International Journal of Language, Linguistics, Literature and Culture
Vol. 04, No. 01; 2025
ISSN: 2583-6560

REVITALIZATION OR RECLAMATION? REFRAMING THE RECOVERY OF INDIGENOUS LANGUAGES IN LATIN AMERICA: A HISTORICAL AND AI-DRIVEN APPROACH

Dianala M. Bernard*
University of the People, Pasadena, California, United States
ORCiD: https://orcid.org/0000-0003-3643-0435
Maren A. Benn
Independent Researcher, Ypsilanti, Michigan, United States

https://doi.org/10.59009/ijlllc.2025.0103

ABSTRACT
Indigenous languages of Latin America have faced significant decline due to colonization, globalization, and sociopolitical factors. While some languages remain endangered, others have entirely disappeared, leaving behind limited historical records or, in some cases, none at all. This study explores the historical transmission of these languages, the current state of documentation, and the role of artificial intelligence (AI) in their recovery, including revitalization and reclamation. Focusing on endangered languages such as Bribri, Cabécar, Maléku, Ngäbere, and Kuna, alongside extinct languages such as Muisca, Cumanagoto, Lenca, Charrúa, and Puelche, this research examines how AI-driven natural language processing (NLP), Optical Character Recognition (OCR), and Text-to-Speech (TTS) synthesis contribute to Indigenous language reconstruction and learning. Furthermore, the study explores the emerging role of text-to-video AI technologies, which can generate immersive audiovisual learning materials to facilitate oral language transmission, contextualize linguistic structures, and support culturally embedded storytelling practices. In employing a qualitative historical analysis combined with digital linguistics, this research highlights AI's potential to bridge critical language gaps, develop culturally relevant teaching materials, and enhance Indigenous-led language recovery initiatives in Latin America.

Keywords: Endangered Languages, Extinct Languages, Latin America, Language Revitalization, Language Reclamation, Artificial Intelligence.

1. INTRODUCTION
Indigenous languages are critical to cultural identity, history, and knowledge transmission (Shen, 2024; UNESCO, 2022). According to research by the World Bank (2023), many Indigenous Peoples continue to speak distinct languages, accounting for over 4,000 of the world's 7,000 languages. However, approximately 60% of these languages have become critically endangered and facing extinction by the year 2100 due to historical and contemporary pressures, including forced assimilation, language suppression policies, and the dominance of colonial languages (Bernard, 2024; Carpenter et al., 2024; Cruz, 2021; Mendecka, 2023; World Bank, 2019, 2023). A similar phenomenon can be observed with Mekatelyu, an English Creole language spoken in Port Limón, Costa Rica, which has experienced a language shift due to sociolinguistic stigma, governmental policies, and the dominance of Spanish (Benn, 2010), which parallels external pressures experienced by Indigenous languages in Latin America, leading to linguistic erosion and community-wide language loss.

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Despite increasing awareness and revitalization and reclamation efforts, many Indigenous languages continue to face the threat of disappearance (Rahima, 2024). This ongoing linguistic crisis has sparked debates over the appropriate terminology to describe efforts to sustain these languages. Various scholars use *language revitalization* and *language reclamation* to describe efforts to reverse these losses (Cruz, 2021; Mendecka, 2023). While revitalization and reclamation are commonly used, this study employs the broader term *language recovery* to encompass both processes, reflecting the diversity of efforts needed to sustain Indigenous languages (Ajani et al., 2024; Henne-Ochoa, 2020; Wang & Bai, 2024). Indigenous languages are inseparable from the traditions, histories, and environmental knowledge they encode, serving as vessels for worldviews, ecological wisdom, and intergenerational cultural continuity. Their recovery represents the revitalization and reclamation of Indigenous linguistic diversity and the strengthening of critical knowledge systems (Barkaskasi & Gladwin, 2021; Ciucci, 2021; Gautam, 2019; Gonçalves, 2024; Redvers et al., 2023).

The World Bank's 2019 infographic, *Lenguas Indígenas: Legado en Extinción*, highlights the urgent state of Indigenous languages in Latin America and the Caribbean. The region is home to approximately 42 million Indigenous Peoples who collectively speak around 560 Indigenous languages, or 14% relative to the world's Indigenous languages. However, language loss remains a growing concern, with 20% of Indigenous populations losing their native languages in recent decades. Additionally, about 26% of these languages risk disappearing entirely. AI has successfully aided in language translation, documentation, and linguistic pattern recognition for some under-resourced languages (Carpenter et al., 2024; Chiruzzo et al., 2024; Prieto et al., 2024a, b; Sánchez-Martínez et al., 2024).

While traditional recovery methods rely on oral transmission and documentation, AI advancements have opened new possibilities for language preservation and reconstruction (Pinhanez et al., 2024). Beyond text-based AI approaches, text-to-video AI technologies now offer immersive, audiovisual solutions that can visually depict phonetic articulation, traditional storytelling, and nonverbal linguistic cues such as gestures, expressions, and cultural contexts (Rojas & Martínez-Cano, 2024; Longpre et al., 2024). These advancements can support Indigenous-led revitalization efforts by generating culturally authentic educational materials that reinforce oral traditions and interactive language learning, bridging the gap between historical documentation and active language use (Ghildyal et al., 2024).

Research Objectives

Accordingly, this study seeks to investigate the potential of AI in Indigenous language recovery in the Latin American region by addressing the historical, technological, and socio-linguistic dimensions of language loss and recovery (Carpenter et al., 2024; Ermolova et al., 2024; Srivastava & Upadhyay, 2024; Wang, 2024). Specifically, the study aims to:

1. Analyze the historical transmission methods of lost Indigenous languages.
2. Assess the availability of documentation and learning resources for these languages.
3. Explore AI's potential in reconstructing extinct languages and supporting endangered ones.

The first objective focuses on analyzing the historical transmission methods of lost Indigenous languages, as traditionally, Indigenous languages have been preserved through oral traditions, intergenerational storytelling, ceremonial use, and community-based education (Bernard, 2024; Wang & Bai, 2024). However, colonization, forced assimilation, and globalization have significantly disrupted these transmission pathways (Benn, 2010; Shen, 2024). Examining how these languages were historically passed down provides a foundation for understanding the mechanisms contributing to linguistic continuity and decline. Additionally, understanding past

Vol. 04, No. 01; 2025
ISSN: 2583-6560

transmission methods will inform AI-driven language recovery strategies, ensuring that new models align with traditional Indigenous ways of learning and knowledge-sharing (Patil, 2024; Pinhanez et al., 2024; Srivastava & Upadhyay, 2024; Sujatna et al., 2024; Tennell & Chew, 2024).

The second objective assesses the availability of documentation and learning resources for these languages. Many Indigenous languages lack comprehensive documentation, particularly extinct languages such as Muisca, where primary sources are limited to colonial texts and lexicons compiled by early European observers (Avila, 2015; Chiruzzo et al., 2024; Pinhanez et al., 2024; Prieto et al., 2024a, b). This study focuses on the extent, quality, and accessibility of linguistic resources, including written texts, grammatical frameworks, dictionaries, audio recordings, and linguistic reconstructions. In identifying gaps in existing documentation, the study will inform AI models on how best to supplement these resources through computational linguistics and data synthesis (Gartlehner et al., 2023; Ghanim, 2024; Giner-Miguelez et al., 2024; Semrl et al., 2023; Ziems et al., 2023).

The third objective explores AI's potential in reconstructing extinct languages and supporting endangered ones. AI has demonstrated machine translation, TTS synthesis, and NLP capabilities for low-resource languages (Chiruzzo et al., 2024; Prieto et al., 2024a, b). In leveraging AI to reconstruct extinct languages such as Muisca, this study examines how deep learning models, phonetic pattern recognition, and comparative linguistic algorithms can help rebuild lost grammar structures and vocabulary. For endangered languages, AI-generated video content can create visual representations of spoken language, reinforcing oral transmission through immersive, interactive materials (Longpre et al., 2024). These tools can simulate cultural storytelling, pronunciation guides, and conversational contexts, supporting Indigenous educators and language learners with dynamic, multimedia learning environments (Rojas & Martínez-Cano, 2024). Additionally, this study considers the ethical implications of AI-driven language reconstruction, ensuring that Indigenous communities maintain control over their linguistic and cultural heritage (Pinhanez et al., 2023).

The findings of this study have significant implications for linguists, educators, AI researchers, and Indigenous communities. For linguists, the study provides insights into how historical transmission patterns influence contemporary language survival. For educators and policymakers, it offers a framework for integrating AI into Indigenous language education programs. AI researchers will benefit from understanding AI models for under-documented languages, contributing to ethical and decolonial approaches in computational linguistics. For Indigenous communities, these insights can empower local efforts to revitalize, reclaim, and sustain linguistic heritage, offering tools and strategies to preserve their cultural identity in the face of modern challenges (Carpenter et al., 2024). Accordingly, this study seeks to address the following research questions:

- How were Indigenous languages of Latin America historically transmitted?
- What learning resources and documentation currently exist?
- How can AI-driven linguistic models help reconstruct extinct languages and support endangered languages of Latin America?

2. THEORETICAL FRAMEWORK

The present study is grounded in three interconnected theoretical perspectives—Linguistic Anthropology (Ahearn, 2021) and Cultural Transmission (Trueba, 2022) Theories, Digital Humanities (Luhmann & Burghardt, 2022) and AI in Linguistics (Groenewald et al., 2024), and Decolonial Approaches to Language Revitalization and Reclamation (Akumbu et al., 2024; Guerrettaz & Engman, 2023)—each contributing a unique yet complementary lens to

Vol. 04, No. 01; 2025
ISSN: 2583-6560

understanding the complexities of Indigenous language decline and the role that AI can play in their recovery. The combined framework allows for a context-aware application of AI in reconstructing extinct languages while supporting endangered languages in ways that respect Indigenous traditions, cultural autonomy, and knowledge transmission (Arul Dayanand et al., 2023; Ray et al., 2024). This interdisciplinary approach is essential for bridging linguistic heritage with modern AI tools in a sustainable, inclusive, and community-driven manner (Arul Dayanand et al., 2023; Campbell, 2024; Pradhan & Dey, 2023; Ray et al., 2024).

2.1 Linguistic Anthropology & Cultural Transmission Theories

Linguistic anthropology examines how language functions within cultural and social contexts, particularly in transmitting linguistic knowledge across generations (Ahearn, 2021; Trueba, 2022). Thus, intergenerational language transmission, influenced by parental proficiency, emotional attachment, and cultural identity, is essential for preserving linguistic diversity, especially for endangered languages. Cultural Transmission Theory explains how traditions and norms are shared through socialization, enculturation, vertical (parent-to-child), and horizontal (peer-to-peer) transmission. Researchers employ ethnographic methods and archival studies to understand language use in daily life. Hymes (1974) and Hymes, as cited in Sabrina et al. (2023), provide that Ethnography of Communication is a foundational framework for analyzing the sociocultural mechanisms through which Indigenous languages have been passed down, including oral storytelling, ritualistic practices, and community-based learning.

Additionally, Joshua Fishman's (1991) Reversing Language Shift (RLS) model is instrumental in understanding endangered languages' decline and potential recovery. Soylu and Şahin (2024) agree, emphasizing intergenerational transmission as the most critical factor in language survival and offering insights into how AI-driven recovery efforts must integrate community-centered approaches rather than focusing solely on technological interventions. Thus, linguistic anthropology and cultural transmission theories provide valuable information about how language and culture are intertwined and passed down through generations, contributing to our understanding of human communication and social dynamics.

2.2 Digital Humanities and AI in Linguistics

The digital humanities perspective integrates computational tools and AI technologies to advance language documentation, preservation, and reconstruction, particularly for endangered languages (Bird, 2020; Ray et al., 2024). Computational linguistics and NLP play a critical role in automating annotation, creating digital corpora, and developing lexical databases, significantly improving the efficiency of language documentation (Arul Dayanand et al., 2023; Wang, 2024). AI-powered tools further support language preservation through speech-to-translation alignment, automated transcription, and machine translation, enhancing accessibility and comprehension of lesser-known languages (Prieto et al., 2024a, b). These advancements enable researchers to document and study languages more effectively, even in cases where resources and fluent speakers are scarce (ELA, n.d.-b; Gessler & von der Wense, 2024).

Digital humanities initiatives further contribute to language revitalization and reclamation through digital archives and community engagement. Archives such as the DoBeS archive at the Max-Planck-Institute (MPI) store valuable linguistic data, offering multimedia corpora for researchers and language communities (Arul Dayanand et al., 2023; MPI, 2025). Community-based digital tools, like the Kubishi Suite, developed for the Owens Valley Paiute language, provide online dictionaries and translation systems to support language learning (Coleman et al., 2024; Kubishi Research Group, 2024). Similar resources are also available with the Open

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Language Archives Community (OLAC), a collaborative network of institutions and individuals dedicated to collecting and sharing resources related to the world's languages, providing access to linguistic materials, including primary texts, lexical resources, and language descriptions, through its online databases, supporting linguistic research, language preservation, and recovery efforts (OLAC, 2025a).

Additionally, the Archive of the Indigenous Languages of Latin America (AILLA, 2025) is a digital repository preserving recordings and texts in Indigenous languages across the region. Its collection includes audio and video materials in various genres, often transcribed and translated into Spanish, English, or Portuguese, and extensive language documentation resources. AILLA also supports bilingual education and language revitalization through teaching materials and is a joint initiative of LLILAS Benson Latin American Studies, the Department of Linguistics, and the University of Texas at Austin's Digital Library Services Division (AILLA, 2025). These tools complement traditional recovery and preservation methods, making language education more accessible to younger generations and ensuring intergenerational transmission (ELA, n.d.-b; Ermolova et al., 2024).

2.3 Decolonial Approaches to Indigenous Language Recovery

Decolonial theories challenge Eurocentric models of language documentation and advocate for Indigenous knowledge sovereignty in revitalization or reclamation efforts (Akumbu et al., 2024; Bird, 2020; Carpenter et al., 2024; Cruz, 2021). In Costa Rica, speakers of Mekatelyu have historically faced discrimination, reinforcing the connection between linguistic prestige and sociopolitical power structures (Benn, 2010). Linda Tuhiwai Smith's (2012) emphasis on Indigenous epistemologies, self-determination, and cultural continuity in language revitalization remains a foundational perspective. Furthermore, scholars argue for reclamation rather than revitalization, yet both processes fall within the broader framework of language recovery, which acknowledges historical disruptions and ongoing efforts to sustain Indigenous languages (Wang & Bai, 2024). Indigenous language reclamation prioritizes relational epistemologies, recognizing language as a means of communication and nurturing community relationships (Henne-Ochoa, 2020). This approach acknowledges the broader cultural and identity reclamation efforts required to support linguistic survival, reinforcing that Indigenous languages are inseparable from the traditions, histories, and environmental knowledge they encode (Ajani et al., 2024).

Self-determination also plays a crucial role in language reclamation, with successful programs centering on community needs and goals to foster intergenerational transmission (Noels et al., 2020). Effective language programs avoid overgeneralization and are instead tailored to the unique linguistic and cultural characteristics of each Indigenous group (Guerrettaz & Engman, 2023). Many of these initiatives incorporate Indigenous curricula and holistic pedagogies to decolonize education and immerse learners in their ancestral knowledge systems (Barkaskasi & Gladwin, 2021). Additionally, Indigenous languages encode crucial ecological knowledge, playing a vital role in environmental stewardship and biodiversity conservation (Gautam, 2019; Gonçalves, 2024; Redvers et al., 2023; UNESCO, 2021).

3. LITERATURE REVIEW

3.1 Historical Transmission of Indigenous Languages

Indigenous languages in Latin America have historically been passed down through oral traditions, ceremonial practices, and community-based education, serving as vital conduits for cultural knowledge, intergenerational wisdom, and social identity (Ahearn, 2021). Similarly, Mekatelyu has remained primarily an oral language in the Caribbean region of Costa Rica, with

Vol. 04, No. 01; 2025
ISSN: 2583-6560

limited written documentation contributing to its vulnerable status as younger generations shift to Spanish (Benn, 2010). Unlike written languages that rely on texts for documentation, Indigenous languages thrived through storytelling, song, and religious rituals, reinforcing linguistic continuity and cultural cohesion across generations (Bradley & Bradley, 2019). These oral traditions were deeply embedded in everyday life, with narratives serving as means of entertainment, historical memory, moral instruction, and cosmological knowledge (Bernard, 2024). Songs and chants carried spiritual and social significance, accompanying ceremonies, rites of passage, and communal gatherings where rhythmic and melodic structures facilitated memorization and retention of linguistic expressions (Hymes, 1974; Sabrina et al., 2023).

In the Andes and Western Amazonia, Indigenous language survival is deeply linked to social organization, territoriality, and land relationships, highlighting the inextricable connection between linguistic transmission and Indigenous sovereignty (Campbell, 2024). Additionally, Indigenous education systems were community-centered and experiential, with language learning occurring through participation in daily activities, storytelling circles, and mentorship by elders (Bernard, 2024). Children acquired linguistic competence by observing, listening, and engaging in structured oral exchanges, reinforcing language's social and relational functions within kinship networks and broader tribal affiliations (Henne-Ochoa et al., 2020). In some communities, linguistic expressions were reserved for sacred contexts, such as religious ceremonies, legal proceedings, and diplomatic exchanges, demonstrating the deep connection between language, governance, and Indigenous legal traditions (Bradley & Bradley, 2019).

However, the forced introduction of Western colonial education models and linguistic hierarchies disrupted these traditional knowledge systems, undermining Indigenous epistemologies (Sallabank & Austin, 2023). Colonial policies, including, in some cases, the removal of Indigenous children to boarding schools, disrupted intergenerational linguistic transmission and inflicted profound psychological trauma, necessitating language reclamation efforts that incorporate healing and community well-being (Henne-Ochoa et al., 2020). Imposing Spanish, Portuguese, English, and French through colonial education systems led to widespread language suppression and cultural erosion (Benn, 2010; Bernard, 2024).

Today, the legacy of colonial linguistic suppression remains evident (Wang & Bai, 2024). However, resistance to linguistic erasure persists through community-driven revitalization and reclamation efforts, youth engagement, and transnational Indigenous movements (UNESCO, n.d.). The International Decade of Indigenous Languages (2022-2032), proclaimed as a key outcome of the 2019 International Year of Indigenous Languages and announced in 2021, represents an opportunity to mobilize global policy interventions, educational reforms, and linguistic rights initiatives (UNESCO, n.d.; 2021). The intent is to safeguard these languages and their cultural heritage by promoting policies, educational initiatives, and community-led revitalization and reclamation programs. Indigenous languages are integral to preserving traditional knowledge, biodiversity conservation, and cultural identity (Gautam, 2019; Gonçalves, 2024; Redvers et al., 2023; UNESCO, 2021).

3.2 AI & Language Documentation

AI has become essential in preserving and analyzing under-documented languages (Ermolova et al., 2024; Ray et al., 2024). Existing AI-powered technologies, such as NLP, OCR, and TTS, have proven highly effective in digitizing, analyzing, and reconstructing endangered and extinct languages (Ghanim, 2024; Pinhanez et al., 2024). NLP is a critical AI-driven tool for Indigenous language research, enabling machines to analyze, process, and understand human language. NLP can play a key role in reconstructing syntax and grammar for extinct languages

Vol. 04, No. 01; 2025
ISSN: 2583-6560

by drawing upon machine learning algorithms trained on related linguistic structures (Coleman et al., 2024). Furthermore, NLP-powered speech recognition models are increasingly being developed to transcribe and analyze spoken Indigenous languages, preserving oral histories and ensuring their long-term accessibility (Tonja et al., 2024).

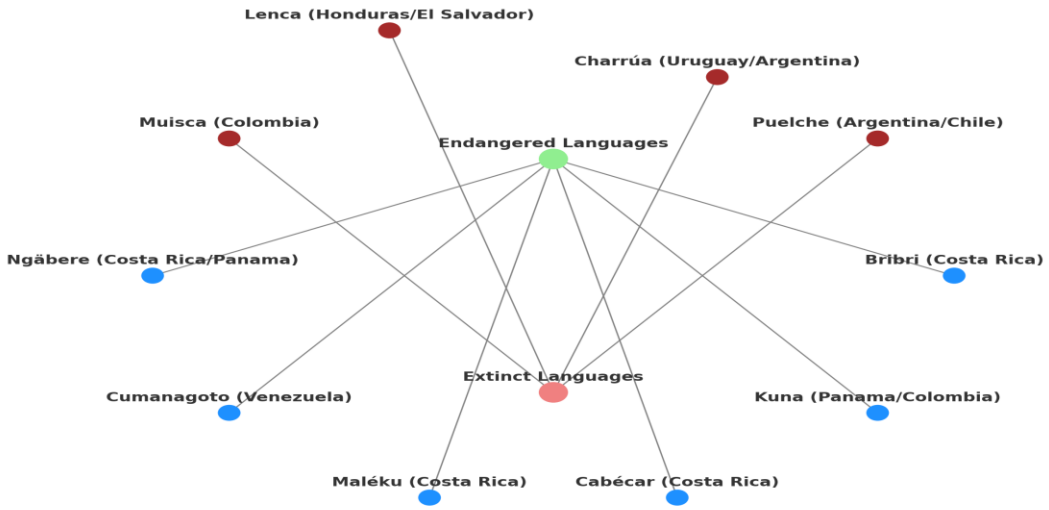
Another promising AI application for Indigenous language recovery is OCR, which converts scanned images of historical texts, handwritten manuscripts, and printed documents into machine-readable text (Agarwal & Anastasopoulos, 2024). For extinct languages such as Muisca, where written documentation is limited to colonial-era lexicons and missionary records, OCR provides a means of recovering and digitizing linguistic data for further analysis (Sánchez Carrera et al., 2024). Lastly, TTS is another transformative AI technology supporting Indigenous language learning and preservation (Pine et al., 2022). AI-powered speech synthesis models can create audio-based dictionaries, pronunciation guides, and virtual language tutors, providing valuable resources for individuals and educational institutions engaged in language revitalization and reclamation (Pinhanez et al., 2024).

3.3 Endangered and Extinct Languages

This study examines a selection of endangered and extinct Indigenous languages in Latin America, focusing on their historical transmission, current status, and the potential for AI-driven linguistic reconstruction and preservation. Endangered languages, such as Bribri, Cabécar, Maléku, Ngäbere, and Kuna, face declining intergenerational transmission due to language shift, requiring AI-powered educational tools, digital documentation, and speech synthesis models to support preservation (World Bank, 2019). Extinct languages, including Muisca, Cumanagoto, Lenca, Charrúa, and Puelche, no longer have fluent speakers but retain varying historical documentation (Ancient Origins, 2016; Avila, 2015).

Muisca is a prime candidate for AI reconstruction, as colonial-era lexicons and their connection to Chibchan languages provide a foundation for NLP-driven grammatical modeling and phonetic reconstruction (Avila, 2015; OLAC, 2025d). Other extinct languages present more significant challenges due to limited records. However, AI-based pattern recognition and computational linguistics promise partial revival, accentuating the intersection of technology and Indigenous language recovery (Prieto et al., 2024a, b).

Figure 1 Network of Endangered and Extinct Languages in Latin America



Vol. 04, No. 01; 2025
ISSN: 2583-6560

Note. Network visualization of some endangered and extinct languages in Latin America.

3.4 Endangered Languages (Languages Still Spoken)

Bribri (Costa Rica)

Bribri is a tribal name from a word meaning *mountains* in their language. The Bribri language is also called *Su Uhtuk*, which means *our language* (Native Languages, 2020). Bribri, a Chibchan language spoken by approximately 6,000 to 11,000 people in the Talamanca region along the Lari, Telire, and Uren rivers in the Port Limón Province of Costa Rica, as well as in the Buenos Aires canton of the Puntarenas Province, remains one of the country's most resilient Indigenous languages, though it is considered vulnerable or threatened due to declining speaker numbers and intergenerational transmission challenges (ELA, n.d.-a, n.d.-b; ELP, n.d.-a). Bilingual education policies and the increasing use of Spanish among younger generations have weakened its intergenerational transmission (ELA, n.d.-a, n.d.-b; ELP, n.d.-a).

As Spanish became the dominant language in schools, workplaces, and media, fewer Bribri speakers passed it down to their children. Additionally, a lack of formal literacy materials in Bribri has made it difficult for new learners to acquire the language in written form (Sánchez Avendaño & Angulo-Jiménez, 2023). A parallel linguistic blending is evident in Mekatelyu, which evolved through the contact of English-based Creoles with Spanish in the Port Limón province (Benn, 2010). Without sustained revitalization and reclamation efforts, Bribri may continue to decline, necessitating digital preservation strategies and AI-supported learning tools (Coto-Solano, 2021; Feldman & Coto-Solano, 2020; Karson & Coto-Solano, 2024). Resources from OLAC can support linguistic research, language preservation, and revitalization efforts (OLAC 2025b; 2025d).

Cabécar (Costa Rica)

Closely related to Bribri, Cabécar is spoken by approximately 9,000 people in Costa Rica, primarily in the Turrialba region (ELP, n.d.-b). The name Cabécar comes from a tribal name, *Kabekada*. The people also refer to themselves as *Ditsä Si* (meaning *the real people*) and to their language as *Sektu* (*our language*) (Native Languages, 2020). Cabécar is still primarily oral, relying on storytelling, songs, and spoken traditions for transmission (ELP, n.d.-b; Sánchez Avendaño & Angulo-Jiménez, 2023). Notably, Cabécar is the only language in Costa Rica with monolingual speakers, predominantly women, allowing it to preserve traditional grammar and phonetics (OLAC, 2025c). However, their isolation has resulted in limited written documentation and literacy resources, making it challenging for younger generations to learn the language formally (Coto-Solano et al., 2024; ELP, n.d.-b). Additionally, as Spanish education systems expand, younger speakers increasingly struggle to maintain fluency in Cabécar. Thus, the lack of structured educational materials and technological learning tools further complicates revitalization and reclamation efforts (ELP, n.d.-b; Native Languages, 2020).

Maléku (Costa Rica)

Spoken by Indigenous groups in northern Costa Rica, Maléku is critically endangered, with fewer than 1,000 active speakers (ELP, n.d.-c). The leading cause of its decline is the shift toward Spanish, which is the dominant language in education, business, and government. As younger generations prioritize Spanish for socioeconomic mobility, fewer children learn Maléku as their first language, leading to an accelerated decline in intergenerational transmission (Sánchez Avendaño & Angulo-Jiménez, 2023). Additionally, formal linguistic documentation of Maléku is scarce, making it challenging to integrate into modern education

Vol. 04, No. 01; 2025
ISSN: 2583-6560

systems or develop structured literacy programs (Coto-Solano et al., 2024). Limited access to written materials, teacher training, and digital learning platforms further exacerbates the language's vulnerability. Without immediate intervention through language preservation programs, community-led revitalization and reclamation efforts, and AI-supported digital tools such as NLP-powered text generation and TTS synthesis, Maléku risks becoming dormant within a generation (Pinhanez et al., 2024).

Ngäbere (Costa Rica and Panama)

Ngäbere, spoken by the Ngäbe-Buglé people, has a more extensive speaker base than many other Indigenous languages in Costa Rica and Panama. Despite this, Spanish continues encroaching on traditional Ngäbere-speaking communities, particularly in the education and economic sectors (Solano Acuña, 2021). Younger generations increasingly choose Spanish over Ngäbere due to societal pressures and a lack of institutional support for Indigenous language learning (Guerrettaz & Engman, 2023). Furthermore, there are few digital resources for learning Ngäbere, limiting its accessibility outside of oral transmission within Indigenous communities (OLAC, 2025h). There is a living dictionary specific to the Ngäbere language, however, which should be explored (Living Dictionary, n.d.). The absence of structured educational materials further hinders language retention among younger speakers. AI-powered speech recognition and language-learning applications could play a crucial role in preventing further decline by providing accessible language-learning tools and revitalization and reclamation strategies (Pinhanez et al., 2024).

Kuna (Panama and Colombia)

The Kuna language remains relatively stable but faces challenges related to digital representation and literacy development. Kuna is still spoken among Indigenous communities in Panama and Colombia, yet younger generations are increasingly drawn toward Spanish due to educational and economic integration (Sherzer, 2009). The lack of modern literacy tools, digital dictionaries, and formal learning materials limits the ability of Kuna speakers to maintain proficiency in written contexts (Guerrettaz & Engman, 2023). Additionally, the absence of AI-powered speech recognition and NLP-based transcription tools further hinders Kuna's adaptation to digital platforms, restricting its accessibility beyond oral traditions. AI-driven linguistic documentation and TTS applications could support its long-term survival by expanding digital access to learning resources and creating interactive language-learning models using OLAC resources, for example (OLAC, 2025f; Pinhanez et al., 2024).

3.5 Extinct Languages (Disappeared Languages)

Muisca (Colombia) – A Prime Candidate for AI-Driven Recovery

Muisca, once spoken by the Muisca Confederation in central Colombia, became extinct due to Spanish colonization and language suppression policies (Avila, 2015; OLAC, 2025d). Unlike many extinct Indigenous languages, Muisca has a relatively strong linguistic record, with colonial-era dictionaries, grammar structures, and comparative Chibchan studies providing valuable documentation (Avila, 2015; Sánchez Carrera et al., 2024). These historical texts offer a foundation for AI-driven reconstruction, utilizing NLP, phonetic modeling, and machine learning algorithms to reconstruct lost grammatical structures and pronunciation patterns (Prieto et al., 2024a, b). In leveraging computational linguistics and cross-linguistic analysis with related Chibchan languages, AI technologies could significantly contribute to the recovery of Muisca, making it more accessible for linguistic and cultural restoration efforts.

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Cumanagoto (Venezuela)

Cumanagoto, an extinct Cariban language, was once spoken along Venezuela's Caribbean coast. The language disappeared following Spanish conquest and forced displacement, losing its intergenerational transmission (Ancient Origins, 2016). Only limited linguistic documentation remains, primarily in fragmented word lists and missionary records, which provide an incomplete picture of its grammatical structure and phonetic system (OLAC, 2025e). However, AI-driven phonetic analysis can compare Cumanagoto with related Cariban languages like Pemón and Wayuu to infer missing grammatical structures, phonological patterns, and syntactic rules (Prieto et al., 2024a, b). Additionally, NLP and machine learning models trained on surviving Cariban languages (see OLAC, 2025e) could aid in reconstructing plausible word formations and sentence structures, offering new pathways for linguistic revitalization and reclamation efforts.

Lenca (Honduras and El Salvador)

Lenca, an isolated Indigenous language spoken in Honduras and El Salvador, disappeared due to colonial displacement and the lack of formal transmission systems (Ancient Origins, 2016). Unlike languages in larger linguistic families, Lenca has no known relatives, making reconstruction particularly difficult (Mendecka, 2023). The absence of closely related languages limits comparative linguistic analysis, requiring AI applications to rely on pattern recognition, phonetic modeling, and historical linguistic reconstruction techniques (Prieto et al., 2024a, b). Additionally, computational methods such as NLP-based syntactic inference and machine learning algorithms trained on archival materials could help identify structural patterns in surviving Lenca word lists and historical records (OLAC, 2025g), offering a potential avenue for digital preservation and partial recovery. However, the scarcity of extensive linguistic documentation necessitates further field research and collaboration with local Indigenous communities to validate AI-generated reconstructions and ensure cultural relevance in revitalization and reclamation efforts (Guerrettaz & Engman, 2023).

Charrúa (Uruguay and Argentina)

The Charrúa language, once spoken by Indigenous groups in Uruguay and Argentina, disappeared following mass displacement and genocide during the colonial period (Ancient Origins, 2016). Spanish settlers and military campaigns eradicated much of the Charrúa population, leaving behind only fragmented word lists and oral testimonies. The lack of extensive linguistic documentation makes language recovery—whether through revitalization, reclamation, or reconstruction—challenging. However, AI could assist in data extraction from historical texts, phonetic reconstruction, and the development of speech synthesis models by drawing inferences from related Indigenous languages such as Guaraní and Puelche (Prieto et al., 2024a, b). Additionally, computational linguistics and deep learning models could be employed to analyze phonetic consistency within fragmented records, helping reconstruct possible pronunciation patterns. Given the cultural sensitivity surrounding the Charrúa identity, it is crucial to integrate AI-driven approaches with community-led initiatives to ensure ethical and culturally aligned language revitalization and reclamation efforts (Guerrettaz & Engman, 2023).

Puelche (Argentina and Chile)

Puelche, an isolated language of Argentina and Chile, became extinct due to cultural assimilation and the dominance of Spanish in the region. With few surviving linguistic records, Puelche remains one of the more challenging Indigenous languages to reconstruct. Historical

Vol. 04, No. 01; 2025
ISSN: 2583-6560

accounts contain only fragmented vocabularies and sparse grammatical documentation, limiting traditional linguistic comparative methods. However, AI-driven deep learning models could analyze existing linguistic fragments and search for potential syntactic and phonetic parallels with regional Indigenous languages such as Tehuelche and Mapudungun (Prieto et al., 2024a, b). NLP algorithms could also be leveraged to identify recurring morphological patterns in the limited textual data available (OLAC, 2025i). Given the highly endangered nature of Puelche, community-led language documentation initiatives and AI-enhanced preservation tools could provide a multi-faceted approach to recovery (Guerrettaz & Engman, 2023).

4. METHODOLOGY

4.1 Research Design

This study employs a qualitative historical analysis combined with digital linguistics methodologies to examine Indigenous language transmission, documentation gaps, and the potential role of AI in revitalization or reclamation efforts. The historical analysis investigates how these languages were traditionally passed down, the sociopolitical factors that contributed to their decline, and the extent of their linguistic documentation (Bradley & Bradley, 2019). This method involves critically examining research on archival records, colonial lexicons, missionary texts, and ethnographic studies, which are primary sources for reconstructing extinct languages. Meanwhile, the digital linguistics approach leverages AI-powered tools such as NLP, OCR, and TTS can contribute to Indigenous language recovery, including revitalization and reclamation efforts (Agarwal & Anastasopoulos, 2024).

Additionally, machine learning models trained on related languages could enhance the accuracy of grammatical and phonetic reconstruction by identifying structural similarities in morphosyntax and phonology. In integrating these approaches, this study ensures that AI methodologies remain culturally and linguistically authentic, aligning with Indigenous epistemologies and language sovereignty (Pinhanez et al., 2024). This interdisciplinary framework underscores the ethical implications of AI-driven language recovery, advocating for community-led digital preservation strategies to reinforce Indigenous linguistic and cultural autonomy (Guerrettaz & Engman, 2023).

4.2 Data Collection Methods

This study incorporates literature on archival research, ethnographic sources, and AI-driven computational analysis approaches to assess the potential of Indigenous language documentation and recovery (AILLA, 2025; Kubishi Research Group, 2024; MPI, 2025; OLAC, 2025a).

4.2.1 Archival Research

Archival research focuses on historical grammar structures, dictionaries, and linguistic fieldwork reports, particularly those compiled by missionaries and colonial administrators (Bradley & Bradley, 2019). While these records often reflect colonial biases, they provide valuable linguistic data that AI models can process to identify phonetic, syntactic, and morphological patterns. Additionally, computational linguistics tools can cross-reference linguistic elements with related Indigenous languages to enhance reconstruction accuracy (Prieto et al., 2024a, b).

Vol. 04, No. 01; 2025
ISSN: 2583-6560

4.2.2 Ethnographic Accounts & Oral Histories

Oral histories and ethnographic research are crucial in understanding linguistic change and recovery strategies (Sallabank & Austin, 2023). Benn (2010) employed interviews and surveys to assess the sociolinguistic perceptions of Mekatelyu speakers in Costa Rica, a methodology that can similarly inform Indigenous language revitalization and reclamation efforts. Oral histories and Indigenous knowledge systems also preserve linguistic elements, such as ceremonial speech and storytelling structures, essential for AI-driven modeling (Jafari, 2023). Incorporating community-driven ethnographic research ensures that revitalization and reclamation efforts align with Indigenous epistemologies and reinforce cultural authenticity (Guerrettaz & Engman, 2023).

4.2.3 Digital Linguistics and AI-Based Methods

The digital linguistics approach applies AI-driven tools to extract, analyze, and reconstruct Indigenous languages. OCR technology facilitates the digitization of historical texts, enabling researchers to recover linguistic data from missionary records and colonial lexicons (Sánchez Carrera et al., 2024). Once digitized, NLP algorithms analyze phonological and syntactic patterns, offering insights into possible grammatical structures and vocabulary reconstructions (Romero et al., 2024). Further, phonetic modeling and TTS synthesis contribute to reconstructing pronunciation and spoken forms of extinct languages by training AI on surviving audio data from related Indigenous languages (Romero et al., 2024). Lastly, deep learning techniques support identifying linguistic correlations across fragmented records, enhancing the precision of AI-driven language reconstruction efforts (Pinhanez et al., 2024).

4.3 Data Analysis

This study applies thematic analysis to identify patterns in linguistic transmission, documentation gaps, and AI feasibility in Indigenous language recovery. Thematic analysis is a qualitative method that categorizes recurring linguistic and historical patterns (Braun & Clarke, 2006). In employing this approach, the study systematically examines critical Indigenous language revitalization and reclamation initiatives.

4.3.1 Linguistic Transmission

This study analyzes how Indigenous languages were historically passed down through oral traditions, ceremonial practices, and intergenerational learning (Bradley & Bradley, 2019). These traditional knowledge systems ensured linguistic continuity, embedding cultural narratives and epistemologies within Indigenous communities. However, external pressures, such as colonial suppression and language shift toward dominant national languages, disrupted these transmission pathways.

4.3.2 Documentation Gaps

The study assesses how well Indigenous languages—particularly extinct ones like Muisca—have been recorded in colonial texts, grammar structures, and linguistic fieldwork studies (Avila, 2015; Bradley & Bradley, 2019). Many archival materials contain missionary lexicons and linguistic descriptions that, while biased, offer foundational data for reconstruction. With Muisca, Avila (2015) provides that the language's history was preserved in 16th- and 17th-century texts. It was driven toward extinction by King Charles III of Spain, who banned its use to control the Indigenous population. This ban remained in effect until changes to Colombia's 1991 Constitution. Additionally, computational methods can assist in identifying patterns

Vol. 04, No. 01; 2025
ISSN: 2583-6560

within fragmented linguistic records to enhance documentation accuracy (Prieto et al., 2024a, b).

4.3.3 AI Feasibility in Language Recovery

NLP, phonetic modeling, and speech synthesis contribute to linguistic revival while also addressing challenges such as data limitations, AI biases, and ethical concerns related to Indigenous knowledge sovereignty (Pinhanez et al., 2024). Researchers can use AI-driven techniques to generate context-aware linguistic models that respect Indigenous epistemologies while supporting practical language learning and recovery efforts (Guerrettaz & Engman, 2023). Through this interdisciplinary approach, the study bridges historical linguistic research with AI-driven language recovery, ensuring that technological applications remain culturally appropriate and scientifically rigorous (Jafari, 2023).

Emerging text-to-video AI technologies enhance Indigenous language recovery by generating immersive, audiovisual learning materials depicting real-life conversational contexts, traditional storytelling, and cultural practices. These tools can visually represent phonetic articulation, gestures, and situational language use, making language learning more engaging and accessible for Indigenous communities and language learners (Rojas & Martínez-Cano, 2024). Furthermore, AI-generated video synthesis can provide interactive simulations of real-life dialogues, embedding pronunciation guidance, tonal variations, and socio-cultural nuances into learning materials (Ghildyal et al., 2024). This approach enhances digital immersion, allowing learners to visualize conversational structure, prosody, and nonverbal communication alongside linguistic elements.

Additionally, ongoing research in AI-generated image and video quality prediction highlights the need for high-fidelity audiovisual materials that respect cultural accuracy and pedagogical effectiveness (Ghildyal et al., 2024). Researchers can amplify language revitalization efforts by integrating AI-generated video content with linguistic modeling while preserving the cultural depth embedded in oral traditions. However, ethical considerations, including data provenance, bias mitigation, and AI's role in shaping knowledge representation, must remain at the forefront of these innovations (Ch'ng, 2024).

5. FINDINGS & DISCUSSION

5.1 Historical Analysis of Lost & Endangered Indigenous Languages

The historical transmission of Indigenous languages varied significantly depending on whether they were primarily oral or written, shaping the degree to which they could be preserved, revitalized, and reclaimed in the face of external pressures (Mirdha, 2020). Many Indigenous languages in Latin America, including Muisca, Lenca, and Charrúa, relied primarily on oral traditions. In contrast, some cultures developed writing systems or adapted to colonial scripts, allowing for a degree of textual documentation. The mode of transmission played a critical role in language survival, particularly as colonization, forced assimilation, and socio-political pressures disrupted Indigenous linguistic practices. Understanding the differences between oral and written transmission provides insight into why certain languages were lost entirely, why others remain critically endangered, and how AI-driven tools can aid reconstruction and preservation (Ray et al., 2024).

5.2 Oral Transmission and Language Vulnerability

For many Indigenous groups, language was passed down through oral traditions, including storytelling, ceremonial speech, songs, and communal interactions (Mirdha, 2020). These transmission methods were highly effective in tightly knit communities where knowledge was

Vol. 04, No. 01; 2025
ISSN: 2583-6560

shared across generations. Oral traditions often embed cultural, historical, and spiritual knowledge, reinforcing language retention within social structures. However, because oral languages lacked standardized written records, they were particularly vulnerable to language shift and loss when external forces disrupted intergenerational learning (Arul Dayanand et al., 2023).

The impact of colonization was especially severe on orally transmitted languages. European colonial powers imposed Spanish and Portuguese as dominant languages, erasing Indigenous linguistic traditions through forced assimilation, missionary education, and linguistic suppression policies (Sallabank & Austin, 2023). Missionary schools, for example, prohibited children from speaking their native tongues, replacing Indigenous oral knowledge with Christian catechisms and Western educational models. As Indigenous populations were displaced or reduced due to disease, warfare, and social marginalization, their languages suffered rapid attrition, with many becoming dormant or extinct within a few generations. This phenomenon is evident in languages such as Puelche (Argentina and Chile) and Cumanagoto (Venezuela), which lacked written documentation and disappeared as Indigenous communities were assimilated or displaced (Langdon, 2020).

5.3 Written Language and Preservation Potential

In contrast, Indigenous languages with writing systems or documented records had a higher chance of partial survival or reconstruction. Some civilizations, such as the Maya, Mixtec, and Aztecs, developed sophisticated hieroglyphic or logographic writing systems, while others adapted Latin scripts introduced by Spanish missionaries. Even among primarily oral cultures, colonial-era documentation, such as word lists, grammar structures, and lexicons, played a role in preserving linguistic data that could be used for later study (Avila, 2015; Bradley & Bradley, 2019).

A prime example is Muisca (Colombia), an extinct language from the Chibchan family. Though no longer spoken, Muisca words, grammatical structures, and phonetic information survive in 16th- and 17th-century colonial manuscripts, including early grammar structures and missionary dictionaries (Avila, 2015; Bradley & Bradley, 2019). These texts allow linguists and AI-driven models to analyze patterns, compare with related Chibchan languages, and reconstruct missing elements (Prieto et al., 2024a, b).

5.4 Impact of Colonization on Language Extinction and Endangerment

The decline and extinction of many Indigenous languages are directly tied to colonial policies, forced assimilation, and economic pressures that devalued Indigenous linguistic identity. Colonization led to a language shift, where Indigenous speakers gradually abandoned their native tongues in favor of colonial languages to access education, economic opportunities, and political participation (Dorzheeva, 2021). Over time, Indigenous languages became associated with stigma and social disadvantage, further accelerating their decline. The case of Mekatelyu in Costa Rica also illustrates this phenomenon, where governmental policies favoring Spanish and negative sociolinguistic attitudes have contributed to its decline as a spoken language (Benn, 2010).

5.5 AI and Indigenous Language Recovery: Possibilities and Challenges

AI in Indigenous language recovery offers significant potential but presents complex challenges. AI-driven tools such as NLP, phonetic modeling, and machine learning algorithms have already been successfully applied to low-resource languages, providing new opportunities for both documentation and reconstruction (Ray et al., 2024). One approach involves AI-

Vol. 04, No. 01; 2025
ISSN: 2583-6560

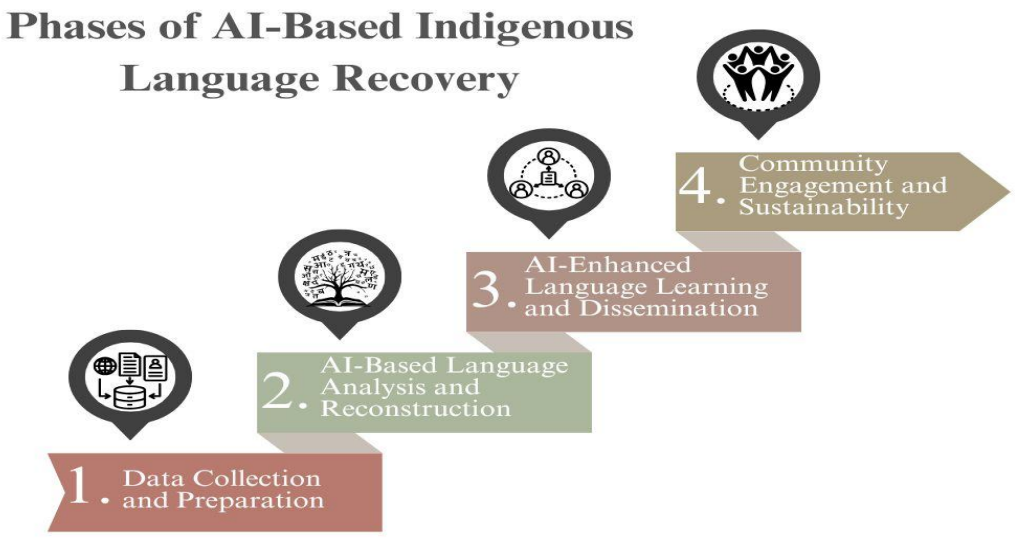
sourced linguistic data extraction from historical texts, utilizing OCR technology to digitize and process archival materials (Agarwal & Anastasopoulos, 2024; Pinhanez et al., 2024). Many of the earliest recorded Muisca words and grammatical structures exist in Spanish-translated manuscripts written by missionaries and colonial officials. In applying machine learning algorithms trained on digitized texts, AI can detect recurring patterns, phonetic variations, and syntactic structures that provide insight into the language's original form (Avila, 2015; Jafari, 2023).

Emerging text-to-video AI technologies introduce an additional layer of support by generating immersive audiovisual learning materials that depict real-life conversational contexts, traditional storytelling, and cultural practices, allowing Indigenous language learners to visualize phonetic articulation, gestures, and situational language use, thereby improving accessibility and engagement (Rojas & Martínez-Cano, 2024). However, challenges remain, including bias in AI training datasets, limited access to high-quality linguistic recordings, and ethical AI governance to ensure that AI-generated linguistic materials reflect Indigenous epistemologies (Pinhanez et al., 2024). While AI presents promising solutions, its successful implementation in Indigenous language recovery requires collaborative efforts between linguists, Indigenous language speakers, AI developers, and policymakers. Ensuring that AI-generated materials are culturally authentic and community-approved is critical to preventing linguistic misrepresentation and reinforcing Indigenous data sovereignty.

6. ROADMAP FOR AI-BASED INDIGENOUS LANGUAGE RECOVERY

Revitalizing or reclaiming Indigenous languages using AI requires a structured, multi-phase approach that combines historical linguistic research, computational analysis, educational dissemination, and community engagement (Jafari, 2023). This roadmap outlines the key phases necessary for AI-driven Indigenous language recovery initiatives, ensuring the process is technologically feasible and culturally sustainable (Pinhanez et al., 2024). Each phase addresses specific challenges related to data collection, linguistic reconstruction, AI-assisted learning, and long-term community integration to facilitate the preservation and potential revival of endangered and extinct Indigenous languages (Souter et al., 2024).

Figure 2 Roadmap for AI-Based Indigenous Language Recovery



Vol. 04, No. 01; 2025
ISSN: 2583-6560

Note. This figure presents a structured, multi-phase roadmap for recovering Indigenous languages through AI. Arrows indicate the progression and interconnectivity between phases, emphasizing the cyclical and community-driven nature of language recovery.

Phase 1: Data Collection and Preparation

Phase 1 focuses on collecting historical data for Indigenous language recovery, supporting revitalization and reclamation as complementary approaches. This phase focuses on digitizing historical records, transcribing oral histories, and creating a structured linguistic corpus that serves as the basis for computational analysis (Pinhanez et al., 2024). Many Indigenous languages, particularly those that are extinct, exist only in fragmented colonial-era texts, ethnographic records, or oral traditions, making their documentation a critical first step (Jafari, 2023).

OCR tools can digitize historical manuscripts and linguistic field notes, converting them into machine-readable formats for further analysis (Agarwal & Anastasopoulos, 2024). Additionally, transcribing oral histories from community elders, linguistic scholars, and Indigenous knowledge keepers ensures that spoken language elements, pronunciation, and storytelling traditions are preserved (Daigneault & Anderson, 2023). Researchers can then develop a robust, AI-compatible linguistic corpus that enables subsequent computational modeling and reconstruction by aggregating and organizing these linguistic resources (Flavelle & Lachler, 2023).

Phase 2: AI-Based Language Analysis and Reconstruction

Once a comprehensive dataset has been assembled, AI-powered linguistic analysis and reconstruction can be initiated. NLP models play a key role in syntax and grammar analysis, helping identify sentence structures, morphological patterns, and linguistic rules from existing texts and comparative Indigenous languages (Pinhanez et al., 2024). In analyzing grammatical similarities with related Indigenous languages, AI can make predictions about missing linguistic elements, helping to fill in the gaps for partially documented or extinct languages like Muisca (Avila, 2015). Phonetic modeling and speech synthesis are applied to reconstruct pronunciation, intonation, and phonological patterns, ensuring linguistic accuracy (Romero et al., 2024). AI-driven phonetic modeling allows for creating synthetic speech models, which can be validated through scholars' comparisons with existing Indigenous languages and linguistic reconstructions (Tonja et al., 2024).

Phase 3: AI-Enhanced Language Learning and Dissemination

Following language reconstruction, the next step is developing accessible AI-powered learning tools to support language dissemination among Indigenous communities, educators, and researchers (Jafari, 2023). AI-powered learning platforms and chatbots can simulate language immersion experiences, enabling learners to practice Indigenous languages interactively (Nanduri & Bonsignore, 2023). TTS technology and AI-generated educational materials can create digital dictionaries, pronunciation guides, and language learning applications tailored for community use and academic study (Pinhanez et al., 2024). These tools bridge computational language models and real-world language adoption, ensuring recovery efforts move beyond linguistic research into practical language acquisition and usage (Daigneault & Anderson, 2023).

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Phase 4: Community Engagement and Sustainability

For AI-driven language recovery to be effective, Indigenous communities must play a central role in validating linguistic outputs, refining AI-generated models, and guiding sustainable implementation efforts (Jafari, 2023). This phase emphasizes crowdsourcing linguistic validation, where Indigenous speakers, scholars, and language advocates review AI-generated linguistic materials to ensure accuracy, cultural appropriateness, and usability (Cooper et al., 2024). Furthermore, long-term sustainability requires integrating Indigenous languages into education policies, school curricula, and cultural initiatives to reinforce language usage across generations (Pradhan & Dey, 2023). Digital language tools and AI models must be designed with Indigenous ownership and governance, ensuring communities retain control over their linguistic data, AI-generated language outputs, and teaching resources (Pinhanez et al., 2023).

7. Implications & Recommendations

Integrating AI into Indigenous language recovery has significant implications for language preservation, policy development, technology accessibility, and interdisciplinary collaboration (Jafari, 2023; Pinhanez et al., 2024). While AI presents unprecedented opportunities for reconstructing and supporting endangered and extinct Indigenous languages, its application must be ethically guided, community-driven, and aligned with Indigenous linguistic and cultural sovereignty (Ray et al., 2024; Srivastava & Upadhyay, 2024). This section outlines key recommendations for policy implementation, technological advancements, interdisciplinary partnerships, and future research directions to ensure that AI-driven language recovery efforts are sustainable, inclusive, and impactful (Arul Dayanand et al., 2023).

7.1 Policy Recommendations: AI Integration in Indigenous Language Education

Governments, educational institutions, and Indigenous advocacy groups must develop comprehensive policies integrating AI-driven tools into Indigenous language education (Jafari, 2023; Pinhanez et al., 2024). National language policies should recognize AI as a supportive resource for Indigenous language revitalization, ensuring that computational tools are accessible to schools, Indigenous educators, and community-led language programs (Pradhan & Dey, 2023). Specific policy recommendations include:

- Incorporating AI-driven Indigenous language learning into national curricula, ensuring students can access AI-powered dictionaries, speech synthesis tools, and interactive chatbots for language learning (Jafari, 2023).
- Funding digital language preservation initiatives that use machine learning, NLP, and phonetic modeling to reconstruct and document Indigenous languages (Pinhanez et al., 2024; Rana, 2024).
- Aligned with decolonial frameworks (see Section 2.3), Indigenous data sovereignty is a core consideration in AI development for language recovery. This study follows ethical AI governance principles to ensure community-led validation (Ajani et al., 2024).
- Creating grants and scholarships for Indigenous scholars and language advocates to engage in AI research and development, ensuring community leadership in AI language projects (Pradhan & Dey, 2023).

7.2 Technology Development: Open-Access Platforms for Endangered Languages

To maximize the impact of AI-driven language recovery, there is an urgent need to strengthen open-access platforms that house Indigenous language datasets, AI-powered linguistic tools, and community-driven learning resources (Arul Dayanand et al., 2023; Daigneault &

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Anderson, 2023). Many Indigenous languages, particularly those endangered or extinct, lack comprehensive linguistic corpora, making their preservation dependent on centralized, accessible, and ethically governed digital platforms (Jafari, 2023; Pinhanez et al., 2024). Key technological recommendations include:

- Developing AI-powered Indigenous language repositories that store digitized linguistic data, colonial texts, speech recordings, and AI-generated learning resources in an open-access format (Arul Dayanand et al., 2023; Daigneault & Anderson, 2023).
- Expanding AI-powered translation and speech synthesis tools, enabling Indigenous speakers and learners to translate and hear Indigenous words, phrases, and stories in real-time (Prieto et al., 2024a, b).
- Creating decentralized, Indigenous-led digital language databases, ensuring that Indigenous communities retain control over AI-generated linguistic materials rather than external institutions (Pinhanez et al., 2024).
- Integrating AI-based Indigenous language tools into mobile applications, allowing widespread accessibility to language learning resources, pronunciation guides, and interactive AI tutors for community members and students (Jafari, 2023).

7.3 Collaborative Efforts: Partnerships Between Linguists, AI Researchers, and Indigenous Communities

Success in AI-driven Indigenous language recovery depends on interdisciplinary collaboration between linguists, AI developers, policymakers, and Indigenous leaders (Flavelle & Lachler, 2023; Jafari, 2023). Without active community involvement, AI-driven projects risk extracting linguistic data without cultural consent or misrepresenting Indigenous languages (Pinhanez et al., 2023). Collaborative partnerships should ensure that Indigenous voices are central in shaping AI language models, validating AI-generated linguistic outputs, and managing digital linguistic resources (Pinhanez et al., 2024).

- Establishing AI-Indigenous research collectives, where Indigenous linguists, AI researchers, and language advocates co-develop computational models for language documentation and reconstruction (Jafari, 2023; Pinhanez et al., 2024).

- Providing training in AI and computational linguistics for Indigenous researchers and language activists, ensuring that AI tools are developed and managed by Indigenous communities rather than external entities (Srivastava & Upadhyay, 2024).

- Creating AI-integrated Indigenous language research centers where scholars and community members can collaborate on NLP models, phonetic analysis, and AI-assisted language immersion programs (Ajani et al., 2024; Pradhan & Dey, 2023).

- Encouraging corporate and institutional responsibility by requiring tech companies and AI research institutions to consult Indigenous communities before implementing AI language projects (Gellman, 2021; Pinhanez et al., 2023).

Through equitable and ethical collaboration, AI-driven language recovery can become a model for community-led digital sovereignty, ensuring that technology is a tool for empowerment rather than linguistic displacement (Lewis et al., 2020).

Vol. 04, No. 01; 2025
ISSN: 2583-6560

7.4 AI and Indigenous Sign Languages: Expanding AI-Driven Recovery to Indigenous Sign Languages

Indigenous sign languages, like spoken Indigenous languages, face endangerment due to historical suppression, lack of documentation, and accessibility barriers (Prieto et al., 2024a, b; Pinhanez et al., 2024). Many Indigenous Deaf communities have developed unique sign languages that are not yet widely recognized or supported by AI-driven linguistic tools. However, AI technologies such as NLP), computer vision, and machine learning offer promising solutions for documenting, preserving, and recovering these visual-gestural languages.

Key recommendations for integrating Indigenous sign languages into AI-based recovery efforts include:

- Developing AI-powered sign language recognition tools that analyze Indigenous sign language gestures, facial expressions, and body movements to aid linguistic preservation and translation (Flavelle & Lachler, 2023; Daigneault & Anderson, 2023).
- Expanding AI-driven video-based learning tools to create digital sign language dictionaries, interactive learning modules, and real-time translation applications for Indigenous sign language users (Jafari, 2023; Pinhanez et al., 2024).
- Ensuring Indigenous sign languages are included in national AI language policies, supporting accessibility initiatives in education, media, and Indigenous governance structures (Pradhan & Dey, 2023).
- Creating community-led Indigenous sign language documentation projects, where Deaf Indigenous speakers collaborate with AI researchers to design culturally appropriate AI models that reflect linguistic and cultural authenticity (Farooq et al., 2021).

Including Indigenous sign languages in AI-driven recovery efforts can help researchers and Indigenous communities ensure that language preservation remains inclusive, addressing the linguistic needs of both spoken and signed language users. Indigenous Deaf communities must be central stakeholders in AI research and policy, reinforcing the cultural significance of sign languages alongside their spoken counterparts.

7.5 Future Research Directions: Expanding AI Models for Linguistic Diversity

Although AI has been successfully applied to major world languages, there remains a significant gap in AI research for low-resource and endangered languages (Ray et al., 2024). Many Indigenous languages lack structured datasets, formal grammatical documentation, and AI training models, limiting their inclusion in machine learning-based linguistic reconstruction (Prieto et al., 2024a, b). Future research should focus on expanding AI capabilities for Indigenous language processing, ensuring that computational linguistics serves diverse linguistic traditions rather than reinforcing Western-centric language hierarchies (Pinhanez et al., 2024). Priority areas for future research include:

- Developing AI models tailored for polysynthetic, agglutinative, and morphologically complex Indigenous languages, which differ significantly from Indo-European linguistic structures (Pinhanez et al., 2024).

- Advancing AI-driven phonetic modeling for unwritten and undocumented languages, allowing speech synthesis models to predict and generate pronunciation patterns based on linguistic relatives (Pinhanez et al., 2024).

Vol. 04, No. 01; 2025
ISSN: 2583-6560

- Integrating AI-driven conversational models with Indigenous oral storytelling traditions, ensuring that AI-generated language tools reflect Indigenous grammatical structures and narrative conventions (Flavelle & Lachler, 2023; Pinhanez et al., 2024).

- Enhancing AI-generated text-to-video capabilities for visual language learning, allowing learners to see and hear real-time facial expressions, gestures, and pronunciation cues to improve fluency and comprehension (Ghildyal et al., 2024).

- Developing ethical AI frameworks prioritizing Indigenous data sovereignty, preventing AI-driven linguistic extraction without community governance and ensuring that AI-generated resources align with Indigenous pedagogical priorities (Pinhanez et al., 2023).

Expanding AI research into text-to-video synthesis for language learning will allow researchers to create interactive, culturally embedded digital materials that reinforce oral traditions and community-based learning. Future studies should focus on developing high-fidelity audiovisual learning tools, incorporating gesture recognition, pronunciation correction, and contextualized storytelling to enhance the accessibility and effectiveness of Indigenous language revitalization efforts. As AI continues to evolve, its role in Indigenous language recovery must remain rooted in ethical, community-led research that prioritizes Indigenous knowledge, sovereignty, and linguistic diversity. Further exploration of AI's role in audiovisual language transmission can enhance Indigenous-led language sustainability efforts, ensuring that oral traditions remain central to digital language recovery initiatives.

8. CONCLUSION

This study highlights the transformative potential of AI in preserving endangered Indigenous languages and reconstructing lost languages such as Muisca. AI-driven tools, including NLP, phonetic modeling, OCR, and TTS synthesis, provide new avenues for linguistic documentation, reconstruction, and accessibility. In addition to text-based AI technologies, emerging text-to-video AI capabilities offer further potential by creating immersive audiovisual materials that reinforce oral language transmission and cultural storytelling. These tools can simulate pronunciation, gestures, and conversational contexts, making Indigenous language learning more interactive and accessible for scholars and community members. Researchers leveraging AI can analyze historical linguistic data, infer missing grammatical structures, and develop multimodal learning platforms that enhance academic research and community-based language recovery efforts.

However, while AI presents innovative and scalable solutions, its success in Indigenous language preservation depends on collaboration, ethical considerations, and community engagement. For AI-driven language recovery to be sustainable and effective, Indigenous communities must play a significant role in developing, validating, and governance of AI-generated linguistic models. Many Indigenous languages encode unique cultural, historical, and epistemological knowledge systems, making it essential that recovery efforts prioritize Indigenous linguistic sovereignty and community consent. AI must serve as a supportive tool for Indigenous language recovery—encompassing revitalization, reclamation, and reconstruction—rather than as an extractive technology. Indigenous knowledge holders must retain control over language data, digital archives, and AI-generated resources to prevent the commodification or misrepresentation of their linguistic heritage.

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Moreover, successfully implementing AI in language recovery and preservation requires interdisciplinary collaboration between linguists, AI researchers, educators, and Indigenous language advocates. While AI can aid recovery, it cannot fully capture the oral traditions, nonverbal linguistic cues, and contextual knowledge embedded in Indigenous languages. Therefore, a holistic approach is necessary—one that integrates AI methodologies with traditional linguistic research, oral history documentation, and community-led education programs. Text-to-video AI can bridge technological innovation and Indigenous pedagogical traditions, reinforcing learning through audiovisual storytelling, digital language immersion, and culturally relevant media.

While this study underscores the exciting possibilities of AI in Indigenous language recovery, it also acknowledges the challenges that must be addressed. Issues such as bias in AI models, data scarcity, and ethical concerns regarding Indigenous intellectual property rights must be carefully navigated to prevent the reinforcement of colonial linguistic hierarchies in digital spaces. Future research should expand AI capabilities for low-resource languages, improve phonetic modeling for unwritten Indigenous languages, and develop AI frameworks that align with Indigenous cultural and pedagogical priorities. Further exploration of the role of text-to-video in language transmission can enhance Indigenous-led language sustainability efforts, ensuring that oral traditions remain at the center of digital language recovery initiatives.

REFERENCES

Agarwal, M., & Anastasopoulos, A. (2024). A concise survey of OCR for low-resource languages. In *Proceedings of the 4th workshop on natural language processing for Indigenous languages of the Americas (AmericasNLP 2024), pages 88–102, Mexico City, Mexico.* Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.americasnlp-1.10

Ahearn, L. M. (2021). *Living language: An introduction to linguistic anthropology*. John Wiley & Sons.

Ajani, Y. A., Oladokun, B. D., Agboola, O. S., Bashorun, M. T., & others. (2024). Revitalizing Indigenous knowledge systems via digital media technologies for sustainability of Indigenous languages. *Preservation, Digital Technology & Culture*, *53*(1), 35-44. https://doi.org/10.1515/pdtc-2023-0051

Akumbu, P. W., Mazzoli, M., Outakoski, H. M., & Sippola, E. (2024). Decolonization and linguistics: A conversation on opportunities and challenges. *Journal of Postcolonial Linguistics*, *10*, 26-40. https://www.diva-portal.org/smash/get/diva2:1924316/FULLTEXT01.pdf

Ancient Origins. (2016, February 3). *The last of the Charrúa: Honored warriors of the lost tribe of Uruguay.* Ancient Origins. https://www.ancient-origins.net/history-famous-people/last-charrua-honored-warrior-tribe-uruguay-003528

Archive of the Indigenous Languages of Latin America – AILLA (2025). *Language archives.* http://www.language-archives.org/archive/ailla.utexas.org

Arul Dayanand, S., Uma Devi, M., & Kumar, R. (2023). Digital archives and preservation techniques for revitalising endangered languages. *Shanlax International Journal of English, 12*(1), 174-182. https://doi.org/10.34293/rtdh.v12iS1-Dec.133

Avila, E. (2015, May 9). *Bringing the Muisca language back from extinction, one word at a time.* Rising Global Voices. https://rising.globalvoices.org/blog/2015/05/09/bringing-the-muisca-language-back-from-extinction-one-word-at-a-time/

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Barkaskasi, P., & Gladwin, D. (2021). Pedagogical talking circles: Decolonizing education through relational Indigenous frameworks. *Journal of Teaching and Learning*, *15*(1), 20–38. https://doi.org/10.22329/JTL.V15I1.6519

Benn, M. (2010). *Spanish-Creole characteristics in the dialect of people with African origins in Port Limón, Costa Rica* (Master's thesis). Eastern Michigan University, Ypsilanti, MI.

Bernard, D. M. (2024). Reimagining Indigenous education in Costa Rica through Paulo Freire's pedagogical lens. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades 5*(4), 3415–3437. https://doi.org/10.56712/latam.v5i4.2503

Bird, S. (2020). *Decolonising speech and language technology.* In *Proceedings of the 28th international conference on computational linguistics* (pp. 3504–3519). Barcelona, Spain (Online). International Committee on Computational Linguistics. https://doi.org/10.18653/v1/2020.coling-main.313

Bradley, D., & Bradley, M. (2019). Language reclamation. In *Language endangerment* (pp. 208–227). chapter, Cambridge: Cambridge University Press.

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

Campbell, L. (2024). *The Indigenous languages of the Americas*. Oxford Academic.

Carpenter, C., Lyon, J., Thorogood, M., & Armstrong, J. (2024). Seeding alignment between language technology and Indigenous methodologies: A decolonizing framework for endangered language revitalization. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @* LREC-COLING 2024, pp. 318–324, Torino, Italia. ELRA and ICCL. https://aclanthology.org/2024.sigul-1.38.pdf

Chiruzzo, L., Denisov, P., Molina-Villegas, A., Fernandez-Sabido, S., Coto-Solano, R., Agüero-Torales, M., Alvarez, A., Canul-Yah, S., Hau-Ucán, L., Ebrahimi, A., Pugh, R., Oncevay, A., Rijhwani, S., Von Der Wense, K., & Mager, M. (2024). Findings of the AmericasNLP 2024 shared task on the creation of educational materials for Indigenous languages. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas* (AmericasNLP 2024). https://aclanthology.org/2024.americasnlp-1.27.pdf

Ch'ng, L. K. (2024). Standing on the shoulders of generative AI. In *Transforming education with generative AI: Prompt engineering and synthetic content creation* (pp. 1–21). IGI Global. https://doi.org/10.4018/979-8-3693-1351-0.ch001

Ciucci, L. (2021). From fieldwork to reconstruction: Historical issues in hotspots of linguistic diversity. *Studia Linguistica, 75*(2), 165-174. https://doi.org/10.1111/stul.12160

Coleman, J., Krishnamachari, B., Rosales, R., & Iskarous, K. (2024). LLM-assisted rule-based machine translation for low/no-resource languages. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas* (AmericasNLP 2024) (pp. 67–87). Association for Computational Linguistics. https://aclanthology.org/2024.americasnlp-1.9.pdf

Cooper, N., Heldreth, C., & Hutchinson, B. (2024). "It's how you do things that matters"—Attending to process to better serve Indigenous communities with language technologies. *Journal of Indigenous Technology, 39*(4), 121–143. https://aclanthology.org/2024.eacl-short.19.pdf

Coto-Solano, R. (2021). Explicit tone transcription improves ASR performance in extremely low-resource languages: A case study in Bribri. In *Proceedings of the AmericasNLP 2021 Workshop.* https://aclanthology.org/2021.americasnlp-1.20.pdf

Coto-Solano, R., Kim, T. W., Jones, A., & Loáiciga, S. (2024). Multilingual models for ASR in Chibchan languages. In *Proceedings of NAACL 2024*. https://aclanthology.org/2024.naacl-long.471.pdf

Cruz, J. D. (2021). Colonial power and decolonial peace. *Peacebuilding, 9*(3), 274–288. https://doi.org/10.1080/21647259.2021.1906516

Daigneault, A.L., & Anderson, G.D.S. (2023). Living dictionaries: A platform for Indigenous and under-resourced languages. *Dictionaries: Journal of the Dictionary Society of North America 44*(2), 57-74. https://dx.doi.org/10.1353/dic.2023.a915065

Dorzheeva, A. (2021). Preservation of Indigenous languages in the United States: Challenges and policy recommendations. *Journal of Linguistic Anthropology, 42*(1), 32-56. https://doi.org/10.15405/epsbs.2021.06.03.34

Endangered Language Alliance – ELA. (n.d.-a). *Bribri*. Retrieved from https://www.elalliance.org/languages/bribri

Endangered Language Alliance - ELA. (n.d.-b). *Forging a future for languages*. Retrieved from https://www.elalliance.org/

Endangered Languages Project - ELP. (n.d.-a). *Bribri language status and transmission*. Retrieved from https://www.endangeredlanguages.com/lang/687

Endangered Languages Project – ELP. (n.d.-b). *Cabécar language status and transmission*. Retrieved from https://www.endangeredlanguages.com/lang/cjp

Endangered Languages Project – ELP. (n.d.-c). *Maléku language status and transmission*. Retrieved from https://www.endangeredlanguages.com/lang/4246

Ermolova, T., Savitskaya, N., Dedova, O., & Guzova, A. (2024). The potential of artificial intelligence in preserving linguistic diversity for future generations. *Science for Education Today, 14*(1), 184–198. http://dx.doi.org/10.15293/2658-6762.2401.09

Farooq, U., Rahim, M., Sabir, N., Hussain, A., & Abid, A. (2021). Advances in machine translation for sign language: Approaches, limitations, and challenges. *Neural Computing and Applications, 33*, 14357–14399. https://doi.org/10.1007/s00521-021-06079-3

Feldman, I., & Coto-Solano, R. (2020). Neural machine translation models with back-translation for the extremely low-resource Indigenous language Bribri. In *Proceedings of COLING 2020*. https://aclanthology.org/2020.coling-main.351.pdf

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

Flavelle, A., & Lachler, J. (2023). Strengthening relationships between Indigenous communities, documentary linguists, and computational linguists in the era of NLP-assisted language revitalization. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP* (C3NLP), pages 25–34, Dubrovnik, Croatia. Association for Computational Linguistics. https://aclanthology.org/2023.c3nlp-1.4.pdf

Gartlehner, G., Kahwati, L., Hilscher, R., Thomas, I., Kugley, S., Crotty, K., Viswanathan, M., Nussbaumer-Streit, B., Booth, G., Erskine, N., Konet, A., & Chew, R. (2023). Data extraction for evidence synthesis using a large language model: A proof-of-concept study. *Research Synthesis Methods, 15*, 576-589. https://doi.org/10.1002/jrsm.1710

Gautam, S. K. (2019). The role of Indigenous knowledge in biodiversity conservation: Integrating traditional practices with modern environmental approaches. *Environmental Reports, 1*(2), 1–3. https://doi.org/10.51470/ER.2019.1.2.01

Gessler, L., & von der Wense, K. (2024). NLP for language documentation: Two reasons for the gap between theory and practice. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas* (AmericasNLP 2024) (pp. 1–6). Association for Computational Linguistics. https://aclanthology.org/2024.americasnlp-1.1.pdf

Ghanim, H. (2024). *Artificial intelligence and the spatial documentation of languages*. Cornell University. https://doi.org/10.48550/arXiv.2404.01263

Ghildyal, A., Chen, Y., Zadtootaghaj, S., Barman, N., & Bovik, A. C. (2024). *Quality prediction of AI generated images and videos: Emerging trends and opportunities.* ArXiv. Cornell University. https://doi.org/10.48550/arXiv.2410.08534

Giner-Miguelez, J., Gómez, A., & Cabot, J. (2024). *Using large language models to enrich the documentation of datasets for machine learning.* Cornell University. https://doi.org/10.48550/arXiv.2404.15320

Gonçalves, D. R. (2024). Language is the earth: How Indigenous languages safeguard the planet. *Cultural Survival Quarterly [Online]*. https://www.culturalsurvival.org/publications/cultural-survival-quarterly/language-earth-how-indigenous-languages-safeguard

Groenewald, E. S., Pallavi, P., Rani, S., Singla, P., Howard, E., & Groenewald, C. A. (2024). Artificial intelligence in linguistics research: Applications in language acquisition and analysis. *Naturalista Campano, 28*(1), 1253-1262. https://www.museonaturalistico.it/index.php/journal/article/download/239/190

Guerrettaz, A. M., & Engman, M. M. (2023). Indigenous language revitalization. In *Oxford Research Encyclopedia of Education*. Retrieved January 15, 2025, from https://doi.org/10.1093/acrefore/9780190264093.013.559

Henne-Ochoa, R., Elliott-Groves, E., Meek, B. A., & Rogoff, B. (2020). Pathways Forward for Indigenous Language Reclamation: Engaging Indigenous Epistemology and Learning by Observing and Pitching in to Family and Community Endeavors. *The Modern Language Journal*, *104*(2), 481–493. http://www.jstor.org/stable/45286674

Hymes, D. H. (1974). *Foundations in sociolinguistics: An ethnographic approach.* University of Pennsylvania Press.

Jafari, Z. (2023). The role of AI in supporting Indigenous languages. *AI and Tech in Behavioral and Social Sciences, 1*(2), 4–11. https://doi.org/10.61838/kman.aitech.1.2.2

Karson, J., & Coto-Solano, R. (2024). Morphological tagging in Bribri using Universal Dependency features. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 56–66, Mexico City, Mexico. Association for Computational Linguistics. https://aclanthology.org/2024.americasnlp-1.8.pdf

Kubishi Research Group (2024). *Research topics.* Retrieved January 8, 2025, from https://research.kubishi.com/

Langdon, C. (2020). Language documentation and revitalization among the Southern Cone Indigenous peoples: Challenges and achievements. *The Journal de la Société des Américanistes*, 129–150. https://doi.org/10.4000/jsa.18933

Living Dictionaries (n.d.). *Living dictionaries: Ngäbere (Guaymí).* Retrieved January 15, 2025, from https://livingdictionaries.app/ngabere/about

Longpre, S., Singh, N., Cherep, M., Tiwary, K., Materzynska, J., Brannon, W., Mahari, R., Dey, M., Hamdy, M., Saxena, N., Anis, A. M., Alghamdi, E. A., Chien, V. M., Yin, D., Qian, K., Li, Y., Liang, M., Dinh, A., Mohanty, S., . . . Kabbara, J. (2024). *Bridging the data provenance gap across text, speech, and video.* ArXiv. Cornell University. https://doi.org/10.48550/arXiv.2412.17847

Luhmann, J., & Burghardt, M. (2022). Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of the Association for Information Science and Technology*, *73*(2), 148-171. https://doi.org/10.1002/asi.24533

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Max-Planck-Institute – MPI (2025). *Research initiatives*. Retrieved January 31, 2025, from https://www.mpi.nl/research

Mendecka, K. (2023). The right to mother tongue education for Indigenous Peoples: An overview of international and regional standards. *Studia Juridica*, *96*, 208-231. https://doi.org/10.31338/2544-3135.si.2023-96.11

Mirdha, P. (2020). Translating orality: Empowering the Indigenous culture and literary traditions in India. *Journal of Indigenous Cultural Studies*, *14*(1), 23–44. https://www.researchgate.net/publication/349140591_Translating_Orality_Empowering_the_Indigenous_Culture_and_Literary_Traditions_in_India.

Nanduri, D., & Bonsignore, E. (2023). *Revitalizing endangered languages: AI-powered language learning as a catalyst for language appreciation.* Cornell University. https://doi.org/10.48550/arXiv.2304.09394

Native Languages (2020). *Bribri*. Retrieved January 15, 2025, from https://www.native-languages.org/bribri.htm

Native Languages (2020). *Cabecar*. Retrieved January 15, 2025, from https://www.native-languages.org/cabecar.htm

Noels, K. A., Lou, N. M., Vargas Lascano, D. I., Chaffee, K. E., Dincer, A., Zhang, Y. S. D., & Zhang, X. (2020). Self-determination and motivated engagement in language learning. In M. Lamb, K. Csizér, A. Henry, & S. Ryan (Eds.), *The Palgrave handbook of motivation for language learning (pp. XX–XX).* Palgrave Macmillan. https://doi.org/10.1007/978-3-030-28380-3_5

Open Language Archives Community – OLAC. (2025a). *OLAC resources for languages in the Americas.* Retrieved from http://www.language-archives.org/area/americas

Open Language Archives Community – OLAC. (2025b). *OLAC resources in and about the Bribri language.* Retrieved from http://www.language-archives.org/language/bzd

Open Language Archives Community – OLAC. (2025c). *OLAC resources in and about the Cabécar language.* Retrieved from http://www.language-archives.org/language/cjp

Open Language Archives Community – OLAC. (2025d). *OLAC resources in and about the Chibcha language.* Retrieved from http://www.language-archives.org/language/chb

Open Language Archives Community – OLAC. (2025e). *OLAC resources in and about the Cumanagoto language.* Retrieved from http://www.language-archives.org/language/cuo

Open Language Archives Community – OLAC. (2025f). *OLAC resources in and about the Kuna language.* Retrieved from http://www.language-archives.org/language/kvn

Open Language Archives Community – OLAC. (2025g). *OLAC resources in and about the Lenca language.* Retrieved from http://www.language-archives.org/item/oai:ailla.utexas.org:len

Open Language Archives Community – OLAC. (2025h). *OLAC resources in and about the Ngäbere language.* Retrieved from http://www.language-archives.org/language/gym

Open Language Archives Community – OLAC. (2025i). *OLAC resources in and about the Puelche language.* Retrieved from http://www.language-archives.org/language/pue

Patil, A. (2024). Leveraging AI for historical linguistics. *International Journal For Multidisciplinary Research*, *6*(2), 1–6. https://www.ijfmr.com/papers/2024/2/16926.pdf

Pine, A., Wells, D., Brinklow, N., Littell, P., & Richmond, K. (2022). Requirements and motivations of low-resource speech synthesis for language revitalization. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 7346–7359, Dublin, Ireland.* Association for Computational Linguistics. https://aclanthology.org/2022.acl-long.507.pdf

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Pinhanez, C., Cavalin, P., Vasconcelos, M., & Nogima, J. (2023). Balancing social impact, opportunities, and ethical constraints of using AI in the documentation and vitalization of Indigenous languages. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23) Special Track on AI for Good, 6174-6182*. https://doi.org/10.24963/ijcai.2023/685

Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., Vasconcelos, M., Domingues, P., De Souza Mizukami, P., Grell, N., Gongora, M., & Gonçalves, I. (2024). *Harnessing the power of artificial intelligence to vitalize endangered Indigenous languages: Technologies and experiences.* Cornell University. https://doi.org/10.48550/arXiv.2407.12620

Pradhan, U., & Dey, J. (2023). Language, artificial education, and future-making in Indigenous language education. Learning, Media and Technology, 1-14. Pradhan, U., & Dey, J. (2023). Language, artificial education, and future-making in Indigenous language education. *Learning, Media and Technology*, 1–14. https://doi.org/10.1080/17439884.2023.2278111

Prieto, J., Martínez, C., Robles, M., Moreno, A., Palacios, S., & Manrique, R. (2024a). Translation systems for low-resource Colombian Indigenous languages, a first step towards cultural preservation. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*. https://doi.org/10.18653/v1/2024.americasnlp-1.2

Prieto, C., Martínez, L., & Gómez, J. (2024b). Translation systems for low-resource Colombian Indigenous languages: Challenges and advancements. *In Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024), pages 7–14, Mexico City, Mexico.* Association for Computational Linguistics. https://aclanthology.org/2024.americasnlp-1.2.pdf

Rahima, A. (2024). Language documentation as a revitalization effort regional languages are starting to extinct. *International Journal of Multidisciplinary Sciences and Arts, 3*(3), 72-76. 10.47709/ijmdsa.v3i3.4123

Ray, S., Vidhate, D., Singla, P., , P., Grover, S., & Ray, -. (2024). Exploring the role of artificial intelligence in language documentation and endangered language preservation. *Tuijin Jishu/Journal of Propulsion Technology, 45*(2), 1630–1639. https://doi.org/10.52783/tjjpt.v45.i02.6122

Redvers, N., Aubrey, P., Celidwen, Y., & Hill, K. (2023). Indigenous Peoples: Traditional knowledges, climate change, and health. *PLOS Global Public Health, 3*(10), e0002474. https://doi.org/10.1371/journal.pgph.0002474

Rojas, R. V. B., & Martínez-Cano, F. J. (Eds.). (2024). Revolutionizing communication: The role of artificial intelligence. CRC Press.

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

Sabrina, A., Maharani, S., Gunawan, M., & Pujiati, H. (2023). Discourse analysis of a news text using cohesion and ethnography of speaking theories. Stairs. *English Language Education Journal, 4*(1), 62-71. https://doi.org/10.21009/stairs.4.1.6

Sallabank, J., & Austin, P. K. (2023). Endangered languages. In J. H. Hulstijn & R. F. Young (Eds.), *The Routledge handbook of applied linguistics* (2nd ed.). Routledge. https://doi.org/10.4324/9781003082637-31

Sánchez Avendaño, C., & Angulo-Jiménez, H. (2023). Thematic picture dictionaries and other visual resources for Costa Rican Indigenous languages: Beyond bilingual equivalencies. *Dictionaries: Journal of the Dictionary Society of North America, 44*, 239 - 258. https://doi.org/10.1353/dic.2023.a915073

Vol. 04, No. 01; 2025
ISSN: 2583-6560

Sánchez Carrera, S., Zariquiey, R., & Oncevay, A. (2024). Unlocking knowledge with OCR-driven document digitization for Peruvian Indigenous languages. In *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas* (AmericasNLP 2024). https://aclanthology.org/2024.americasnlp-1.11.pdf

Sánchez-Martínez, F., Pérez-Ortiz, J., Sánchez-Cartagena, V., Lou, A., García-Romero, C., Jiménez, A., & Esplà-Gomis, M. (2024). Lightweight neural translation technologies for low-resource languages. , 4-5. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation* (Volume 2), pages 4–5, Sheffield, UK. European Association for Machine Translation (EAMT). https://aclanthology.org/2024.eamt-2.3.pdf

Semrl, N., Feigl, S., Taumberger, N., Bracic, T., Fluhr, H., Blockeel, C., & Kollmann, M. (2023). AI language models in human reproduction research: exploring ChatGPT's potential to assist academic writing. *Human Reproduction, 38*(12), 2281-2288. https://doi.org/10.1093/humrep/dead207

Shen, N. (2024). The status of Indigenous languages under globalization. *Transactions on Social Science, Education and Humanities Research, 13*, 267–274. https://doi.org/10.62051/8ze04c21

Sherzer, J. (2009). *Verbal art in San Blas: Kuna culture through its discourse.* Cambridge University Press.

Solano Acuña, R. (2021). Voces, imágenes y escritos que rememoran el pasado: Un análisis del lenguaje y la identidad en comunidades indígenas. *Journal of Indigenous Linguistic Studies, 34*, 56-72. https://doi.org/10.5944/rdh.43.2021.29349

Souter, C., Sammons, A., & Hutchins, L. (2024). Creating digital learning and reference resources for endangered Indigenous languages: Lessons from recent AI integration projects. *Computational Linguistics & Language Education, 22*(1), 93–116. https://aclanthology.org/2024.computel-1.10.pdf

Soylu, D., & Şahin, A. (2024). The role of AI in supporting Indigenous languages. *AI and Tech in Behavioral and Social Sciences, 2*(4), 11-18. https://doi.org/10.61838/kman.aitech.2.4.2

Srivastava, S., & Upadhyay, P. (2024). Digital empowerment for Indigenous communities using generative artificial intelligence. *GATR Global Journal of Business and Social Science Review (GATR-GJBSSR) 12*(2), 74–82. https://doi.org/10.35609/gjbssr.2024.12.2(3)

Sujatna, M., Astarina, A., & Heryono, H. (2024). AI for language learning: Friend or foe?. *Linguistic, English Education and Art (LEEA) Journal, 7*(2), 456-476. https://doi.org/10.31539/leea.v7i2.9331

Tennell, C., & Chew, K. A. (2024). Perspectives on relationality in online Indigenous language learning. *AlterNative: An International Journal of Indigenous Peoples, 20*(3), 512–520. https://doi.org/10.1177/11771801241261699

Tonja, A., Balouchzahi, R., & Méndez, P. (2024). NLP progress in Indigenous Latin American languages: Challenges and innovations. *Natural Language Processing Review, 17*(2), 31-48. https://doi.org/10.48550/arXiv.2404.05365

Trueba, H. T. (2022). The dynamics of cultural transmission. In *Healing multicultural America* (pp. 10-28). Routledge.

Tuhiwai Smith, L. (2012). *Decolonizing methodologies: Research and Indigenous Peoples*, 2nd ed. Zed Books.

United Nations Educational, Scientific and Cultural Organization - UNESCO. (n.d.). *International Decade of Indigenous Languages (2022–2032).* UNESCO. https://www.unesco.org/en/decades/indigenous-languages

United Nations Educational, Scientific and Cultural Organization - UNESCO. (2020, October 28). *Cutting edge: Protecting and preserving cultural diversity in the digital era.* Retrieved

Vol. 04, No. 01; 2025
ISSN: 2583-6560

January 6, 2025, from https://www.unesco.org/en/articles/cutting-edge-protecting-and-preserving-cultural-diversity-digital-era

United Nations Educational, Scientific and Cultural Organization - UNESCO. (2021). *The global action plan of the International Decade of Indigenous Languages is presented at UNESCO and living heritage is mobilized (2022-2032). UNESCO.* Retrieved January 7, 2025, from https://ich.unesco.org/en/news/the-global-action-plan-of-the-international-decade-of-indigenous-languages-is-presented-at-unesco-and-living-heritage-is-mobilized-13348

United Nations Educational, Scientific and Cultural Organization - UNESCO. (2022, January 5). *Cutting edge: Indigenous languages—Gateways to the world's cultural diversity.* UNESCO. Retrieved January 6, 2025, from https://www.unesco.org/en/articles/cutting-edge-indigenous-languages-gateways-worlds-cultural-diversity

Wang, L. (2024). Artificial intelligence's role in the realm of endangered languages: Documentation and teaching. *Applied and Computational Engineering, 48*, 123–129. https://doi.org/10.54254/2755-2721/48/20241249

Wang, N., & Bai, W. (2024). Indigenous language revitalization movements: Resistance against colonial linguistic domination. *Communications in Humanities Research, 35*, 69-74. https://doi.org/10.54254/2753-7064/35/20240015

World Bank (2019, February 22). *Lenguas indígenas: Legado en extinción [Infographic]*. The World Bank. https://www.worldbank.org/en/news/infographic/2019/02/22/lenguas-indigenas-legado-en-extincion

World Bank (2023). *Indigenous Peoples*. The World Bank. https://www.worldbank.org/en/topic/indigenouspeoples

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can large language models transform computational social science?. *Computational Linguistics, 50*(1), 237-291. https://aclanthology.org/2024.cl-1.8.pdf