International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 7 No. 5 January – February - 2026

Artificial Intelligence and Preservation of Indigenous and Regional Languages for Sustainable Development

Laxmikant Jaganrao Bante
Gramin Vikas Vidyalya and Jr. College Salwa, Taluka: Mouda, Dist.: Nagpur
DOI - 10.5281/zenodo.18897908

Abstract:

India is one of the most linguistically diverse countries in the world, home to hundreds of indigenous and regional languages that embody rich cultural knowledge, traditional wisdom, and collective identity. However, rapid globalization, the dominance of major languages, and accelerating digital transformation have placed many of these languages at risk of marginalization and extinction. Linguistic erosion threatens not only cultural identity but also the intergenerational transmission of ecological knowledge, oral histories, and community-based wisdom systems.

In this context, Artificial Intelligence (AI) has emerged as a powerful technological intervention capable of supporting language preservation, revitalization, and sustainable educational development. This conceptual paper critically examines the role of Artificial Intelligence in preserving indigenous and regional languages, with special reference to the Indian education system and language research. The study explores how AI-driven tools such as Natural Language Processing (NLP), speech recognition, machine translation, and digital language archives can contribute to inclusive education, multilingual learning, and sustainable development goals. Further, the paper discusses Indian initiatives and case examples where AI has been applied to linguistic documentation, translation, and educational access. By integrating AI with language education and research, the paper argues that technological innovation can act as a catalyst for linguistic sustainability, cultural preservation, social inclusion, and equitable knowledge dissemination in a multilingual society like India.

Keywords: Artificial Intelligence, Indigenous Languages, Regional Languages, Language Preservation, Sustainable Development, Education System, India

Introduction:

Language is not merely a medium of communication; it is a repository of culture, history, identity, worldview, and traditional knowledge. Indigenous and regional languages play a crucial role in shaping social values, community relationships, and educational foundations. They encode ecological knowledge, agricultural practices, medicinal traditions, and ethical systems that have evolved over centuries within specific environments.

In India, linguistic diversity is a defining feature of national heritage, with the Constitution recognizing multiple scheduled languages and thousands of dialects spoken across regions. Yet, despite this richness, many indigenous and regional languages face serious threats due to urbanization, migration, globalization, and the growing dominance of English and other major languages in education, administration, and digital communication. The erosion of linguistic diversity not only weakens cultural identity but also disrupts knowledge systems essential for sustainable living.

The advent of Artificial Intelligence has transformed sectors such as education, healthcare, governance, and research. In the education system, AI-driven tools are increasingly used for

personalized learning, automated assessment, content generation, and language instruction. When aligned with sustainable development objectives, AI offers immense potential to address linguistic inequalities, enhance educational access, and support the preservation and revitalization of endangered languages. This paper situates language preservation within the broader framework of sustainable development, emphasizing inclusive education, cultural sustainability, and equitable access to knowledge.

Artificial Intelligence in the Education System:

Artificial Intelligence refers to the simulation of human intelligence in machines capable of learning from data, recognizing patterns, making decisions, and improving performance over time. In education, AI technologies such as adaptive learning platforms, intelligent tutoring systems, speech recognition software, learning analytics, and automated translation tools have significantly transformed teaching and learning processes.

AI enables personalized education by adapting content to learners' linguistic backgrounds, learning pace, and cognitive styles. Students receive customized feedback, language support, and learning resources aligned with their individual needs. In multilingual societies like India, this adaptability is particularly valuable, as learners often come from diverse linguistic and cultural backgrounds.

AI-based language learning applications, voice-enabled educational tools, and real-time translation systems help students access educational content in their mother tongue or preferred language. This reduces comprehension barriers, improves learning outcomes, and promotes educational inclusion. Such technological interventions align with the principles of inclusive and equitable education emphasized in global sustainable development agendas and national policies such as the National Education Policy (NEP) 2020, which advocates mother-tongue instruction and multilingual education at foundational levels.

Indigenous and Regional Languages: The Need for Preservation:

Indigenous and regional languages are carriers of traditional ecological knowledge, social norms, belief systems, and community practices. They shape collective memory and cultural continuity. The loss of a language results in the loss of unique knowledge systems and social identities.

UNESCO has repeatedly warned that many indigenous languages worldwide are endangered, and India is no exception. Several tribal and regional languages are spoken by shrinking populations and often lack written scripts, standardized orthographies, institutional support, or digital representation. As younger generations shift toward dominant languages for economic mobility and social prestige, native languages gradually disappear from daily use.

Language preservation is therefore a critical component of sustainable development. Sustainable development extends beyond economic growth to include social equity, cultural continuity, environmental responsibility, and intergenerational justice. Preserving indigenous languages ensures that communities maintain their identity, knowledge systems, and agency while participating in modern educational and technological systems.

Role of Artificial Intelligence in Language Preservation:

Artificial Intelligence provides innovative tools for documenting, preserving, revitalizing, and promoting endangered languages. AI-powered systems can record, analyze, and store linguistic data efficiently, even for languages with

limited written resources or small speaker populations.

Natural Language Processing enables the creation of digital dictionaries, language corpora, morphological analyzers, and grammatical models. Speech recognition and text-to-speech technologies allow oral languages to be documented, transcribed, and taught digitally. Machine translation enhances accessibility by converting indigenous languages into widely spoken ones and vice versa, increasing visibility and usability.

AI-driven conversational agents and chatbots can support language learning, practice, and daily use, especially among younger generations who interact primarily through digital platforms. These tools transform language preservation from static archiving into active language revitalization.

Indian Case Studies of AI in Language Preservation:

1. Bhashini -National Language Translation Mission:

The Government of India's **Bhashini** platform uses AI to provide translation, speech recognition, and text-to-speech services across Indian languages. It enables citizens to access digital content and government services in regional and indigenous languages. From an educational perspective, Bhashini supports multilingual learning by allowing students and teachers to access academic resources in their mother tongue, promoting linguistic inclusion, digital equity, and educational access.

2. AI-Based Documentation of Tribal Languages:

Several Indian universities and research institutions use AI tools to document endangered tribal languages such as Gondi, Santhali, Bhili, Kurukh, and Ho. Speech recognition and corpus-building technologies record oral narratives, folk traditions, medicinal knowledge, and historical accounts. These digital archives preserve linguistic heritage and serve as educational resources for tribal schools and researchers.

3. AI Translation in Educational Technology:

Indian EdTech platforms increasingly use AI-based translation, subtitles, and voice tools to deliver learning materials in regional languages. These platforms reduce linguistic barriers in higher education, vocational training, and digital literacy programs, enabling learners from non-English backgrounds to participate meaningfully in the digital economy.

4. Digitization of Sanskrit and Classical Languages:

AI is used to digitize and analyze classical languages such as Sanskrit and Pali. Machine learning assists in manuscript analysis, semantic indexing, searchable databases, and automated translation. These initiatives preserve India's classical intellectual heritage and integrate it into modern research and education systems.

5. Community-Centered Language Revitalization:

Community-based AI projects involve local speakers in developing digital storytelling tools, voice applications, and mobile language-learning platforms. Community participation ensures cultural sensitivity, ethical data use, and sustainable language revitalization, transforming speakers from passive subjects into active knowledge producers.

AI, Education, and Sustainable Development:

The integration of AI in language education supports Sustainable Development Goals, particularly SDG 4 (Quality Education) and SDG 11 (Sustainable Communities). By promoting mother-tongue instruction and multilingual education, AI reduces educational inequalities, improves learning outcomes, and supports cultural sustainability.

AI-driven language preservation initiatives also promote social inclusion by giving marginalized linguistic communities a digital presence and academic recognition. This fosters cultural pride, social cohesion, and intergenerational knowledge transmission.

Challenges and Ethical Concerns:

Despite its potential, AI-based language preservation faces challenges such as data scarcity for minority languages, algorithmic bias favoring dominant languages, ethical concerns related to cultural ownership, consent, and data privacy, and unequal access to digital infrastructure.

Without inclusive policies, ethical guidelines, and community participation, AI initiatives risk reinforcing existing inequalities. Therefore, human oversight, participatory approaches, and ethical governance frameworks are essential.

Conclusion:

The accelerating decline of indigenous and regional languages represents not merely a linguistic loss, but a profound erosion of cultural identity, traditional knowledge systems, and social memory. In multilingual societies such as India, language is deeply intertwined with community life, ecological wisdom, ethical values, and educational foundations. The marginalisation of minority languages therefore threatens both cultural diversity and sustainable development. This paper has argued that Artificial Intelligence, when thoughtfully integrated into education and research, offers a powerful means to counteract this decline and to transform language preservation from a passive archival activity into a dynamic, participatory, and sustainable process.

Through tools such as Natural Language Processing, speech recognition, machine translation, and digital archiving, AI enables the systematic documentation, analysis, and dissemination of endangered languages. These technologies allow oral traditions to be recorded, linguistic structures to be modelled, and learning resources to be created in ways that were previously impossible at scale. In the educational context, AI-supported multilingual platforms facilitate mother-tongue-based learning, improve comprehension, and enhance inclusion for learners from linguistically marginalised communities. In research, AI accelerates linguistic documentation and opens new possibilities for interdisciplinary studies linking language, culture, ecology, and society.

The Indian case studies discussed in this paper demonstrate that AI is already playing a significant role in promoting linguistic inclusion and sustainability. Initiatives such as the Bhashini platform, AI-based tribal language documentation projects, multilingual EdTech tools, digitisation of classical languages, and community-driven language revitalisation efforts illustrate how technology can be aligned with social and cultural goals. These examples show that AI can function not as a force of linguistic homogenisation, but as an instrument of diversity, inclusion, and cultural empowerment — provided it is guided by ethical principles and social responsibility.

However, the paper has also highlighted that technological potential alone is insufficient. Without inclusive policies, community participation, ethical safeguards, and equitable digital infrastructure, AI-driven language initiatives risk reproducing existing inequalities and marginalising the very communities they seek to support. Issues such as data bias, cultural ownership, consent, privacy, and access must therefore remain central to the design and implementation of AI systems. Language preservation must be understood not as a purely technical task, but as a socio-cultural process that

requires collaboration between technologists, educators, linguists, policymakers, and local communities.

In conclusion, Artificial Intelligence should not be viewed as a replacement for human linguistic and cultural practices, but as a supportive tool that enhances human capacity to preserve, transmit, and revitalise linguistic heritage. The future of language preservation lies in the harmonious integration of technology, education, culture, and ethical governance. When AI is embedded within inclusive educational frameworks and grounded in community participation, it can contribute meaningfully to linguistic sustainability, cultural continuity, and equitable development. In this way, AI becomes not only a technological innovation, but a catalyst for sustaining the linguistic and cultural diversity that is essential to the social and intellectual richness of human civilisation.

References:

1. Bird, S. (2020). Decolonising speech and language technology. *Proceedings of the 28th International Conference on Computational Linguistics*, 3504–3519.
2. Bender, E. M. (2019). The dangers of stochastic parrots: Can language models be too big? *Proceedings of the ACL*.
3. Government of India. (2020). *National Education Policy 2020*. Ministry of Education.
4. Government of India. (2022). *Bhashini: National Language Translation Mission*. Ministry of Electronics and Information Technology.
5. Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *Proceedings of ACL*, 6282–6293.
6. Mohanty, A. K. (2019). *Multilingualism and language policy in India*. Oxford University Press.
7. UNESCO. (2018). *Safeguarding endangered languages*. UNESCO Publishing.
8. UNESCO. (2022). *Global education monitoring report*. UNESCO Publishing.
9. United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*.
10. Varshney, L. R., & Sun, J. (2020). Artificial intelligence and cultural sustainability. *AI & Society, 35*(2), 389–401. https://doi.org/10.1007/s00146-018-0841-y