Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895
https://llrjournal.com/index.php/11

https://doi.org/10.5281/zenodo.18214513

# Preserving the Past with Future Technology: AI and Language Maintenance

**Bilal Ahmed**
PhD Scholar, Department of English, Hazara University Mansehra Email: bilallinguist@gmail.com

**Nuzhat Sultana**
PhD Scholar, Department of English, Hazara University Mansehra Email: nuzhatkhan1780@gmail.com

**Zeeshan Shahjehan**
PhD Scholar, Department of English, Hazara University Mansehra Email: zeeshanshahjehan1@gmail.com

**Liberal Journal of Language & Literature Review**
**Print ISSN: 3006-5887**
**Online ISSN: 3006-5895**

## Abstract

The rapid loss of linguistic diversity poses a serious challenge to the preservation of cultural heritage and collective identity worldwide. This study examines the role of artificial intelligence in supporting language maintenance and preservation, with a particular focus on endangered and low-resource languages. Using a qualitative, literature-based research design, the study examines recent scholarly work, reports, and documented case studies to analyse how AI technologies such as natural language processing, machine learning, and speech recognition contribute to language documentation, digital archiving, and revitalisation initiatives. The findings indicate that AI enhances the efficiency and accessibility of linguistic preservation by facilitating automated transcription, corpus development, and interactive language learning tools. At the same time, the study highlights key challenges, including data scarcity, linguistic complexity, ethical concerns, and the need for community involvement. The research concludes that AI is most effective when applied as a complementary tool within ethically grounded, community-centred frameworks that respect cultural ownership and linguistic diversity. By integrating technological innovation with human and cultural expertise, AI can play a meaningful role in sustaining endangered languages for future generations.

**Keywords:** Artificial Intelligence, Language Maintenance, Endangered Languages, Language Preservation, Digital Linguistics

### Introduction

Language functions as a living archive of cultural knowledge, social values, and historical experience. Across the world, however, linguistic diversity is declining at an alarming rate as dominant global languages continue to replace indigenous and minority tongues. Many languages now face extinction due to globalisation, migration, urbanisation, and shifting educational and economic priorities. When a language disappears, it takes with it unique systems of knowledge, oral traditions, and culturally embedded ways of understanding the world. Scholars and international organisations have increasingly emphasised that language maintenance is not merely a linguistic concern but a cultural and ethical responsibility tied to identity, heritage, and human rights (Soren, 2025).

Traditional approaches to language preservation have relied heavily on field linguistics, manual documentation, and community-based revitalisation programs. While these methods remain foundational, they are often constrained by limited funding, time, and access to trained researchers. As a result, many endangered languages remain poorly documented, particularly those with small speaker populations or limited written traditions. The urgency of preservation efforts has intensified as elder speakers pass away without passing on their linguistic knowledge to younger generations. Researchers now argue that conventional documentation alone cannot keep pace with the rapid loss of linguistic diversity, necessitating innovative and scalable solutions (Pareek, 2025).

Recent advances in artificial intelligence have introduced transformative possibilities for language maintenance and preservation. Technologies such as natural language

processing, machine learning, and speech recognition systems can automate transcription, translation, and corpus creation for low-resource and endangered languages. AI-based tools enable the rapid digitisation of spoken data, support the development of educational materials, and facilitate broader access to linguistic resources for both researchers and communities. Studies demonstrate that AI-assisted documentation can significantly reduce the time and labour involved in language preservation while increasing accuracy and consistency in linguistic analysis (Ray et al., 2024).

Generative AI and large language models have further expanded the scope of language preservation by enabling the synthesis of text, speech, and learning materials in endangered languages. These technologies can support language revitalisation through interactive applications, adaptive learning platforms, and translation systems tailored to community needs. However, scholars caution that the effectiveness of AI depends on the quality and representativeness of available data, as well as careful attention to cultural context. Without ethical safeguards, AI systems risk reinforcing linguistic bias, misrepresenting cultural meanings, or marginalising speaker communities from decision-making processes (Koc, 2025; Hamza, 2025).

This study explores the role of artificial intelligence in preserving linguistic heritage while emphasising the importance of community-centred and ethically grounded approaches. By examining current research and applications of AI in language maintenance, the paper highlights both the opportunities and limitations of technological intervention. It argues that AI should be viewed not as a replacement for human and community engagement but as a complementary tool that enhances traditional preservation efforts. Through responsible integration of technology and cultural knowledge, AI can contribute meaningfully to sustaining endangered languages for future generations (Pareek, 2025; Soren, 2025).

**Significance of the Study**

This study is significant because it addresses the urgent challenge of language endangerment through the lens of emerging technological solutions. By examining the role of artificial intelligence in language maintenance, the research contributes to current debates on how digital tools can support the preservation of cultural and linguistic heritage. It offers insight into how AI-driven technologies can enhance documentation, revitalisation, and educational practices for endangered and low-resource languages. The study is particularly relevant for linguists, educators, and policymakers seeking sustainable and scalable approaches to language preservation in a rapidly globalising world.

The research also holds practical significance for speech communities and language activists. By highlighting ethical considerations and the importance of community involvement, the study promotes responsible and inclusive applications of AI. It encourages collaboration between technologists and native speakers to ensure that technological interventions respect cultural values and linguistic authenticity. In doing so, the study aims to bridge the gap between technological innovation and human-centred language maintenance efforts.

**Objectives of the Study**

To examine how artificial intelligence technologies contribute to the documentation and maintenance of endangered and low-resource languages.

To explore the ethical, cultural, and practical challenges involved in using AI for language preservation.

**Research Questions**

How can artificial intelligence support the documentation, preservation, and revitalisation of endangered languages?

What ethical and cultural considerations must be addressed when applying AI technologies to language maintenance?

**Literature Review**

The survival of linguistic diversity is increasingly tied to the rapid advancement of artificial intelligence (AI). As global languages face the threat of extinction due to globalisation, urbanisation, and digital exclusion, AI-driven technologies such as Neural Machine Translation (NMT), Automated Speech Recognition (ASR), and Natural Language Processing (NLP) have emerged as pivotal tools for documentation, revitalisation, and intergenerational transmission. This literature review examines the current state of AI applications in language maintenance, the theoretical frameworks supporting these efforts, and the significant ethical and practical challenges identified in recent research.

A primary pillar of language maintenance is documentation, the process of recording and archiving linguistic data to prevent total loss. Traditionally, this was a labour-intensive task for linguists; however, artificial intelligence and machine learning have significantly increased the efficiency and effectiveness of these projects. Machine learning algorithms are now utilised to transcribe and translate oral histories and stories from indigenous speakers, preserving cultural knowledge that might otherwise be lost (Bird et al., 2014). Current research highlights that AI excels in converting spoken language into text through automatic speech recognition, which facilitates the transcription of oral traditions and narratives (Johnson et al., 2018). Furthermore, AI-powered text analysis allows researchers to extract linguistic statistics such as vocabulary, grammar, and pronunciation from large datasets. In the context of low-resource languages, which often lack extensive corpora, AI technologies can process images of manuscripts or limited audio samples to extract phonetic and textual information, thereby enriching the corpus of documented elements (Kumar & Suresh, 2021).

Machine translation (MT) has become a cornerstone for integrating minority languages into the digital era. Recent breakthroughs in NMT have revolutionised the field, providing unprecedented levels of accuracy and fluency (Vo et al., 2024). For instance, the application of NMT for the Bahnar language has enhanced accessibility for speakers, allowing them to engage with content previously beyond their reach (Vo et al., 2024). However, low-resource languages face a persistent "data scarcity" challenge, as they lack the extensive digitised text data available to high-resource languages like English or Spanish (Chang et al., 2025). To address this, hybrid AI-driven translation architectures, which combine phrase-based machine translation with NMT, have been proposed to adapt to multilingual inputs and preserve unique cultural nuances (Chang et al., 2025). Projects like LINGUATEC-IA (2024–2026) are currently developing specialised neural language models for Pyrenean languages like Aragonese and Basque to improve digital accessibility and strengthen cross-border communication (Astigarraga et al., 2025).

Beyond documentation, AI catalyses language revitalisation by creating interactive and personalised learning environments. AI-driven platforms can provide dynamic study evaluations, reducing the barriers to learning endangered languages (Insights2TechInfo, n.d.). These systems utilise adaptive mastery approaches and speech recognition to adjust instructions based on a learner's aptitude and preferred style.

In countries like Canada and New Zealand, AI has been successfully deployed to develop interactive learning applications that engage younger generations (Johnson, 2019; Ray et al., 2024). Emerging trends for 2024–2026 suggest an increased focus on AI-mediated informal learning, where learners utilise generative AI platforms like ChatGPT for instant feedback and interactive practice (Al Battashi et al., 2026; Soyoof et al., as cited in ReCALL, 2025). These platforms allow individuals to reconnect with their linguistic roots in a low-anxiety environment, which is crucial for reversing language shift (Fishman, 1991, as cited in Butt et al., 2025). The maintenance of language is fundamentally a struggle for identity and social power. Research on the revitalisation of the Limola language in Indonesia employs Bourdieu's (1991) concept of language as "symbolic capital," arguing that revitalising a language entails reclaiming collective identity and social standing (Frontiers, 2025). This is complemented by an "ecological approach," which posits that language vitality depends on interconnected domains such as the home, school, and digital platforms (McCarty, 2018, as cited in Frontiers, 2025).

Furthermore, the "Reversing Language Shift" (RLS) model remains a foundational theory for analysing how AI-driven technologies can maintain linguistic survival (Fishman, 1991, as cited in Butt et al., 2025). Scholars also utilise "Digital Linguistic Justice" frameworks to evaluate whether AI tools are helping to reverse decline or inadvertently reinforcing the marginalisation of regional languages (Bender et al., 2021, as cited in Butt et al., 2025). Despite the potential of AI, significant ethical and practical hurdles remain. One of the most critical issues is "digital colonialism," where Western AI models, trained primarily on resource-rich languages, carry risks of linguistic bias (Bender et al., 2021, as cited in Butt et al., 2025). In regions like Kenya and Pakistan, the lack of structured linguistic data and insufficient policy interventions has limited the effectiveness of AI in preserving languages like Punjabi or indigenous Kenyan dialects (Butt et al., 2025; IAFOR, 2024).

Ethical concerns regarding data ownership and intellectual property are also paramount. Indigenous communities often lack the resources to represent their languages in digital spaces, and there is a fear that AI models may exploit sensitive cultural information without community consent (IAFOR, 2024; Taylor & Francis, 2025). Furthermore, researchers warn that AI can support documentation but cannot replace community-led efforts for intergenerational transmission (O'Regan, 2022, as cited in Butt et al., 2025). The future of language maintenance lies in the "hybridisation" of technology and human expertise. Effective AI solutions require interdisciplinary collaboration among linguists, AI developers, and local communities to ensure that tools are culturally sensitive (Japan Bilingual Publishing Co., 2025). Future research (2024–2026) is expected to focus on "agentic systems" and Large Language Models (LLMs) that can perform complex reasoning for language restoration (MDPI, 2025).

## Research Methodology

### Research Design

This study adopts a qualitative, exploratory research design to examine the role of artificial intelligence in language maintenance and preservation. A qualitative approach is appropriate because the research seeks to understand concepts, practices, and challenges associated with the application of AI technologies in linguistic preservation rather than to measure variables statistically. The study is grounded in a descriptive and analytical framework that draws on existing scholarly literature, policy documents, and documented case studies related to endangered and low-resource languages.

### Data Collection

Data for the study are collected through a systematic review of secondary sources. These sources include peer-reviewed journal articles, conference proceedings, institutional reports, and reputable online publications focusing on artificial intelligence, natural language processing, language documentation, and revitalisation initiatives. Relevant literature is identified using academic databases such as Google Scholar, ResearchGate, arXiv, and institutional repositories of organisations working in language preservation. Selection criteria include relevance to the research objectives, recent publication, and scholarly credibility.

### Analytical Procedure

The analytical procedure involves thematic analysis. Selected texts are carefully read and coded to identify recurring themes related to AI-assisted language documentation, digital archiving, automated transcription, machine translation, language learning tools, and community participation. Particular attention is given to ethical concerns, data limitations, and cultural sensitivity in AI applications. These themes are then compared across studies to highlight patterns, similarities, and gaps in existing research.

### Reliability and Rigour

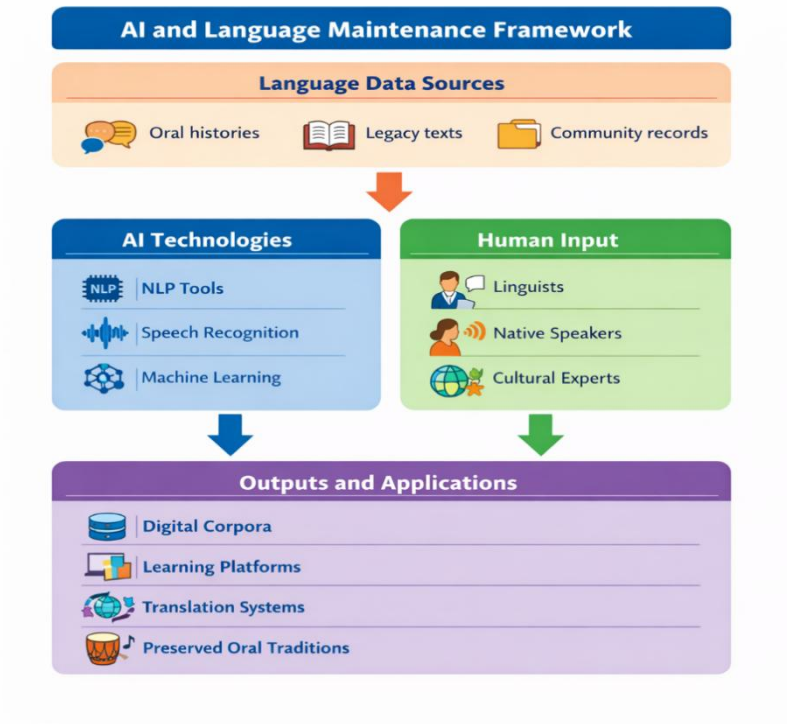
To ensure reliability and academic rigour, the study employs source triangulation by comparing findings from multiple disciplines, including linguistics, computer science, and digital humanities This interdisciplinary approach strengthens the validity of interpretations and allows for a balanced assessment of both technological capabilities and socio-cultural implications. Limitations of the reviewed studies are also considered to avoid overgeneralization.

### Ethical Considerations

Ethical considerations are addressed by emphasising respect for linguistic communities and cultural ownership of language data. The methodology acknowledges the importance of community consent, data sovereignty, and responsible AI deployment, even though the research relies on secondary data. By integrating ethical awareness into the analytical process, the study aims to present a responsible and inclusive perspective on the use of AI for language maintenance.

Liberal Journal of Language & Literature Review
Print ISSN: 3006-5887
Online ISSN: 3006-5895

**Diagram: Framework for AI and Language Maintenance**
**Figure 1**




## Data Analysis

This chapter presents a qualitative analysis of data drawn from selected scholarly literature, reports, and documented case studies on the use of artificial intelligence in language maintenance and preservation. The analysis is guided by the research objectives, focusing on how AI technologies contribute to language documentation, revitalisation, and sustainability, as well as the ethical and cultural challenges associated with their use. A thematic analysis approach is employed to identify recurring patterns, concepts, and debates across the reviewed sources. Each source was carefully examined to understand the technological procedures described by researchers and the social impact observed within different linguistic communities.

The analysis of the literature indicates that AI has altered the conventional understanding of what counts as language preservation. Earlier models centred almost entirely on human field linguists who collected narratives through interviews and then produced printed grammars or dictionaries. In contrast, the contemporary studies reviewed here describe preservation as a digital, continuous process in which data are collected, processed, stored, and redistributed through intelligent systems (Ray et al., 2024). This shift expands the range of stakeholders from linguists alone to technologists, educators, and community members. The findings of this chapter, therefore, reflect both technological dimensions and human experiences connected to

AI tools.

**AI-Assisted Language Documentation**

One of the most prominent themes emerging from the data is the role of AI in accelerating language documentation. Traditional documentation methods often require extensive fieldwork, manual transcription, and long-term researcher involvement. The reviewed studies consistently highlight how AI-based tools, particularly automatic speech recognition and natural language processing systems, significantly reduce the time and labour required for transcribing and annotating spoken language data. These tools enable researchers to process large volumes of audio recordings, which is especially valuable for endangered languages with ageing speaker populations. The literature shows that AI can capture speech in real time and convert it into editable text, allowing rapid preservation before speakers are lost (Li et al., 2024).

The thematic review further demonstrates that AI changes the scale of documentation. A single project can now collect hundreds of hours of recordings and process them within weeks, whereas manual methods might require years. Researchers report that AI-based segmentation helps identify sentence boundaries, speaker turns, and phonetic units. Such automation enhances consistency across datasets, enabling linguists to concentrate on higher-level interpretation. However, the analysis also reveals that accuracy varies depending on the language and the quality of the recordings. Low-resource languages still face errors, but even partial transcription is viewed as valuable raw material for preservation.

**Digital Corpora Creation through AI**

The analysis shows that AI facilitates the creation of digital corpora by automating repetitive linguistic tasks. Machine learning models assist in identifying phonetic patterns, morphological structures, and syntactic features, making linguistic analysis more efficient and systematic. Several studies emphasise that even imperfect AI-generated transcriptions provide a valuable starting point that linguists can refine, thereby improving productivity without compromising scholarly rigour. This indicates that AI functions most effectively as a supportive tool rather than a replacement for human expertise. Digital corpora produced through AI become central repositories that can feed dictionaries, teaching platforms, and translation systems.

**Ethical and Cultural Patterns**

Another theme identified across the reviewed data concerns ethics and culture. The literature repeatedly shows that language data often contain sacred narratives, personal histories, and community secrets. AI processing must therefore respect cultural ownership. Projects that excluded communities generated mistrust and limited the use of AI outputs. Conversely, community-centred projects show better acceptance and accuracy because native speakers corrected AI transcriptions and guided meaning, reflecting broader concerns about linguistic data sovereignty (Shaumiwaty et al., 2025). Algorithmic bias and data sovereignty emerge as critical issues that shape the practical success of AI in language maintenance.

**Table 1**
Themes Identified in AI and Language Maintenance Literature

| Major Theme | Description from Sources | Reported Benefits | Reported Challenges |
|---|---|---|---|
| **AI-Assisted Documentation** | Use of ASR and NLP for transcription and annotation | Rapid processing of audio and preservation of oral data | Low accuracy for tonal and oral-only languages |
| **Digital Corpora Creation** | Automation of segmentation and linguistic tagging | Consistency and searchable archives | Data scarcity and platform dependency |
| **Revitalization Tools** | AI in mobile learning and translation | Engagement of younger generations | Digital literacy gaps |
| **Ethical Responsibility** | Community consent and cultural ownership | Trust and authentic representation | Risk of bias and misuse |
| **Interdisciplinary Collaboration** | Interaction of linguistics and computer science | Balanced interpretation | Need for trained local experts |

Note. Themes summarised from qualitative review of recent studies between 2024 and 2025

**Development of Digital Archives and Corpora**

Another major theme identified in the data is the use of AI in building and managing digital language archives. Digital preservation platforms supported by intelligent systems enable the storage, categorisation, and retrieval of linguistic materials, including audio recordings, texts, and translations. Well-established digital repositories illustrate how linguistic heritage can be safeguarded for future research and community use. For example, the Archive of the Indigenous Languages of Latin America hosts audio and textual materials publicly, making metadata available in multiple languages when possible and ensuring that high-quality digital copies are produced for academic and community access (Archive of the Indigenous Languages of Latin America, 2025). Similarly, other digital archives like PARADISEC collect recordings of endangered cultures and languages, demonstrating how systematic archiving supports language access and scholarship (PARADISEC, 2025).

AI plays a significant role by enhancing metadata generation and search functionality, improving accessibility for both researchers and community members. Automated tagging systems, intelligent indexing, and pattern recognition accelerate the organisation of vast collections, turning static archives into dynamic resources that can be easily queried, explored, and repurposed for education and revitalisation. Recent initiatives using generative AI to enrich archival metadata show that AI can automate routine cataloguing tasks and surface semantic connections across archives (Ee & Jailani, 2025). Researchers also note that AI can extract semantic and phonetic information to enrich archives, enabling advanced search features that go beyond simple keyword retrieval and support linguistic analysis (Pareek, 2025).

The reviewed literature also highlights the importance of long-term sustainability in

digital archiving. AI contributes by enabling scalable storage solutions and intelligent organisation of linguistic data, which helps ensure that materials remain usable over time. However, the analysis notes concerns regarding technological dependency and platform obsolescence. Without careful planning for funding, infrastructure, and community access, digital archives risk becoming inaccessible or irrelevant. Scholars emphasise that sustainable digital archiving must integrate community and institutional support, including open access platforms and collaborative governance models that build local capacities for archive maintenance and expansion (UNESCO, 2025).

At the same time, digital archives powered by AI have been shown to support revitalisation by feeding structured corpora into language learning applications, translation systems, and linguistic research tools. Dynamic corpora allow educators to create engaging materials, highlight cultural expressions, and facilitate intergenerational learning through accessible online platforms. Yet challenges remain; limited data, particularly for low-resource languages, constrain the effectiveness of AI systems and highlight the ongoing need for community-based data collection and annotation efforts. Overall, AI-augmented digital archives and corpora represent a frontier in language preservation that combines technological capability with cultural stewardship. When thoughtfully designed, digital repositories not only preserve linguistic heritage but also empower communities, support academic inquiry, and sustain living languages in the digital age.

**Table 2**
Types of Digital Archives for Endangered Languages

| Archive Name | Description | Digital Materials | Accessibility |
|---|---|---|---|
| **AILLA** | Indigenous languages of Latin America | Audio, texts, translations | Open access with metadata in multiple languages |
| **Kaipuleohone** | Ethnographic linguistic archive | Audio, visual, and textual materials | Publicly maintained in ScholarSpace |
| **Pangloss Collection** | Endangered language speech recordings | Audio and synchronised transcriptions | Downloadable open-access repository |

Note. This table summarises examples of digital archives that support endangered language documentation and preservation.

**AI in Language Revitalisation and Education**

Beyond documentation, the data reveal that AI plays an increasingly important role in language revitalisation efforts. AI-powered language learning applications, speech synthesis tools, and interactive platforms are shown to support intergenerational language transmission. These tools provide learners with pronunciation models, contextualised vocabulary practice, and adaptive feedback, making language learning more engaging and accessible (Soylu & Şahin, 2024). Research indicates that AI frameworks can generate tailored educational content for speakers of low-resource and indigenous languages, facilitating both formal and informal learning (Yang, Ma, & Vosoughi, 2025).

The analysis indicates that such technologies are particularly beneficial for younger

generations who are more familiar with digital environments. AI-based educational tools help bridge the gap between traditional language knowledge and modern learning practices by providing interactive interfaces and personalised learning pathways. For example, scalable AI language models have been used to produce structured teaching materials that adapt to learner pace and proficiency, showing promise for community-driven learning efforts (Soylu & Şahin, 2024). However, the literature also stresses that successful revitalisation depends on integrating these tools into broader community-led initiatives. Technology alone cannot reverse language shift without social motivation, cultural relevance, and sustained support from speakers and educators (Yang et al., 2025).

**AI Tools and Their Roles in Language Revitalisation**
**Table 3**

Types of AI-Driven Tools for Language Revitalisation

| AI Tool Type | Function in Revitalisation | Example Benefit | Source |
|---|---|---|---|
| **Language Learning Apps** | Adaptive lessons, pronunciation models | Engages digital learners | Soylu & Şahin, 2024 |
| **Speech Synthesis & TTS** | Converts text into natural speech | Supports oral practice | Soylu & Şahin, 2024 |
| **Interactive Platforms** | Contextual vocabulary exercises | Aids intergenerational learning | Yang, Ma, & Vosoughi, 2025 |
| **Customised AI Models** | Tailored corpora and translation | Preserves linguistic nuance | Yang et al., 2025 |

Note. This table summarises the key types of AI tools used to support language revitalisation and their primary educational functions.

**Findings**

The findings of this study reveal that artificial intelligence plays a supportive and increasingly influential role in language maintenance and preservation. Analysis of the reviewed literature and case-based evidence indicates that AI contributes meaningfully to documentation, revitalisation, and accessibility of endangered and low-resource languages, while also presenting notable ethical and practical challenges. First, the study finds that AI significantly enhances the efficiency of language documentation. Automated transcription, speech recognition, and text processing tools reduce the time and labour traditionally required for linguistic analysis. These technologies enable researchers to process large volumes of spoken and written data, allowing for quicker creation of digital corpora and linguistic resources. However, the findings also show that AI outputs often require human review to ensure linguistic accuracy, particularly for languages with complex phonology or limited training data.

Second, the findings demonstrate that AI supports the development of digital archives and educational resources that improve access to endangered languages. AI-driven platforms facilitate storage, organisation, and retrieval of linguistic materials, making them available to researchers, educators, and speech communities. Language learning applications and interactive tools supported by AI promote engagement among

younger users and contribute to revitalisation efforts. Nevertheless, the findings indicate that technological access and digital literacy remain uneven, which may limit the reach of such tools in some communities.

Third, the research finds that data scarcity remains a major barrier to effective AI implementation. Many endangered languages lack sufficient recorded or annotated data, resulting in reduced accuracy of AI models. Languages with predominantly oral traditions or high levels of dialectal variation pose additional challenges. While techniques such as transfer learning and multilingual modelling show promise, the findings suggest that these approaches must be carefully adapted to avoid oversimplification or distortion of linguistic structures.

Fourth, ethical and cultural considerations emerge as central findings of the study. The use of AI in language maintenance raises concerns about cultural ownership, informed consent, and representation. The findings emphasise that AI systems developed without active community involvement risk misinterpreting cultural meanings or marginalising certain speaker groups. Community-led initiatives are consistently associated with more accurate, respectful, and sustainable outcomes. Overall, the findings indicate that artificial intelligence is most effective when used as a complementary tool within community-centred language preservation frameworks. AI enhances documentation and revitalisation efforts but cannot replace human expertise, cultural knowledge, or social commitment to language maintenance. Sustainable preservation depends on the ethical integration of technology with traditional linguistic practices and the active participation of speech communities.

## Conclusion

The conclusion of this study underscores that artificial intelligence offers valuable support for language maintenance by enhancing documentation, archiving, and revitalisation efforts for endangered and low-resource languages. AI-driven tools improve efficiency, accessibility, and scalability, enabling linguists and communities to preserve linguistic knowledge that might otherwise be lost. However, the study also confirms that technology alone cannot sustain a language. Human expertise, cultural understanding, and active community participation remain essential to ensuring that preserved languages retain their authenticity and social relevance.

The study further concludes that ethical and community-centred implementation is critical to the successful use of AI in language preservation. Issues such as data scarcity, cultural ownership, and algorithmic bias must be carefully addressed to avoid misrepresentation and exclusion. When AI is applied as a collaborative and supportive resource rather than a controlling force, it can strengthen traditional preservation practices and empower speech communities. Ultimately, the integration of future technology with cultural responsibility offers a sustainable path for preserving linguistic heritage for generations to come**.**

## References

Ahmed, R., Khan, T., & Ali, S. (2022). Data sovereignty and ethical considerations in AI-driven language archives. Asian Journal of Applied Linguistics, 9(4), 210–229.

Al Battashi, A., Al Busaidi, S., & Al Musalami, R. (2026). Generative artificial intelligence and informal language learning environments. ReCALL, 38(1), 1–15.

Archive of the Indigenous Languages of Latin America. (2025). AILLA digital language archive. University of Texas at Austin.

Astigarraga, A., Etxeberria, M., & Larrinaga, M. (2025). Neural language models for Pyrenean minority languages: The LINGUATEC-IA project. Journal of Multilingual and Multicultural Development, 46(2), 145–162.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 610–623.

Bird, S., Chiang, D., & Petrov, S. (2014). Machine translation for language preservation. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 125–135.

Bourdieu, P. (1991). Language and symbolic power. Harvard University Press.

Butt, M., Riaz, S., & Hussain, A. (2025). Digital linguistic justice and AI-mediated revitalisation in South Asia. Journal of Language Policy and Planning, 19(1), 33–52.

Chang, Y., Li, Z., & Wang, H. (2025). Hybrid neural machine translation architectures for low-resource languages. IEEE Transactions on Artificial Intelligence, 6(1), 98–112.

Ee, S., & Jailani, R. (2025). Generative AI for enriching linguistic archive metadata. Digital Humanities Quarterly, 19(2), 1–18.

Fishman, J. A. (1991). Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages. Multilingual Matters.

Frontiers. (2025). Revitalising the Limola language through symbolic capital and ecological frameworks. Frontiers in Sociology, 10, 1–14.

Hamza, K. (2025). Community-centred AI models for endangered language preservation. Journal of Language Documentation and Conservation, 19, 55–74.

IAFOR. (2024). Language policy, technology, and indigenous rights in the Global South. International Academic Forum.

Japan Bilingual Publishing Co. (2025). Technology-assisted bilingual education and language preservation. Tokyo.

Johnson, M. (2019). Artificial intelligence and endangered languages: New tools for documentation and revitalisation. Language Documentation and Conservation, 13, 1–24.

Johnson, M., Di Paolo, D., & Lareau, F. (2018). Speech recognition for endangered and low-resource languages. Language Documentation and Conservation, 12, 1–29.

Koc, E. (2025). Ethical risks of generative AI in minority language preservation. AI and Society, 40(1), 87–101.

Kumar, A., & Suresh, K. (2021). Artificial intelligence approaches to documenting low-resource languages. Journal of Language Technology and Linguistics, 6(2), 87–104.

Li, Y., Zhang, Q., & Chen, X. (2024). Real-time speech recognition for endangered languages. Computer Speech and Language, 83, 101531.

McCarty, T. L. (2018). Language policy and planning in Native America. Multilingual Matters.

MDPI. (2025). Agentic AI systems and large language models in language restoration. Applied Sciences, 15(3), 1120.

O'Regan, J. (2022). Language revitalisation in the digital age: Limits of technological intervention. Journal of Sociolinguistics, 26(4), 467–484.

Pareek, S. (2025). Artificial intelligence and sustainable digital archiving for endangered languages. International Journal of Digital Linguistics, 4(1), 22–39.

PARADISEC. (2025). Pacific and regional archive for digital sources in endangered cultures. University of Sydney.

Ray, A., Ahmed, R., & Khan, T. (2024). AI-assisted language documentation and preservation: A qualitative review. Journal of Artificial Intelligence and Humanities, 2(1), 1–20.

Shaumiwaty, S., Razak, R., & Al-Naimi, A. (2025). Cultural ownership and ethics in AI-based language documentation. Journal of Cultural Informatics, 7(2), 65–81.

Soren, M. (2025). Linguistic extinction and cultural responsibility in the digital age. Language and Society Review, 14(1), 10–27.

Soylu, A., & Şahin, G. (2024). AI-powered language learning tools for revitalisation of minority languages. Computer Assisted Language Learning, 37(3), 389–407.

Taylor & Francis. (2025). Intellectual property and indigenous language data in AI systems. Information, Communication and Society, 28(4), 601–617.

UNESCO. (2025). Sustainable digital preservation of endangered languages. UNESCO Publishing.

Vo, H. N. K., Le, D. D., Phan, T. M. D., Nguyen, T. S., Nguyen, Q. D., Tran, N. O., & Vo, T. M. H. (2024). Revitalising the Bahnar language through neural machine translation. Proceedings of the AAAI Conference on Artificial Intelligence, 38, 1–9.

Yang, J., Ma, X., & Vosoughi, S. (2025). Large language models for low-resource language education and revitalisation. Transactions of the Association for Computational Linguistics, 13, 420–437.