JOURNAL OF INFORMATICS, EDUCATION AND MANAGEMENT (JIEM)
Vol 8 No 1 (2026): September 2025 - February 2026, pp. 1170 ~1185
ISSN: 2716-0696, DOI: 10.61992/jiem.v8i1.306

PRESERVING PROVERBS IN THE DIGITAL AGE: A QUALITATIVE STUDY ON STAKEHOLDER PERCEPTIONS OF ARTIFICIAL INTELLIGENCE FOR ACEHNESE LANGUAGE HERITAGE

Ibnu Hajar[1*], Asmaul Husna[1], Muhammad Kahfi Aulia[1], T. Muntazar[1]

[1]Universitas Bumi Persada, Lhokseumawe, Indonesia

Article Info

Article history:
Received February 5 , 2026
Revised February 20 , 2026
Accepted February 25 , 2026

Keywords:
Acehnese Proverbs, Artificial Intelligence, Cultural Preservation, Endangered Languages, Ethical AI

ABSTRACT

This qualitative study investigates the perceived opportunities and challenges of developing an artificial intelligence application for preserving Acehnese proverbs (Hadih Maja), a component of cultural heritage. Through reflexive thematic analysis of in-depth interviews with academics and cultural custodians in North Aceh and Lhokseumawe, Indonesia, the research reveals a dualistic landscape. Stakeholders recognize AI's transformative potential as a bridge for intergenerational engagement, enabling interactive, mobile-based learning for youth. However, this promise is counterbalanced by formidable socio-technical hurdles, including the fundamental "low-resource" paradox where data scarcity severely limits natural language processing capabilities, and profound ethical risks of cultural decontextualization and erosion of community sovereignty over knowledge. The findings underscore that primary impediments are deeply socio-technical, necessitating an approach that makes community governance foundational. The study concludes that successful initiatives must be reconceptualized from "AI development projects" to "community-led cultural revitalization projects enabled by AI," prioritizing participatory co-design and ethical data governance. This research offers a critical framework for ethically leveraging AI in endangered language contexts globally.

Corresponding Author:
Ibnu Hajar1*| Universitas Bumi Persada, Lhokseumawe, Indonesia
Email: ibnuhajar116@gmail.com

1. INTRODUCTION

The relentless global tide of language endangerment, accelerated by globalization, urbanization, and the dominance of major world languages, represents an irretrievable loss of human intellectual and cultural diversity. The Acehnese language, spoken in the northernmost part of Sumatra, Indonesia, stands as a poignant example within this crisis. While still spoken by millions, it is weakening in intergenerational transmission, particularly in urban centers like Lhokseumawe. Negative attitudes towards the Acehnese language have increased among Acehnese youth, who now prefer using Indonesian in their daily lives for various reasons (Al-Auwal, 2017). Indonesian has evolved from a colonial language to a

vital component of national identity, though during the conflict, there was a gradual shift away from Acehnese (Al-Auwal et al., 2024).

This shift precipitates the erosion of deeply embedded cultural knowledge systems, most critically encapsulated in *peribahasa* (proverbs). These proverbs are not merely linguistic artefacts; they are dense repositories of Acehnese philosophy, historical memory, social norms, and ecological understanding. Many Acehnese proverbs teach cultural values such as hard work, cooperation, honesty, and responsibility (Ibrahim & Usman, 2021), and analyzing popular proverbs is key to understanding a culture (Rousan et al., 2025). The potential disappearance of proverbs indicates a break in the community's knowledge and moral structure, representing a loss that affects cultural identity beyond just language.

Concurrently, the rapid proliferation of artificial intelligence (AI) and digital mobile technologies offers unprecedented, yet double-edged, tools for cultural preservation. On one hand, digital revitalization has evolved from static archives to dynamic, interactive platforms. As Julia and Jeyanthi (2025) highlights the potential of Mobile Assisted Language Learning (MALL) to integrate traditional language teaching with modern technological capabilities, promoting its use in education. The prospect of leveraging AI—through interactive chatbots, personalized learning applications, and authentic text-to-speech systems—presents a tantalizing opportunity to make Acehnese proverbs accessible, engaging, and relevant for a digital-native generation. The use of artificial intelligence can help preserve the native languages of marginalized and ethnic groups while facilitating the dissemination of accurate information in those languages (Horna-Saldaña et al., 2025).

However, the application of advanced AI to a low-resource, indigenous context like Acehnese is fraught with a fundamental socio-technical paradox. The very technologies that promise salvation are predicated on vast, structured, annotated datasets—the "big data" that endangered languages, by definition, lack. This "low-resource" reality creates severe technical bottlenecks for Natural Language Processing (NLP) and speech synthesis, challenges that Gangwal et al. (2024) identify as foundational to AI infrastructure success.

More critically, the risks are not merely technical but profoundly ethical and cultural. The drive to digitize can lead to decontextualization, where rich, performative proverbs are stripped of their cultural gravity and reduced to deconstructed data points or trivialized game elements (Castro et al., 2025). Artificial intelligence can improve metadata work for library professionals, but it also intensifies risks in a late capitalist context where efficiency and profit may compromise the quality of metadata creation (Freeman, 2025). AI systems, often designed with Western, urban biases, may fail to capture the emotional cadence, contextual nuance, and sacred boundaries inherent in traditional knowledge systems.

This tension defines the research gap. While studies on digital language revitalization and the technical constraints of low-resource NLP are growing, there is a scarcity of nuanced, on-the-ground analysis that anticipates these socio-technical complexities before a major AI development project commences. Most research evaluates tools post-development. There is a pressing need for proactive, stakeholder-driven exploration that maps the terrain of opportunities and challenges from the dual perspectives of those who hold the knowledge

(cultural custodians) and those who would build the technology (academics in relevant fields).

This study directly addresses this gap. Its novelty lies in its anticipatory and integrative approach, conducting a qualitative, multi-stakeholder feasibility study that brings ethicists (anthropologists), cultural custodians (*tokoh adat*), linguists, and technologists (AI and data scientists) into a single analytical frame at the conceptual stage. It positions the proposed AI application not as a pre-determined solution, but as a proposition whose viability, desirability, and ethical contours must first be investigated from these multiple, often divergent, standpoints.

The potential benefits of this research are therefore significant and two-tiered. Academically, it will generate a rich, contextual case study contributing to the emerging literature on ethical AI for indigenous knowledge preservation, offering a model for other communities facing similar dilemmas. Practically and socially, its primary benefit is risk mitigation and pathway clarification for the envisioned large-scale project. By foregrounding the perceptions of cultural custodians, it ensures that community sovereignty over knowledge, as mandated by the CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles for indigenous data governance(Carroll et al., 2021), is centered from the outset.

By elucidating the practical concerns of technologists, it provides a realistic assessment of technical feasibility. Ultimately, this research aims to produce a stakeholder-validated framework that can guide the ethical, culturally-grounded, and technically sound design of future AI-driven revitalization tools, ensuring they truly serve as bridges for cultural continuity rather than becoming instruments of unintended harm or yet another unsustainable digital relic.

2. LITERATURE REVIEW

The research on the opportunities and challenges of an AI-driven project for Acehnese proverb preservation sits at the confluence of three critical scholarly domains: the application of digital technology for endangered language revitalization, the specific affordances and limitations of artificial intelligence in low-resource linguistic contexts, and the imperative for ethical, community-centered frameworks in indigenous data governance. This review synthesizes recent literature from these areas to establish the theoretical and practical foundation for the study, highlighting both the transformative potential and the significant socio-technical complexities inherent in such an endeavor.

Digital Revitalization and the Promise of AI for Endangered Languages

The use of digital tools for language preservation and revitalization has evolved from basic archival databases to interactive, engagement-focused platforms. Contemporary scholarship emphasizes moving beyond static "digital museums" towards creating dynamic, participatory spaces that foster everyday language use, particularly among youth. Tang et al. (2025),for instance, conducted a university project on mobile learning for less-commonly taught languages in China and reported that learners had increased interest, improved spoken communication skills, and greater autonomy. Jia et al. (2025) also developed a mobile-

assisted language learning (MALL) speaking model and assessed its effectiveness, revealing that participants viewed the teaching approach positively.

Contemporary scholarship demonstrates the potential of AI in supporting language learning and preservation. They also align with the core objective of the proposed Acehnese project, which aims to leverage a mobile application and chatbot for interactive learning. Furthermore, the creation of a multimodal corpus—integrating audio, text, and narrative context—as envisioned in the project, is supported by Grenoble (2010), who asserts that language documentation addresses the urgent requirement to collect, describe, and archive information on the growing number of endangered languages. Such documenting preserves cultural traditions for future generations and offers valuable resources for researchers and scholars to study and understand culture.

The "Low-Resource" Paradox and Technical-Social Challenges

Despite this promise, the application of AI to endangered languages like Acehnese confronts the fundamental "low-resource" paradox. State-of-the-art NLP and speech technologies typically require massive, annotated datasets, which are precisely what endangered languages lack. As Abdalla et al. (2025) assert that data scarcity is a significant challenge for Artificial Intelligence (AI), posing questions about future solutions to this issue. This scarcity affects every stage, from building accurate speech recognition models to training coherent language generators, often necessitating innovative but imperfect techniques like cross-lingual transfer learning.

Gangwal et al. (2024) further elucidate that the success of AI infrastructure is not solely a function of algorithmic sophistication but is critically dependent on the quality, quantity, and structure of the underlying data. For Acehnese proverbs, challenges such as dialectal variation, non-standardized orthography, and the absence of machine-readable texts compound this data scarcity, creating significant technical hurdles for tasks like proverb recognition and authentic Text-to-Speech (TTS) synthesis.

However, the literature reveals that the challenges are not merely technical; they are profoundly socio-technical. A growing body of work critiques the potential for technological solutions to cause harm through decontextualization and commodification inadvertently. Castro et al. (2025) caution that AI tools are predominantly designed for mainstream, urban educational contexts and may carry inherent biases that render them ineffective or even damaging when applied uncritically to diverse cultural environments. This risk is particularly acute with proverbs and oral traditions, where meaning is deeply embedded in specific cultural, historical, and performative contexts.

The concern that digitization could trivialize sacred or complex knowledge into mere "entertainment" is a valid ethical consideration in the digital era. Moreover, user perception studies indicate public ambivalence towards AI; Gerlich, (2023) finds that trust and perceived vulnerability significantly influence whether AI is seen as a blessing or a curse, while Jiang et al., (2024) note user concerns about AI's lack of emotional depth and cultural nuance—key components in the transmission of proverbial wisdom.

Toward Ethical Governance and Community-Led Frameworks

In response to these risks, recent literature has strongly advocated for ethical frameworks that center on indigenous authority and community benefit. The most prominent of these is the integration of the FAIR (Findable, Accessible, Interoperable, Reusable) principles with the CARE principles for indigenous data governance (Carroll et al., 2021). The CARE principles explicitly assert the rights of indigenous communities to govern data about them, shifting the focus from open data to equitable data. This framework directly informs the necessity of ongoing community custodianship, as demanded by the cultural custodians in the study. It argues against extractive research models and for partnerships built on respect, reciprocity, and shared authority.

This aligns with operational research on community-based methods. Satyawati et al., (2025) demonstrates that data collection led by local leaders enhances ecological validity and data quality, ensuring cultural protocols are respected. Similarly, the concept of AI as a complementary tool rather than a replacement is emphasized. Barnes et al., (2024) show that cultural identity significantly mediates engagement with technology, suggesting that AI systems for revitalization must be co-designed to align with local cultural values and pedagogical traditions. This supports a model of human-AI collaboration, where technology supports, rather than supplants, the role of knowledge keepers and educators (Nguyen et al., 2024).

Finally, the literature stresses the often-overlooked issue of long-term sustainability. Edmond and Morselli (2020) argue that digital humanities projects must plan for sustainability across multiple dimensions—technical, financial, social, and epistemic—from their inception, viewing them as evolving socio-technical systems rather than one-time deliverables. This is crucial to avoid the creation of "digital relics" that become inaccessible after initial funding ends.

3. RESEARCH METHOD

Research Design and Philosophical Underpinning

This preliminary study aims to explore and delineate the opportunities and challenges inherent in a proposed research project aimed at preserving and revitalizing the Acehnese language and wisdom through interactive AI technology. This study adopts a qualitative descriptive research design. It is particularly suited for obtaining straightforward, largely unadorned descriptions of phenomena, especially when the research aims to stay close to the data and provide a comprehensive summary of an event or experience in everyday terms (Creswell & Poth, 2018). This design is appropriate as the primary objective is not to build a grand theory but to produce a detailed, contextual analysis of the perceived opportunities and challenges from the perspectives of directly involved stakeholders.

The study is underpinned by a constructivist paradigm, which posits that reality is socially constructed and that understanding is co-created between the researcher and the participant (Creswell & Poth, 2018). This aligns with the research goal of understanding multiple,

subjective realities held by academics and cultural bearers regarding the intersection of advanced technology and Indigenous language preservation.

Research Locale and Participant Selection

The research is conducted in the regions of North Aceh and the city of Lhokseumawe in Aceh Province, Indonesia. These locales are selected due to their significance as cultural and educational hubs where the Acehnese language is actively used but also faces pressures from globalization and language shift. The sampling strategy is purposive, specifically criterion sampling, to select information-rich cases that can offer profound insights into the research question (Patton, 2015).

The participants comprise eleven individuals divided into two distinct stakeholder groups to ensure triangulation of perspectives. The first group is academicians: two linguists (with expertise in Acehnese philology and sociolinguistics), two anthropologists (focusing on Acehnese culture and oral traditions), two data scientists (skilled in handling unstructured cultural data), and two AI specialists (with knowledge in natural language processing and chatbot development). The second group consists of three local cultural custodians recognized by their communities as holders of traditional knowledge, including peribahasa (proverbs), folklore, and linguistic nuances.

Data Collection Procedure

The primary data collection method is semi-structured, in-depth interviews. This method is chosen for its flexibility, allowing for guided exploration of pre-determined topics while permitting spontaneous probing to follow emerging themes (Brinkmann & Kvale, 2018). Two tailored interview protocols are developed: one for academic respondents and another for cultural custodians, though both cover overlapping core themes.

The protocol for academics focuses on technical and epistemological opportunities (e.g., "What are the potential AI architectures suitable for low-resource languages?"; "How can anthropological theory inform the ethical collection of folklore?") and challenges (e.g., data scarcity, model bias, interdisciplinary collaboration). The protocol for cultural custodians centers on cultural authenticity, ethical engagement, and perceived value (e.g., "What is your view on recording and digitizing sacred or secret proverbs?"; "How can an application ensure the spirit (*makna*) of a proverb is not lost?", "What are your concerns about this technology?").

Each interview, expected to last 30-40 minutes, is conducted in a location of the participant's choice—often university offices for academics and community halls or homes for cultural custodians—to ensure comfort and contextual relevance. With permission, all interviews are audio-recorded using a digital recorder. For interviews conducted in Acehnese or Bahasa Indonesia, meticulous notes are taken simultaneously to capture non-verbal cues and contextual details.

The researcher adopts a stance of a "learner" and "listener," crucial for building rapport, especially with cultural custodians, as emphasized in decolonial and community-based

research methodologies (Smith, 2024). The data collection process is iterative; preliminary insights from early interviews inform slight refinements to questioning in subsequent ones, a hallmark of reflexive qualitative practice.

Data Analysis Technique

The analysis follows the six-phase framework of reflexive thematic analysis as detailed by Braun and Clarke, (2006) which is valued for its theoretical flexibility and systematic rigor. First, all audio recordings are transcribed verbatim, and transcripts in local languages are translated into English by a bilingual researcher with cultural competency, with a subset back-translated to check for conceptual accuracy.

The researcher then immerses themselves in the data by repeatedly reading the transcripts while listening to the recordings to capture tone and emphasis. Initial codes are generated systematically across the entire dataset. For instance, a segment where a linguist discusses "the lack of standardized orthography for certain proverbs" might be coded as "Challenge-Data-Standardization," while a cultural custodian's excitement about "youth accessing stories through phones" might be coded as "Opportunity-Engagement-Accessibility."

These codes are then collated into potential themes. This involves sorting codes into broader patterns of meaning. For example, codes like "Challenge-Data-Standardization," "Challenge-Audio-Quality," and "Challenge-Metadata-Absence" may cluster under a candidate theme "Infrastructural and Data Preparedness Hurdles." Similarly, codes like "Opportunity-Pedagogical-Innovation," "Opportunity-Cross-generational-Dialogue," and "Opportunity-Cultural-Continuity" may form a theme, "Revitalization and Educational Potential."

Themes are reviewed and refined at two levels: first, in relation to the coded extracts to ensure they form a coherent pattern, and second, in relation to the entire dataset to ensure they accurately represent the totality of the data. This refinement process may lead to splitting, combining, or discarding themes. The final stage involves defining and naming themes, weaving them into a coherent narrative that directly addresses the research aim. To ensure trustworthiness, member checking is employed, where summarized findings are shared with a subset of participants for verification, and a detailed audit trail of analytical decisions is maintained (Nowell et al., 2017).

Ethical Considerations

The study adheres to strict ethical protocols. Before interviews, all participants receive and sign an informed consent form explaining the study's purpose, procedures, risks, benefits, confidentiality measures, and their right to withdraw. Particular sensitivity is applied to interactions with cultural custodians, recognizing that traditional knowledge is a form of intellectual property. The consent process for them explicitly details how their knowledge will be used in this study and potentially in the future main project, emphasizing their ongoing agency (Smith, 2024). Anonymity and confidentiality are assured; all participants are assigned pseudonyms, and identifying details are removed from transcripts. Data is stored on encrypted, password-protected devices.

4. FINDINGS AND DISCUSSION

This qualitative study, employing a constructivist, descriptive design, reveals a complex and multifaceted landscape of perceptions held by key stakeholders regarding preserving Acehnese proverbs through interactive AI. The thematic analysis of interviews with academics and cultural custodians in North Aceh and Lhokseumawe crystallized into four overarching thematic clusters: transformative potential for engagement and continuity, foundational data and technical hurdles, imperatives for ethical and cultural integrity, and navigating interdisciplinary and implementation realities. These themes are not siloed but deeply interconnected, painting a picture of a promising yet profoundly challenging endeavor that demands careful, community-led navigation.

Transformative Potential for Engagement and Cultural Continuity

The most salient and unanimous opportunity identified across both stakeholder groups was the project's potential to serve as a dynamic bridge between Acehnese cultural heritage and contemporary youth. Cultural custodians expressed cautious optimism, with one elder noting, "Our stories sleep in old notebooks, and our voices fade. If a phone can make the young hear these proverbs in our accent and they understand the lessons, then the technology is a welcome messenger."

This sentiment aligns with global research on digital language revitalization, where online media and virtual communities have effectively brought the revitalization of language to the forefront of public discussion (Chen & Chen, 2025). The mobile application illustrates the role of community-driven technological interventions in aiding indigenous language revitalization, emphasizing the importance of cultural integrity, intellectual property rights, and meeting practical community needs (Abingosa et al., 2025).

Furthermore, anthropologists and linguists emphasized the project's potential for creating a "multimodal cultural corpus." They argued that by systematically pairing audio, narrative context, and explanatory text, the project could preserve not just lexical items but the "performative and pragmatic essence" of the proverbs. One linguist stated, "A proverb is not a dictionary entry. Its meaning is activated in context, in tone, in the story it alludes to. An AI that can deliver that package is preserving discourse, not just vocabulary."

The statement emphasizes the significant potential of NLP as highlighted by an AI specialist who noted: "NLP can process linguistic data in both a structured and contextually way, which enhances the intuitiveness and naturalness of user interactions with technology." Combining language documentation and language technology presents a significant opportunity for both fields, especially in their overlapping areas, where substantial potential exists for creating mutually beneficial resources (Blokland et al., 2015). Consequently, the younger generation can stay linked to their ancestral heritage while engaging in the modern digital world.

Foundational Data and Technical Hurdles: The "Low-Resource" Reality

Despite the enthusiasm, a stark counter-theme emerged around the significant infrastructural and technical challenges, centering on the "low-resource" nature of Acehnese in the AI paradigm. Data scientists and AI specialists were unequivocal in identifying data scarcity as the primary bottleneck. As one data scientist explained, "For a large language model or even a robust chatbot, we need massive, clean, annotated datasets—millions of words. For Acehnese proverbs, we might start with a few hundred. This is the fundamental asymmetry."

This excerpt highlights the significant challenges in preserving Acehnese cultural heritage, specifically regarding proverbs and oral traditions. A linguist asserted: “The absence of recorded materials—texts, videos, or audio of Acehnese proverbs—underscores the need for ongoing efforts to enrich this cultural heritage in diverse media forms.” An anthropologist further notes: “While Aceh possesses a rich array of oral traditions, they largely exist in an undocumented format, limiting their accessibility.”

The concerns are echoed by a cultural custodian who pinpoint: “The lack of data on Acehnese proverbs is due to the unwillingness of the younger generation to interact with the Acehnese language and the prevailing negative attitude towards it. As a resolution, it needs a systematic initiative to transform these ideologies and attitudes, thereby promoting the revitalization and wider use of the Acehnese language within society.”

The challenge extends beyond volume to quality and structure. Linguists pointed to the lack of a universally standardized orthography for certain dialectal variations and the absence of digitized, machine-readable texts. Ishkhanyan (2025) also found a similar problem in which challenges persist in scalability, dataset availability, and in balancing ethical considerations between privacy protections and AI performance. This scarcity creates a major hurdle for NLP tasks. As Gangwal et al. (2024) confirm, the quality and quantity of input data are crucial components of AI infrastructure, significantly impacting the success of AI-driven efforts, alongside other factors.

The technical challenges are multifaceted. Developing a TTS system with "authentic accent" was cited as a particularly ambitious goal. An AI specialist noted, "Generic TTS models fail to capture the phonological nuances and emotional cadence crucial for proverbs. Building one requires hours of high-fidelity, speaker-balanced audio recordings—a resource we currently don't have." Furthermore, the chatbot’s ability to "recognize" user-inputted proverbs faces the problem of orthographic variation and potential misspellings, necessitating the development of fuzzy matching algorithms and a comprehensive synonym/ variant database—a non-trivial linguistic task.

Imperatives for Ethical and Cultural Integrity: Beyond Technical Solutions

Perhaps the most profound findings emerged from the domain of ethics and cultural integrity, themes championed primarily by the anthropologists and cultural custodians. Here, the project transitions from a technical challenge to a socio-cultural intervention. A major concern was the risk of decontextualization and trivialization. A cultural custodian warned, "A peribahasa like 'adat bak Poe Teumeureuhom, hukom bak Syiah Kuala' [custom lies with the monarch, law with Syiah Kuala] carries the weight of our history and social structure. Turning it into a quiz question in a game risks stripping it of its gravity and reducing our wisdom to entertainment."

This excerpt emphasizes the importance of ethical considerations and cultural integrity in the development and application of AI systems within indigenous communities. An AI specialists asserted: “AI systems in Indigenous communities should not only prioritize technical optimization but must also emphasize ethical accountability, cultural integrity, and community governance.” Another AI specialists highlights: “Values such as fairness, accountability, and transparency emerge from the interplay between AI agents and social institutions, rather than being solely encoded in algorithms.”

In addition, a data scientist also expressed similar concerns and provided examples of several challenges that need to be addressed. He noted: “Critical concerns exist regarding informed consent, the ownership of cultural data, the pursuit of epistemic justice, and the risk of exploitative technological practices that could perpetuate colonial power imbalances.” Another data scientist also confirmed: “Cultural AI emphasizes the necessity for contextual design, institutional alignment, and governance mechanisms that recognize AI as a component of a dynamic sociotechnical ecosystem, rather than viewing it merely as a standalone technological solution.”

The indiscriminate use of digital tools has the potential to commercialize and oversimplify complex knowledge systems. AI enhances accessibility and visibility for languages, but may not preserve their unique nuances, idiomatic expressions, and cultural contexts (Ardalan et al., 2025). AI recommendations sometimes even contradict Indonesian language norms or specific assignment requirements (Muntazar & Hajar, 2025). Besides, digitization has transformed traditional notions of sacred spaces and influenced the evolution of religious practices in the digital age (Battista, 2024).

Faced with these challenges, one anthropologist suggested: “To establish an effective approach, it is essential to integrate participatory design, implement community-led governance models, and ensure alignment with data sovereignty principles alongside culturally responsive AI development.” A linguist also confirms: “Ethical imperatives are essential prerequisites that not only supplement the language used but also establish the legitimacy, sustainability, and social acceptance of AI systems.” Another linguist asserts: “Preserving cultural integrity in AI development requires the integration of relational accountability, contextual sensitivity, and a commitment to long-term community benefits at every stage of the AI lifecycle. This approach transcends basic technical solutions, emphasizing the importance of cultural considerations in the development process.”

Consequently, the theme of community agency and custodianship was paramount. All three cultural custodians stressed that the project must not be an "extractive" exercise. They insisted on ongoing governance roles, such as vetting which proverbs are suitable for public digital dissemination (noting some may be sacred or context-specific) and reviewing the narrative explanations for cultural accuracy. The findings confirm the importance of community-driven data governance in AI ethics and minority language preservation, while also extending existing models through adaptive AI methodologies, interdisciplinary collaboration, and fairness-aware dialectal modeling (Ishkhanyan, 2025). This demand aligns perfectly with the FAIR and CARE principles for indigenous data governance, which emphasize that data must not only be technologically sound but also subject to indigenous authority (Carroll et al., 2021).

In data collection, for instance, this activity needs to be led by local leaders to improve ecological validity and data quality (Satyawati et al., 2025). The anthropologists also argued for embedding FAIR and CARE principles into the project's design, suggesting mechanisms for ongoing consent and benefit-sharing, moving beyond a one-time ethical approval. This strategy aligns with the emphasis on human oversight to maintain academic integrity (Granjeiro et al., 2025) and the proposal for human-AI collaboration in higher education (Nguyen et al., 2024). With careful implementation, AI can serve as a pedagogical partners, supporting effective, creative, and reflective tools for language development and preservation.

Navigating Interdisciplinary and Implementation Realities

The final theme captures the practical and collaborative challenges of executing such a multifaceted project. Interdisciplinary translation emerged as a subtle but significant barrier. As one anthropologist described, "We speak of 'context' and 'meaning,' while our AI colleagues speak of 'training data' and 'feature vectors.' Bridging that epistemological gap is essential to ensure the AI system is built on a foundation of cultural understanding, not just statistical pattern recognition." This requires creating a shared vocabulary and collaborative framework from the outset.

AI specialists and data scientists acknowledge this challenge. One AI specialist emphasized: "Efforts to enhance AI literacy in education must progress through strategic interdisciplinary integration and address various field-based constraints." One of the data scientists also revealed: "AI literacy requires blending insights from computer science, cognitive science, education, and ethics to create a unified teaching framework."

In more detail, linguists and cultural custodians highlighted the challenges encountered and advocated for the participation of key stakeholders. A linguist notes: "This cross-disciplinary approach presents several practical challenges that need to be addressed, including the alignment of curricula, the readiness of teachers, the preparedness of institutions, and ensuring equitable access to technological resources." Another linguist confirms: "Successful implementation requires translating abstract AI concepts into learning experiences that align with students' developmental stages while maintaining conceptual rigor". A cultural custodian also asserts: "Successfully integrating AI education requires a balance between in-depth understanding and practical pedagogical approaches, adapting to

various classroom environments, policy requirements, and current infrastructure challenges."

Through multidisciplinary collaboration, AI serves as a potent enabler, but the heartbeat of the project remains the living community and its self-determined goal of linguistic and cultural perpetuation. Advancing AI literacy in education requires careful consideration of interdisciplinary integration and the constraints of real-world implementation (Allen & Kendeou, 2024). As Atifnigar et al., (2025) conclude, effective revitalization necessitates interdisciplinary collaboration, equitable policies, and global partnerships, with UNESCO and international organizations crucial in protecting linguistic heritage, alongside local efforts and community engagement.

Furthermore, issues of long-term sustainability and infrastructure were raised. Who will maintain, update, and host the mobile application and its AI backend after the research funding ends? How will digital divide issues—such as uneven internet access in rural parts of Aceh—affect the application's reach and equity? A data scientist pragmatically noted, "We can build a brilliant prototype, but if there is no plan for server costs, model updates, and content moderation, it will become a digital relic within two years." A cultural custodian also narrates: "Cultural development initiatives often come to an end when financial backing is withdrawn. Therefore, efforts to digitize proverbs require careful planning to guarantee their long-term sustainability."

This concern is supported by the literature on digital humanities projects, which shows that sustainability planning for large-scale projects must integrate data, technology, community, communications, and process knowledge. It advocates for viewing projects not just as assets but also considering the potential user base and their perceptions of value (Edmond & Morselli, 2020). The costs of digital scholarship include not only human labor, rights abuses, environmental damage, and resource waste in the production lifecycle, but also encompass sustainability as an epistemic concept within complex systems, rather than merely addressing logistical problems (Drucker, 2021). Thus, the success of digital language revitalization projects depends not only on the technical robustness but also on the commitment to community knowledge, Indigenous data governance and the collaborative path required to reach it.

5. CONCLUSION

This qualitative, multi-stakeholder feasibility study elucidates the complex dialectic inherent in proposing AI-driven tools for the preservation of Acehnese proverbs, a critical yet endangered component of intangible cultural heritage. The findings reveal a landscape defined by a compelling duality: transformative potential juxtaposed with profound socio-technical challenges.

Unanimously, stakeholders recognized the opportunity for interactive AI, such as chatbots and mobile applications, to act as a dynamic bridge for intergenerational transmission, potentially revitalizing engagement among digitally native youth by making proverbial wisdom accessible and relevant. This aligns with the broader digital revitalization paradigm seeking to move beyond static archives. However, this promise is rigorously tempered by

significant hurdles. The foundational constraint is the "low-resource" paradox, where the advanced data-hungry architectures of AI confront the scarcity of standardized, machine-readable, and annotated Acehnese linguistic data, creating severe bottlenecks for natural language processing and authentic speech synthesis.

More critically, the analysis foregrounds risks that transcend technical limitations, centering on ethical and cultural integrity. Paramount among these are the dangers of decontextualization, where the rich performative and normative gravity of *peribahasa* could be diminished through gamification or reduction to deconstructed data points, and the violation of community sovereignty over knowledge. The insistence of cultural custodians on ongoing authority in curating, interpreting, and governing digital content underscores the non-negotiable imperative to align any technological intervention with the CARE Principles for Indigenous data governance.

Furthermore, the study exposes the practical complexities of interdisciplinary collaboration, where differing epistemologies between technologists and cultural specialists must be bridged, and the long-term sustainability of such digital initiatives beyond initial funding cycles. Ultimately, the research concludes that an AI application for Acehnese proverbs is not a straightforward technical solution but a complex socio-technical proposition. Its viability is contingent not primarily on algorithmic innovation but on a foundational, community-led process that prioritizes ethical co-design, the building of culturally-grounded datasets, and the establishment of robust governance frameworks before any substantial coding begins.

6. IMPLICATIONS AND RECOMMENDATIONS

The implications of this study are twofold, offering both cautionary guidance and a constructive pathway forward. For researchers and practitioners aiming to develop similar tools for other endangered linguistic and cultural heritage contexts, this case serves as a critical precedent. It demonstrates that a feasibility assessment focusing on socio-technical and ethical dimensions is a necessary precursor to large-scale development, effectively mitigating risks of unintended harm, community alienation, and project failure. The primary implication is that technological prowess must be subordinated to cultural protocol and community agency; otherwise, such projects risk perpetuating digital colonialism under the guise of preservation.

Consequently, the study offers several concrete recommendations. First, for the envisioned Acehnese project, it is recommended to reconceptualize the initiative from an "AI development project" to a "community-led cultural documentation and revitalization project enabled by AI." The immediate priority should be a participatory, ethics-first preliminary phase dedicated to co-designing data collection protocols with cultural custodians, establishing a joint governance council, and systematically building a high-quality, multimodal (text, audio, contextual narrative) and ethically sourced corpus. This corpus itself is a vital outcome of preservation. Subsequent AI development should then proceed iteratively, starting with simpler, low-risk functionalities like a searchable audio repository, and gradually introducing interactive elements like chatbots, with continuous feedback loops embedded within the governance structure.

For future researchers, this study highlights several critical avenues for inquiry. There is a pressing need for applied research into lightweight, low-resource AI model architectures specifically designed for endangered language contexts, potentially leveraging transfer learning and few-shot learning techniques. Furthermore, interdisciplinary methodological research is required to develop best practices for effective collaboration between data scientists, linguists, anthropologists, and community knowledge holders, creating shared vocabularies and collaborative frameworks. Another vital area is longitudinal study on the long-term sustainability of digital revitalization tools, examining models for community-based technical maintenance, funding, and content evolution.

Finally, future work must continue to refine and operationalize ethical frameworks like FAIR and CARE in practical project guidelines, translating principles into actionable steps for access control, benefit-sharing, and ongoing consent in AI development. By addressing these areas, subsequent research can help ensure that the formidable power of artificial intelligence is harnessed as a genuinely respectful and effective instrument for cultural continuity, rather than a disruptive force that accelerates the very erosion it seeks to prevent.

REFERENCES

Abdalla, H. B., Kumar, Y., Marchena, J., Guzman, S., Awlla, A., Gheisari, M., & Cheraghy, M. (2025). The Future of Artificial Intelligence in the Face of Data Scarcity. *Computers, Materials & Continua*, *84*(1), 1073–1099. https://doi.org/10.32604/cmc.2025.063551

Abingosa, D., Bokingkito, P., Pasandalan, S. N., Alovera, J. R. G., & Otano, J. (2025). Digitizing the Higaonon Language: A Mobile Application for Indigenous Preservation in the Philippines. *Informatics*, *12*(3), 90. https://doi.org/10.3390/informatics12030090

Al-Auwal, T. M. R. (2017). Reluctance of Acehnese youth to use Acehnese. *Studies in English Language and Education*, *4*(1), 1–14. https://doi.org/10.24815/siele.v4i1.7000

Al-Auwal, T. M. R., Amery, R., & Green, I. (2024). Language shift in Aceh: The sociolinguistic situation of post-conflict society. *Studies in English Language and Education, 11*(3), 1748–1766. https://doi.org/10.24815/siele.v11i3.39159

Allen, L. K., & Kendeou, P. (2024). ED-AI Lit: An Interdisciplinary Framework for AI Literacy in Education. *Policy Insights from the Behavioral and Brain Sciences*, *11*(1), 3–10. https://doi.org/10.1177/23727322231220339

Ardalan, I. D., Banifatemi, A., Gonzalez, F., Ingram, M., Moradinezhad, R., & Williams, L. (2025). *AI for Community*. Chapman and Hall/CRC. https://doi.org/10.1201/9781003517115

Atifnigar, H., Dawlatzai, A. K., & Habib, S. (2025). Effective Strategies for Preserving Mother Tongues: Approaches to Linguistic and Cultural Sustainability. *European Journal of Contemporary Education and E-Learning*, *3*(2), 3–14. https://doi.org/10.59324/ejceel.2025.3(2).01

Barnes, A. J., Zhang, Y., & Valenzuela, A. (2024). AI and culture: Culturally dependent responses to AI systems. *Current Opinion in Psychology*, *58*, 101838. https://doi.org/10.1016/j.copsyc.2024.101838

Battista, D. (2024). The Digital as Sacred Space: Exploring the Online Religious Dimension. *Academicus International Scientific Journal*, *29*, 21–37. https://doi.org/10.7336/academicus.2024.29.02

Blokland, R., Fedina, M., Gerstenberger, C., Partanen, N., Rießler, M., & Wilbur, J. (2015). Language Documentation meets Language Technology. *Septentrio Conference Series*, (2), 8. https://doi.org/10.7557/5.3457

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brinkmann, S., & Kvale, S. (2018). *Doing Interviews*. SAGE Publications Ltd. https://doi.org/10.4135/9781529716665

Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, *8*(1), 108. https://doi.org/10.1038/s41597-021-00892-0

Castro, A., Díaz, B., Aguilera, C., Prat, M., & Chávez-Herting, D. (2025). Identifying Rural Elementary Teachers' Perception Challenges and Opportunities in Integrating Artificial Intelligence in Teaching Practices. *Sustainability*, *17*(6), 2748. https://doi.org/10.3390/su17062748

Chen, B., & Chen, L. (2025). 'Speak Hokkien': language revitalisation and discursive opportunity structures for Chinese in Penang, Malaysia. *Language and Intercultural Communication*, 1–17. https://doi.org/10.1080/14708477.2025.2471946

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design*. SAGE.

Drucker, J. (2021). Sustainability and complexity: Knowledge and authority in the digital humanities. *Digital Scholarship in the Humanities*, *36*(Supplement_2), ii86–ii94. https://doi.org/10.1093/llc/fqab025

Edmond, J., & Morselli, F. (2020). Sustainability of digital humanities projects as a publication and documentation challenge. *Journal of Documentation*, *76*(5), 1019–1031. https://doi.org/10.1108/JD-12-2019-0232

Freeman, S. B. (2025). Ghost in the Machine: Metadata, Commodification, and the Specter of Artificial Intelligence. *Journal of Library Metadata*, *25*(3), 223–239. https://doi.org/10.1080/19386389.2025.2523717

Gangwal, A., Ansari, A., Ahmad, I., Azad, A. K., & Wan Sulaiman, W. M. A. (2024). Current strategies to address data scarcity in artificial intelligence-based drug discovery: A comprehensive review. *Computers in Biology and Medicine*, *179*, 108734. https://doi.org/10.1016/j.compbiomed.2024.108734

Gerlich, M. (2023). Perceptions and Acceptance of Artificial Intelligence: A Multi-Dimensional Study. *Social Sciences*, *12*(9), 502. https://doi.org/10.3390/socsci12090502

Granjeiro, J. M., Cury, A. A. D. B., Cury, J. A., Bueno, M., Sousa-Neto, M. D., & Estrela, C. (2025). The Future of Scientific Writing: AI Tools, Benefits, and Ethical Implications. *Brazilian Dental Journal*, *36*. https://doi.org/10.1590/0103-644020256471

Grenoble, L. A. (2010). Language documentation and field linguistics. In *Language Documentation* (pp. 289–310). John Benjamins Publishing Company. https://doi.org/10.1075/z.158.28gre

Horna-Saldaña, C. J., Perez Perez, J. E., & Toro Galeano, M. L. (2025). Artificial intelligence in the preservation of native languages and bridging the information access

gap for indigenous peoples. *Journal of Enabling Technologies*, *19*(1), 63–75. https://doi.org/10.1108/JET-09-2024-0063

Ibrahim, I. H., & Usman, J. (2021). Cultural Values in Acehnese Farming-Related Proverbs. *Indonesian Journal of Applied Linguistics*, *11*(2). https://doi.org/10.17509/ijal.v11i2.32323

Ishkhanyan, A. (2025). Ethical considerations in AI-powered language technologies: insights from East and West Armenian. *AI and Ethics*, *5*(4), 4135–4146. https://doi.org/10.1007/s43681-025-00716-6

Jia, S., Lu, Z., & Bava Harji, M. (2025). A mobile-assisted language learning speaking model: development and evaluation. *Computer Assisted Language Learning*, 1–28. https://doi.org/10.1080/09588221.2025.2599150

Jiang, P., Niu, W., Wang, Q., Yuan, R., & Chen, K. (2024). Understanding Users' Acceptance of Artificial Intelligence Applications: A Literature Review. *Behavioral Sciences*, *14*(8), 671. https://doi.org/10.3390/bs14080671

Julia, P., & Jeyanthi, D. B. (2025). IMPACT OF MOBILE-ASSISTED LANGUAGE LEARNING ON ESL LITERACY SKILLS. *Lex Localis - Journal of Local Self-Government*, *23*(S6), 8300–8325. https://doi.org/10.52152/fkpknf66

Muntazar, T., & Hajar, I. (2025). Discursive and Metalinguistic Perspectives on AI-Generated Writing Feedback: A Student Perception Study. *Journal Informatic, Education and Management (JIEM)*, *7*(2), 516–526. https://doi.org/10.61992/jiem.v7i2.168

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, *49*(5), 847–864. https://doi.org/10.1080/03075079.2024.2323593

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis. *International Journal of Qualitative Methods*, *16*(1). https://doi.org/10.1177/1609406917733847

Patton, M. Quinn. (2015). *Qualitative research & evaluation methods: integrating theory and practice*. SAGE Publications, Inc.

Rousan, S. al, Al-Jbour, R., Rousan, R. M. Al, AlRousan, M., & Hassan, H. (2025). Beyond Words: A Linguistic and Cultural Analysis of Tribe's Image in Jordanian Proverbs. *Theory and Practice in Language Studies*, *15*(5), 1658–1666. https://doi.org/10.17507/tpls.1505.30

Satyawati, Astini, N. M. D., Sumarni, N. M., Kanta, K., Safar, J., Lutzenberger, H., Palfreyman, N., & de Vos, C. (2025). The Balinese Homesign Corpus: New insights into corpus development in a rural signing context. *Language Documentation and Conservation*, *19*, 305–327.

Smith, L. Tuhiwai. (2024). *Decolonizing methodologies: research and indigenous peoples*. Bloomsbury Academic.

Tang, J., Zhai, Y., Li, L., Liu, P., & Deng, H. (2025). Mobile learning for less-commonly taught languages: design and application. *Journal of China Computer-Assisted Language Learning*, *5*(1), 23–55. https://doi.org/10.1515/jccall-2024-0006