activities. Humanities Research Sustainability, 1(1), 41–48. https://doi.org/10.70177/jhrs.v1i1.1002

Crist, W., & Soemers, D. J. N. J. (2023). The Digital Ludeme Project: Combining archaeological and computational methods for the study of ancient board games. Journal of Archaeological Science: Reports, 49, 104005. https://doi.org/10.1016/j.jasrep.2023.104005

Demuro, E., & Gurney, L. (2024). Artificial intelligence and the ethnographic encounter: Transhuman language ontologies, or what it means "to write like a human, think like a machine." Language & Communication, 96, 1–12. https://doi.org/10.1016/j.langcom.2024.02.002

Farros, A. A., Wijayanti, I., & Agustiar, A. (2024). Meaning Recognition Terms in Arabic Semantic Studies. Humanities Research Sustainability, 1(1), 23–33. https://doi.org/10.70177/jhrs.v1i1.907

Franco, I., LeCompte, M., Beltrán Ponce, S., Domogauer, J., Dahir, A., & Suneja, G. (2023). Chapter 99—Diversity, equity, and inclusion. Dalam A. E. M. Eltorai, J. A. Bakal, D. W. Kim, & D. E. Wazer (Ed.), Translational Radiation Oncology (hlm. 603–611). Academic Press. https://doi.org/10.1016/B978-0-323-88423-5.00008-X

Galla, C. K., Holmes, A., Schwab Cartas, J., & Rowluck, W. (2023). Community-centered Indigenous language recovery, restoration, revitalization, and renewal. Dalam R. J. Tierney, F. Rizvi, & K. Ercikan (Ed.), International Encyclopedia of Education (Fourth Edition) (hlm. 786–796). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.07038-X

Gao, J., Peng, P., Lu, F., Claramunt, C., Qiu, P., & Xu, Y. (2024). Mining tourist preferences and decision support via tourism-oriented knowledge graph. Information Processing & Management, 61(1), 103523. https://doi.org/10.1016/j.ipm.2023.103523

Geng, X. (2024). Enhancing Translation Education Through Augmented Reality (AR): International Journal of Web-Based Learning and Teaching Technologies, 19(1). https://doi.org/10.4018/IJWLTT.361651

Gerwin, J. (2024). 'Learn Jafaikan in two minutes' – Multicultural London English, enregisterment and ideology in English newspapers. Language & Communication, 97, 13–33. https://doi.org/10.1016/j.langcom.2024.04.002

Han, S., Liu, Z., Deng, Z., Gupta, S., & Mikalef, P. (2024). Exploring the effect of digital CSR communication on firm performance: A deep learning approach. Decision Support Systems, 176, 114047. https://doi.org/10.1016/j.dss.2023.114047

Kabra, G., Dhaundiyal, M., & Saharan, A. (2023). Prioritization of barriers to social media usage by non-governmental organizations (NGOs) in disaster preparedness phase: An integrated methodology. International Journal of Disaster Risk Reduction, 99, 104099. https://doi.org/10.1016/j.ijdrr.2023.104099

Lee-Geiller, S. (2024). Technology married to good governance and diversity: Explaining e-participation preparedness in government. Technological Forecasting and Social Change, 201, 123218. https://doi.org/10.1016/j.techfore.2024.123218

Li, N., Borah, D., Kim, J., & Ji, J. (2023). Internationalization of transnational entrepreneurial firms from an advanced to emerging economy: The role of transnational mixed-embeddedness. International Journal of Entrepreneurial Behavior & Research, 29(3), 707–737. https://doi.org/10.1108/IJEBR-07-2021-0527

Mahmudi, mahmudi, & Khoiruddin, M. A. (2024). Improving Student Life Skills Through the Ma Plus Skills Program at MAN 4 Kediri. Humanities Research Sustainability, 1(1), 16–22. https://doi.org/10.70177/jhrs.v1i1.995

Muawanah, U., Marini, A., & Sarifah, I. (2024). The interconnection between digital literacy, artificial intelligence, and the use of E-learning applications in enhancing the sustainability of Regional Languages: Evidence from Indonesia. Social Sciences & Humanities Open, 10, 101169. https://doi.org/10.1016/j.ssaho.2024.101169

Mukherjee, A. A., Raj, A., & Aggarwal, S. (2023). Identification of barriers and their mitigation strategies for industry 5.0 implementation in emerging economies. International Journal of Production Economics, 257, 108770. https://doi.org/10.1016/j.ijpe.2023.108770

Muqit, F. A., & Putra, A. W. (2024). Delving Into Existential Fatigue and The Search for Meaning in The Play "kapai-kapai" by Arifin C. Noer. Humanities Research Sustainability, 1(1), 34–40. https://doi.org/10.70177/jhrs.v1i1.929

O'Rourke, B., Zhao, Q., Dayán-Fernández, A., Dickson, E., Wang, R., Wang, M., Zhang, L., & Imran, M. (2024). Integrated geospatial methods for multilingual cities: Combining GPS, electronically activated recorder, and map-based ethnographic interviews. Research Methods in Applied Linguistics, 3(3), 100147. https://doi.org/10.1016/j.rmal.2024.100147

Peretz-Andersson, E., Tabares, S., Mikalef, P., & Parida, V. (2024). Artificial intelligence implementation in manufacturing SMEs: A resource orchestration approach. International Journal of Information Management, 77, 102781. https://doi.org/10.1016/j.ijinfomgt.2024.102781

Pothuri, B., Blank, S. V., Myers, T. K., Hines, J. F., Randall, L. M., O'Cearbhaill, R. E., Slomovitz, B. M., Eskander, R. N., Alvarez Secord, A., Coleman, R. L., Walker, J. L., Monk, B. J., Moore, K. N., O'Malley, D. M., Copeland, L. J., & Herzog, T. J. (2023). Inclusion, diversity, equity, and access (IDEA) in gynecologic cancer clinical trials: A joint statement from GOG foundation and Society of Gynecologic Oncology (SGO). Gynecologic Oncology, 174, 278–287. https://doi.org/10.1016/j.ygyno.2023.05.006

Queiroz, L. P., Nogueira, I. B. R., & Ribeiro, A. M. (2024). Flavor Engineering: A comprehensive review of biological foundations, AI integration, industrial development, and socio-cultural dynamics. Food Research International, 196, 115100. https://doi.org/10.1016/j.foodres.2024.115100

Salam, N. E., Yasir, Y., Firdaus, M., Azwa, R., & Lubis, E. E. (2024). Tourism Communication Model in Developing Heritage City Tourism in Siak Sri Indrapura Regency, Riau, Indonesia. Humanities Research Sustainability, 1(1), 1–15. https://doi.org/10.70177/jhrs.v1i1.876

Saka, A., Taiwo, R., Saka, N., Salami, B. A., Ajayi, S., Akande, K., & Kazemi, H. (2024). GPT models in construction industry: Opportunities, limitations, and a use case validation. Developments in the Built Environment, 17, 100300. https://doi.org/10.1016/j.dibe.2023.100300

Sharma, G. D., Taheri, B., Cichon, D., Parihar, J. S., & Kharbanda, A. (2024). Using innovation and entrepreneurship for creating edge in service firms: A review research of tourism and hospitality industry. Journal of Innovation & Knowledge, 9(4), 100572. https://doi.org/10.1016/j.jik.2024.100572

Silva, L. C. e, Sobrinho, Á., Cordeiro, T., Silva, A. P. da, Dermeval, D., Marques, L. B., Bittencourt, I. I., Júnior, J. J. dos S., Melo, R. F., Portela, C. dos S., Souza, M. R. de A., Pereira, R. L., Yasojima, E. K. K., & Isotani, S. (2024). Assessing students' handwritten text

productions: A two-decades literature review. Expert Systems with Applications, 250, 123780. https://doi.org/10.1016/j.eswa.2024.123780

Siminyu, K., Abbott, J., Túbọ̀sún, K., Anuoluwapo, A., Sibanda, B. K., Yeboah, K., Adelani, D., Mokgesi-Selinga, M., Apina, F. R., Mthembu, A. T., Ramkilowan, A., & Oladimeji, B. (2023). Consultative engagement of stakeholders toward a roadmap for African language technologies. Patterns, 4(8), 100820. https://doi.org/10.1016/j.patter.2023.100820

Spoladore, D., Tosi, M., & Lorenzini, E. C. (2024). Ontology-based decision support systems for diabetes nutrition therapy: A systematic literature review. Artificial Intelligence in Medicine, 151, 102859. https://doi.org/10.1016/j.artmed.2024.102859

Tsai, P.-H., Kao, Y.-L., & Kuo, S.-Y. (2023). Exploring the critical factors influencing the outlying island talent recruitment and selection evaluation model: Empirical evidence from Penghu, Taiwan. Evaluation and Program Planning, 99, 102320. https://doi.org/10.1016/j.evalprogplan.2023.102320

Uztürk, D., & Büyüközkan, G. (2024). Industry 4.0 technologies in Smart Agriculture: A review and a Technology Assessment Model proposition. Technological Forecasting and Social Change, 208, 123640. https://doi.org/10.1016/j.techfore.2024.123640

Wang, K. (2024). Precision Teaching of College English Based on Learning Diagnostic Systems. International Journal of Web-Based Learning and Teaching Technologies, 19(1). https://doi.org/10.4018/IJWLTT.349985

Wider, W., Gao, Y., Chan, C. K., Lin, J., Li, J., Tanucan, J. C. M., & Fauzi, M. A. (2023). Unveiling trends in digital tourism research: A bibliometric analysis of co-citation and co-word analysis. Environmental and Sustainability Indicators, 20, 100308. https://doi.org/10.1016/j.indic.2023.100308

Wong, K. L., Kapaanaokalaokeola Nakoa Oliveira, K.-A. R., & Noʻeau Warner, S. L. (2023). Ke Ala o ka Noʻeau: An infusion of Hawaiian into the Hawaiian language. Dalam R. J. Tierney, F. Rizvi, & K. Ercikan (Ed.), International Encyclopedia of Education (Fourth Edition) (hlm. 42–49). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.06079-6

Xie, M., & Zhang, C. (2024). Can linguistic big data empower digital economy?: Evidence from China. Economic Analysis and Policy, 84, 1771–1787. https://doi.org/10.1016/j.eap.2024.11.007

Zhang, L. (2023). The Ethical Turn of Emerging Design Practices. She Ji: The Journal of Design, Economics, and Innovation, 9(3), 311–329. https://doi.org/10.1016/j.sheji.2023.09.002