Research Journal of English Language and Literature (RJELAL)
A Peer Reviewed (Refereed) International Journal
Impact Factor 6.8992 (ICI) http://www.rjelal.com;
Email:editorrjelal@gmail.com; ISSN:2395-2636 (P); 2321-3108(O)

Vol.13.Issue 4. 2025 (Oct-Dec)

RESEARCH ARTICLE

INTERNATIONAL STANDARD SERIAL NUMBER INDIA
2395-2636 (Print):2321-3108 (online)

Preservation and Promotion of Indian Languages for Indigenous Knowledge Transfer

Dr. Ankita Pandey
Assistant Professor, Christ Church College, Kanpur
Email: Ankitapandey0797@gmail.com

DOI: 10.33329/rjelal.13.4.276

Article info
Article Received: 23/10/2025
Article Accepted: 30/11/2025
Published online: 07/12/2025

Abstract

India's vast linguistic diversity forms a critical interface with its Indigenous Knowledge Systems (IKS), weaving together heritage, education, and national identity. The country's endangered languages especially those from smaller families like Tibeto-Burmese face cascading risks of extinction, imperilling cultural memory and specialized local expertise. National schemes, such as SPPEL, Bharata Vani, and Sanchika, document and digitally archive India's linguistic wealth, forming foundational data pipelines for multilingual AI efforts and future Indigenous Large Language Models (LLMs). The National Education Policy (NEP 2020) foregrounds mother tongue-centred pedagogy and integration of IKS into curricula, though its implementation is challenged by teacher shortages, state-level political barriers, and variable digital infrastructure.

Effective language preservation requires robust, ethically governed documentation and targeted inclusion of functional IKS vocabulary, digital resources, and multimodal NLP strategies. Pedagogical innovations like Mother Tongue-Based Multilingual Education (MTB-MLE) and Indigenous Language Immersion programs support intergenerational transfer of knowledge and identity, as demonstrated by both Indian and global models. India's technological roadmap anchored by BharatGen, Adi-Vaani, and Bhashini aims to bridge the resource gap, extend digital access to remote communities, and position India as a global leader in AI-driven linguistic conservation.

To ensure enduring success, policy must integrate language as a factor of community wellness and social equity, prioritize economic relevance and community engagement, and accelerate coordinated action across data infrastructure, pedagogy, and governance. The preservation of India's endangered languages, therefore, is inseparable from the safeguarding of its indigenous wisdom and the social justice of its communities.

Research Journal of English Language and Literature (RJELAL)
A Peer Reviewed (Refereed) International Journal
Impact Factor 6.8992 (ICI) http://www.rjelal.com;
Email:editorrjelal@gmail.com; ISSN:2395-2636 (P); 2321-3108(O)

Vol.13.Issue 4. 2025 (Oct-Dec)

Keywords: Large Language Models, Mother Tongue-Based Multilingual Education (MTB-MLE), Economic relevance, Pedagogy, Governance.

I. The Critical Interface: Language, Knowledge, and Vulnerability

1.1. India's Linguistic Landscape and the Imperative for Preservation

India represents one of the world's most complex and diverse linguistic environments, characterized by four major language families. The Indo-European family, comprising 24 languages, accounts for the largest proportion of speakers, representing approximately 76.89% of the total population. The Dravidian family follows, consisting of 17 languages spoken by about 20.82% of the population. Critically, the Austro-Asiatic family (14 languages) and the Tibeto-Burmese family (66 languages) account for a significantly smaller share of the population, 1.11% and 1.00% respectively, despite the latter containing a large number of distinct languages. This distribution underscores an uneven vitality; while major languages are robust, the numerous smaller indigenous and tribal languages often possess few surviving speakers, placing them at high risk of extinction. The imperative for preservation therefore extends far beyond the 22 Scheduled Languages, focusing intensely on mother tongues spoken by communities of less than 10,000 people, which serve as the primary targets for documentation efforts under national schemes.

1.2. Mapping Endangerment: Status and Distribution

Language loss is formally mapped using established international frameworks. UNESCO defines four critical levels of language endangerment between "safe" and "extinct": Vulnerable, Definitely endangered, Severely endangered, and Critically endangered. When a language loses all native speakers, it becomes extinct.

Analysis of the endangered linguistic population reveals geographically concentrated hotspots. For instance, Critically Endangered languages documented by UNESCO include Kuruba, Naiki, and Nihali. These languages are spread across diverse regions, reflecting the pervasive nature of the crisis. Nihali, for instance, is found in the West Central zone (covering Maharashtra, Madhya Pradesh, and Gujarat), while languages like Toda and Kuruba are found in the Southern zone. The Northeast zone, characterized by a high concentration of languages belonging to the Tibeto-Burmese family (66 languages), is particularly vulnerable, featuring languages like Tangam and Aimol among the endangered.

The existence of 66 Tibeto-Burmese languages representing only 1% of the national population necessitates a decentralized and intensive resource allocation model. Preservation strategies in the Northeast must be hyper-localized and high-cost per speaker to achieve effective coverage, a structural reality that differs significantly from conservation efforts directed at languages within the larger Indo-European or Dravidian families.

Linguistic Endangerment Status and Distribution by Language Family:

Language Family	Approx. Number of Languages	Approx. Population Share	Key Endangered Zones	Example Critically Endangered Languages
Indo-European	24	76.89%	Northern/Central India	Darmiya, Spiti
Dravidian	17	20.82%	Southern India	Kuruba, Toda
Austro-Asiatic	14	1.11%	East Central India	Birhor, Bondo, Gorum
Tibeto-Burmese	66	1.00%	Northeast India	Nihali, Tangam, Aimol

1.3. Language as an Epistemic Container (The IKS Link)

The vitality of the Indian Knowledge System (IKS) is inextricably linked to linguistic diversity. Regional languages function not merely as communication tools but as crucial epistemological containers, preserving cultural heritage and ensuring the seamless transmission of indigenous knowledge.

The decline of linguistic vitality directly precipitates the erosion of IKS. For example, the shift to contemporary healthcare systems has been accompanied by the depletion of traditional medicinal knowledge, particularly concerning the healing qualities of local flora. This knowledge is typically embedded in specific regional languages, making its loss unavoidable when the native language declines. This linguistic erosion is rapid, driven primarily by globalization and urbanization, which interrupt intergenerational knowledge transfer as young people relocate to metropolitan areas, detaching themselves from their ancestral communities and traditions.

A case study on the Ahirani language, an Indo-Aryan language spoken in the Khandesh region of Maharashtra, illustrates this symbiotic relationship clearly. Ahirani has historically served as a critical medium for oral traditions, folk literature, and socio-cultural practices. Survey data indicates that the language is profoundly functional in its community context: 92% of respondents connect Ahirani to indigenous agricultural knowledge, and 80% link it to cultural and religious rituals. This functional indispensability means that documentation mandates must extend beyond mere grammar and vocabulary to capture specific, domain-related terminology such as agricultural terms, craft methods, or medicinal plant names that are functionally embedded within the language. If documentation focuses only on basic linguistic structure, the essential IKS context is lost, rendering the preservation effort incomplete. Therefore, effective documentation protocols require mandatory integration of IKS experts and community practitioners alongside technical linguists.

The successful preservation effort must be measured not only by the creation of an archive but by its ability to engage the dispersed, urbanized youth audience who are currently detached from the oral traditions of their communities, thus countering the effects of internal migration.

II. Policy Architecture and Institutional Responses

2.1. The National Education Policy (NEP 2020) Mandate

The National Education Policy (NEP 2020) provides a robust policy framework

emphasizing the intrinsic value of India's native knowledge traditions. The policy recognizes Indian languages explicitly as carriers of indigenous knowledge systems, advocating for the promotion and preservation of these linguistic assets. To foster a comprehensive approach to education, the NEP encourages universities to offer courses that combine traditional and contemporary knowledge.

Multilingualism is promoted as a fundamental principle of education, designed to enable students to access traditional texts and knowledge in their original languages. To support this vision, the policy proposes the creation of new specialized institutions, including the Indian Institute of Translation and Interpretation and specialized National Institutes for Pali, Persian, and Prakrit, underscoring a national commitment to deep linguistic engagement.

However, the path from policy formulation to implementation is fraught with challenges. Linguistic policy frequently encounters political barriers at the state level. The implementation of the multi-language formula, for example, has faced resistance, notably from states like Tamil Nadu, which has not implemented the formula. This state-level resistance demonstrates that linguistic policy, rather than being purely a pedagogical tool, can become a political barrier in education, stalling nationwide implementation of the policy's aims.

2.2. Key Preservation and Documentation Schemes (CIIL)

The Central Institute of Indian Languages (CIIL) manages several critical schemes that form the backbone of the national preservation effort.

The **Scheme for Protection and Preservation of Endangered Languages (SPPEL)**, instituted in 2013, directly addresses the deteriorating situation of lesser-known languages. SPPEL focuses rigorously on documenting and archiving languages spoken by fewer than 10,000 speakers, or languages that have not been previously studied by linguists. The scheme aims to accomplish documentation in the form of grammar, dictionary, and ethno-linguistic profiles for approximately 500 such lesser-known languages in the coming years.

The **Bharatavani Project (BvP)** serves as India's leading initiative for language diversity and digital archiving, aligning with the Government's ongoing Digital India mission. Bharatavani functions as a single-point online repository and clearing house for knowledge in and about all Indian languages. It hosts over 5,500 resources in approximately 118 Indian languages, covering the 22 officially recognized languages and 96 other mother tongues. The project's mandate is broad, providing a digital, multi-modal platform offering resources such as audio-video recordings, monolingual and multilingual dictionaries, and language learning materials. This comprehensive digital repository is vital for fostering intergenerational and intercultural communication and promoting the languages' relevance in domains like research, education, and cultural preservation. The platform's commitment to providing multilingual content has a long-lasting measurable impact on community engagement and digital content creation.

2.3. Digital Infrastructure and AI Foundations

The advancement of digital infrastructure is essential for scaling preservation efforts. The digital documentation schemes are strategically linked to India's broader AI objectives.

Sanchika: Digital Repository for Indian Languages, managed by CIIL, aggregates foundational data dictionaries, primers, storybooks, and multimedia resources for Scheduled and tribal languages. Sanchika is fundamentally important because it acts as a centralized archive that provides the foundational, linguistically categorized datasets necessary for emerging Artificial Intelligence (AI) and Natural Language Processing (NLP) applications.

These documentation projects are not viewed merely as cultural expenditure but as critical national data infrastructure. Their efficiency dictates the quality and scope of ambitious advanced AI projects. This strategic alignment is evident in initiatives such as **Adi-Vaani** (AI for Tribal Language Inclusion), **BharatGen** (AI Models for Indian Languages), and **Bhashini** (AI-Driven Multilingual Translation). BharatGen, for instance, develops advanced text-to-text and text-to-speech translation models for the 22 Scheduled Languages, leveraging data gathered through SPPEL and Sanchika. Bhashini aims to ensure digital content is accessible across the diverse linguistic landscape, even powering applications like the GeMAI multilingual assistant for the Government e-Marketplace (GeM).

The eventual goal of this technological push, as articulated in national roadmaps, includes the launch of India's first dedicated Tribal Large Language Model (LLM) by 2027 and the digitization of over 50 endangered languages by 2030. Furthermore, India is positioning itself as a global leader by planning to host a UN-backed AI repository for endangered languages, with Adi-Vaani designated as a Digital Public Good. This leadership role necessitates adherence to high international standards of data governance, ensuring research ethics and collaborative oversight with indigenous communities, which is paramount for sustainable preservation.

Core Government Initiatives for Language Preservation and Digitization:

Initiative/Project	**Managing Authority**	**Primary Objective**	**Scope and Targets**	**Role in Indigenous Knowledge Transfer (IKS)**
SPPEL	CIIL	Documenting endangered languages.	Focus on languages <10,000 speakers (approx. 500 planned).	Capturing grammar, dictionary, and ethno-linguistic profiles before language extinction.
Bharatavani Project (BvP)	CIIL	Single-point, multilingual knowledge repository.	Hosts 5500+ resources in 118 languages; aims for digital revolution.	Disseminating cultural resources, folklore, and language learning materials worldwide.
Sanchika	CIIL	Centralized Digital Repository/Data Source.	Aggregates dictionaries, multimedia resources for tribal and scheduled languages.	Provides foundational, categorized data for training language models for IKS accessibility.
BharatGen/Bhashini	IndiaAI Mission	Develop AI for multilingual translation.	Advanced models for 22 Scheduled Languages, leveraging SPPEL/Sanchika data.	Ensures digital content and IKS resources are translated and accessible across India's major linguistic landscape.

III. Pedagogical Pathways for Intergenerational Transfer

3.1. Mother Tongue-Based Multilingual Education (MTB-MLE) in the Indian Context

The academic effectiveness of linguistic preservation is critically dependent on pedagogical innovation. Mother Tongue-Based Multilingual Education (MTB-MLE) is widely recognized as a highly effective educational strategy, particularly for ethnic minority children who struggle with the national language used in schools. The MTB-MLE model dictates that children learn exclusively in their native language (L1) during early grades, with other languages introduced later.

International and regional studies, including those relevant to tribal learners in India, confirm that MTB-MLE significantly improves students' academic achievement and retention. Research findings show attainment of higher scores, with gains of 15% to 35% across subjects like Reading, Math, Science, and Social Studies. This success stems from providing teaching and learning that is easily understood and closely related to the students' lives.

For tribal learners, such as those in Odisha, a lack of familiarity with the textbook background (which is often in the regional majority language, Odia) severely hinders comprehension and learning outcomes. MTB-MLE directly addresses this deficiency by connecting instruction to the rich knowledge networks (L1) the child already possesses. This process, known as cumulative bilingualism, enhances the intellectual development of tribal learners, allowing them to comprehend subjects and connect them with their culture.

3.2. Integrating IKS and Cultural Curriculum

Beyond mere language instruction, successful revitalization requires robust culture-based curricula. The goal is to embed Indigenous ways of knowing directly into the educational framework, moving beyond isolated language subjects to permeate areas such as history, science, and social studies.

A salient example is found in Northeast India, where Nagaland successfully developed indigenous language curricula for 14 tribal languages. These curricula explicitly incorporate traditional elements, including festivals, agricultural practices, and oral histories, establishing language as a vehicle for essential, domain-specific cultural knowledge. This approach recognizes that true IKS transfer demands that language and culture be embedded across all subjects for instance, designing math problems around local economic concepts or science lessons rooted in traditional ecological knowledge to make the language functionally relevant beyond the classroom.

A significant challenge to scaling this pedagogical model is the acute shortage of trained educators. The implementation of effective MTB-MLE is hampered by a lack of teacher knowledge regarding appropriate methodology, materials, and training needed to manage diverse classroom contexts. Furthermore, the availability of teaching personnel remains a critical operational obstacle even where curriculum development has succeeded, such as in Nagaland. The scalability of this strategic pedagogical pathway is thus fundamentally constrained by the capacity of the teacher workforce.

3.3. Global Models for Language Immersion and Reclamation

International case studies offer compelling blueprints for holistic language revitalization. Indigenous Language Immersion (ILI) schooling, which provides most or all instruction in the Indigenous language combined with a strong culture-based curriculum, has proven effective in supporting high language acquisition, academic achievement, and well-being for minoritized groups.

The Māori revitalization strategy in New Zealand demonstrates the necessity of community-led immersion starting from the earliest stages of life. The establishment of *Kōhanga Reo* (language nests) provides language classes for mothers with infants, ensuring the babies absorb the sounds and cadences of their tribal language from infancy.

The Māori experience underscores that revitalization transcends simple vocabulary instruction; it frames language promotion as a comprehensive sociopolitical project. *Te reo Māori* is explicitly linked to environmental knowledge, cultural values, and identity. Crucially, the process requires engaging the wider community, media, and private sector to understand the language's intrinsic value, rather than relying solely on slow governmental adoption. This demonstrates that preservation efforts must adopt a "National Language Pride Mission" to counter the common societal perception that tribal languages lack economic or social utility, ensuring widespread, sustained commitment.

Comparative Effectiveness of Language Revitalization Pedagogies:

Pedagogical Model	Core Strategy	Proven Benefits	Relevance to IKS Transfer	Indian/Global Example
MTB-MLE (Mother Tongue-Based Multilingual Education)	L1 instruction in early grades, bridging to L2/L3 later.	Higher academic achievement, increased retention, cumulative bilingualism.	Connects existing rich L1 knowledge (IKS/Culture) to complex L2 concepts, enhancing intellectual development.	Odisha Tribal Schools, Philippines, Vietnam
ILI (Indigenous Language Immersion)	Majority or all instruction occurs in the indigenous language, with culture-based curriculum.	High linguistic fluency, strong cultural identity, well-being benefits.	Direct, holistic transmission of ancestral knowledge embedded within the curriculum.	Hawaiian Immersion (ILI), Māori Kōhanga Reo

IV. Digital Frontiers: AI, NLP, and Bridging the Resource Gap

4.1. The Digital Divide and Resource Scarcity

The ambition to leverage technology for language preservation is significantly hampered by the "low-resource language problem." Advancements in Natural Language Processing (NLP) are overwhelmingly concentrated in high-resource languages, which possess abundant, well-annotated data. Indigenous Indian languages, spoken by smaller or marginalized communities, suffer from a scarcity of standardized tools and linguistic resources required for training complex, data-driven models. The effectiveness of even sophisticated Large Language Models (LLMs) diminishes significantly for extremely low-resource languages due to this lack of labelled data.

These languages often present unique technical hurdles, including complex grammatical structures and diverse vocabularies, which pose additional challenges for standard NLP techniques. Compounding the data scarcity is the physical "Digital Divide": limited access to technology and internet infrastructure in remote communities obstructs efforts to digitally document, archive, and transmit endangered languages, hindering

accessibility for the very communities they are intended to serve.

4.2. India's Technological Response and Indigenous LLMs

In response to the resource gap, India is executing a strategic program focusing on technological autonomy in linguistic AI. Under the India AI Mission, Sarvam is tasked with building India's sovereign LLM. This model is being developed from scratch using local infrastructure and Indian talent, designed specifically to be fluent in Indian languages and optimized for voice, serving as a vital step toward *Atmanirbhar Bharat* (Self-Reliant India). This focus confirms a governmental recognition that global, frontier models, primarily trained on high-resource data, are inadequate for addressing India's unique linguistic and IKS requirements.

The national roadmap emphasizes rapid progress in this domain. Key initiatives include the Adi-Vaani program for AI targeting Tribal Language Inclusion. By 2027, the plan includes launching India's first dedicated Tribal Large Language Model (LLM), which aligns with the goal of digitizing over 50 endangered languages and enabling AI-powered translation in sectors like healthcare, potentially benefiting 10 crore citizens by 2030.

Projects like BharatGen and Bhashini demonstrate the application of this strategy. BharatGen develops translation models for the 22 Scheduled Languages, directly utilizing the data generated by preservation schemes like SPPEL and Sanchika. Bhashini integrates AI-driven multilingual translation across government services, ensuring content delivery and seamless communication in the diverse linguistic landscape.

4.3. Innovative NLP Strategies for Resource Mobilization

To overcome the inherent data scarcity, innovative computational strategies are being adopted. Transfer learning, where models pre-trained on high-resource languages (such as Multilingual Bidirectional Encoder Representations from Transformers (Mbert) or Cross-Lingual Models,XLM-R) are adapted to low-resource settings, has shown significant promise.

Another crucial strategy is the mobilization of diverse, non-traditional data sources. Researchers are exploring integrating multimodal approaches, which combine textual data with rich media formats such as images, audio, or video, to enhance NLP performance where text data is scarce. This approach is essential for utilizing the rich oral traditions and cultural practices archived through platforms like Bharatavani.

Furthermore, the data collection process is increasingly reliant on community engagement. Researchers are actively collecting and curating datasets, utilizing community platforms to expand data resources. The success of digital archiving, however, hinges on community awareness, outreach to remote areas, and the ethical treatment of languages, knowledge, and cultural materials. The data collected via SPPEL and Sanchika must therefore be governed by stringent ethical frameworks, ensuring collaborative oversight by native speakers and respecting intellectual property rights.

There is a negative feedback loop where the digital divide (lack of connectivity) hinders data collection, which in turn limits the training of robust LLMs. If the resulting LLMs fail to generate truly useful, culturally sensitive content for remote communities, the divide is perpetuated. This structural challenge confirms that technological development (AI) and physical infrastructure investment (connectivity and hardware) must be simultaneously prioritized to ensure the success of the envisioned "Digital Kranti".

V. Strategic Challenges and Implementation Barriers

5.1. Socio-economic Drivers of Decline

The primary forces accelerating language endangerment are socio-economic. The incessant currents of globalization, urbanization, and socio-economic transformation pose existential challenges to most tribal languages. The "urban pull" motivates young people to relocate to metropolitan areas, which destabilizes the traditional context necessary for the oral transmission of IKS.

This migration is often reinforced by the marginalization of indigenous languages in contemporary life. The dominance of mainstream languages (both national and transnational) in mass media and technology limits the functional utility of tribal languages in modern contexts, making them less appealing and relevant to younger generations. This dynamic creates a vicious cycle where a lack of digital resources limits usage, reinforcing the perception of low socio-economic value.

5.2. Institutional and Political Hurdles

Despite strong national policy mandates, institutional support for indigenous languages often remains insufficient. The mainstream Indian education system routinely prioritizes dominant languages, leaving tribal languages with little or no formal support in the curriculum. Without formal education in their native tongues, children from these communities are at severe risk of losing their linguistic heritage.

Implementing progressive pedagogical models like MTB-MLE is severely hampered by practical constraints. Teachers often lack the requisite knowledge in methodology, materials, and training needed to effectively manage linguistically diverse classroom settings. Furthermore, the political dimension of language policy, demonstrated by resistance to the multilingual formula in certain states, highlights that central policy success is undermined by decentralized inertia and political friction.

The digital divide continues to limit the reach of national digital initiatives. Although platforms like Bharatavani host extensive resources, limited access to internet connectivity and technology obstructs digital documentation and transmission efforts, particularly hindering accessibility in remote regions where endangered languages are most prevalent. The technological solutions risk becoming symbolic successes rather than functional tools if the necessary last-mile infrastructure is not addressed.

5.3. Ethical Considerations and Sustainability

Digital documentation, while vital, presents ethical dilemmas. The ethical treatment of languages, indigenous knowledge, and cultural materials during archiving is a non-negotiable prerequisite for sustainable success. This demands active community awareness, robust outreach to remote areas, and explicit community engagement in the documentation and subsequent use of the data.

The specialized IKS embedded within indigenous languages extends beyond oral tradition to technical domains like crafts and traditional occupations. For instance, the knowledge related to crafts like Kondapalli Dolls, Bidriware, or specialized weaving is intrinsically linked to the language used to transmit technical instructions, material sourcing protocols, and specialized terminology. Therefore, the loss of language means the loss of the technical manual. Preservation programs must integrate specialized vocabulary and technical glossaries related to IKS domains alongside general linguistic documentation to ensure the continuity of these skills.

Finally, ensuring long-term sustainability requires political commitment and sustained funding beyond initial project launches. Global

revitalization models demonstrate that maintaining generational momentum requires overcoming public backlash and embedding language preservation within a broader framework of community wellness and social justice, rather than treating it merely as a short-term cultural project.

VI. Conclusion and Strategic Recommendations

6.1. Synthesis of Findings

The preservation of Indian languages is a national priority directly linked to the survival of Indigenous Knowledge Systems (IKS). The analysis confirms the severe vulnerability of languages, particularly those in the Tibeto-Burmese family, to extinction, which simultaneously threatens ancestral knowledge related to agriculture, medicine, and cultural identity (as exemplified by Ahirani).

The national policy structure is robust, underpinned by the IKS mandate of NEP 2020 and institutional efforts like SPPEL, Sanchika, and Bharatavani.

However, this architecture is constrained by significant implementation gaps: critical data scarcity for low-resource languages limits the efficacy of advanced AI (BharatGen/Tribal LLM); the digital divide restricts accessibility; and a profound shortage of specialized teachers undermines proven pedagogical models like MTB-MLE. Overcoming these barriers requires a coordinated, multi-sectoral strategy that treats linguistic preservation as both a critical data infrastructure project and a matter of social equity.

6.2. Action Agenda: High-Impact Interventions for IKS Transfer

6.2.1. Policy and Governance Recommendations (Institutional Strengthening)

1. **Mandate Functional IKS Integration and Curriculum Audit:** The Ministry of Education should mandate and fund comprehensive "whole-school approach" curriculum audits across relevant states, requiring the integration of indigenous knowledge into science, social studies, and commerce subjects, thereby making native languages functionally indispensable across the curriculum, not just in language class.
2. **Establish a Dedicated Teacher Corps for Multilingual Education:** A National Mission for Multilingual Education Teacher Training must be established to certify specialists in MTB-MLE methodology and IKS curriculum delivery. This mission must prioritize recruiting and training teachers from indigenous communities to address the acute operational challenge of teacher availability and proficiency in methodology.
3. **Enforce Ethical Data Governance and Community Reciprocity:** Implement mandatory ethical frameworks for all documentation undertaken by SPPEL and Sanchika. These protocols must explicitly guarantee community intellectual property rights and mandate collaborative oversight by native speakers, ensuring that data is not merely collected but is treated as a shared resource that benefits the source community.

6.2.2. Technological and Data Infrastructure Recommendations (Scaling AI)

1. **Accelerate High-Quality Data Pipelines:** Resource allocation for SPPEL and Sanchika must be significantly increased and prioritized, recognizing that these projects are the critical data backbone for the future of Sovereign AI. The focus must shift to rapidly generating high-quality, annotated, and culturally contextualized datasets required for

training the upcoming Tribal LLM (Adi-Vaani).

2. **Bridging the Last-Mile Digital Divide:** AI development programs (like Bhashini and BharatGen) must be coordinated with governmental programs providing subsidized digital hardware and internet connectivity in documented low-resource language zones. This dual investment is necessary to break the negative feedback loop between access and data scarcity, ensuring that digital resources are functionally accessible to remote communities.

3. **Invest in Multimodal NLP for Oral Traditions:** Substantial R&D investment should target multimodal NLP strategies that combine textual data with rich audio and video recordings. This approach leverages the wealth of oral tradition archived through initiatives like Bharatavani, overcoming the fundamental limitation of purely textual data scarcity for these languages.

6.2.3. Pedagogical and Community Recommendations (Revitalization)

1. **Scale Community Immersion Models:** Government and CIIL must actively support and scale community-led Indigenous Language Immersion (ILI) programs, drawing heavily on successful global models such as the Māori *Kōhanga Reo*. Priority should be given to establishing immersion environments during early childhood to ensure strong mother tongue fluency and robust cultural identity formation.

2. **Create Economic and Social Relevance:** Language revitalization efforts must be strategically partnered with indigenous economic sectors, such as traditional crafts, organic farming, and localized tourism. By integrating preserved linguistic and technical knowledge into viable economic opportunities, the preservation strategy counteracts the pressures of urbanization and migration, ensuring that the indigenous language retains functional relevance for younger generations.

3. **Frame Language as Wellness and Rights:** Policy communication must frame language promotion not as a mere cultural endeavour, but as an essential component of community well-being, equity, and the recognition of ancestral rights, leveraging the proven correlation between linguistic identity and improved socio-cultural outcomes.

Conclusion

India's future as a vibrant knowledge society hinge on its capacity to preserve, promote, and dynamically employ its remarkable linguistic diversity as a core interface for Indigenous Knowledge Systems (IKS). Endangered languages, particularly those belonging to smaller families like Tibeto-Burmese, face acute threats, with cascading consequences for cultural heritage, specialized expertise, and social equity. National initiatives including SPPEL, Bharatavani, and Sanchika have laid foundational frameworks for documentation, digital archiving, and AI-driven innovation, but persistent challenges remain related to implementation, teacher workforce, and the digital divide.

Pedagogical models such as Mother Tongue-Based Multilingual Education (MTB-MLE) and Indigenous Language Immersion programs reinforce the intergenerational transmission of heritage, language, and identity. Their effectiveness is amplified when supported by robust infrastructure, functional domain-specific vocabulary, and ethically governed data platforms. Technology, especially AI and Natural Language Processing, is increasingly

central, enabling scalable solutions for documentation, translation, and revitalization, provided that multimodal data strategies and last-mile connectivity are prioritized.

Systematic integration of indigenous knowledge into mainstream education and economic sectors ensures the functional relevance of native languages while promoting community wellness and social justice. Ethical stewardship, sustainable funding, and community engagement are non-negotiable prerequisites for success the loss of a language equates to the disappearance of irreplaceable reservoirs of wisdom and practice.

For India, the path forward demands coordinated action across governance, technology, and pedagogy. This effort must transcend symbolic gestures and result in lasting change embedding indigenous language preservation in formal education, digital resource ecosystems, economic participation, and policy communication. Only through such multipronged strategies can India secure both its linguistic legacy and the full breadth of wisdom contained within its Indigenous Knowledge Systems for generations to come.

References

[1]. Press Information Bureau. (n.d.). *Documenting India's endangered languages*. https://www.pib.gov.in/FeaturesDeatils.aspx?NoteId=155013&ModuleId=2

[2]. Central Institute of Indian Languages. (n.d.). *Schemes, projects, regional centres, and classical centres*. https://www.ciil.org/download/Schemes_Projects_Regional_Centres_and_Classical_Centres.pdf

[3]. Wikipedia. (n.d.). *List of endangered languages in India*. https://en.wikipedia.org/wiki/List_of_endangered_languages_in_India

[4]. Patil, A. (2024). *Ahirani and the Indian Knowledge System: A case study on the role of regional languages in cultural preservation* [PDF]. ResearchGate. https://www.researchgate.net/publication/390532376_Ahirani_and_the_Indian_Knowledge_System_A_Case_Study_on_the_Role_of_Regional_Languages_in_Cultural_Preservation

[5]. Chinsembu, K. C., et al. (2024). *The role of indigenous knowledge in advancing the therapeutic use of medicinal plants: Challenges and opportunities*. *Journal of Ethnobiology and Ethnomedicine*, 20(XX). https://pmc.ncbi.nlm.nih.gov/articles/PMC11633201/

[6]. Ekka, A. (2016). *Language preservation and its socio-cultural impact in tribal communities of India*. *International Journal of Research in Social Sciences, Management and Literature*, 1(3), 10–20. https://ijrsml.org/wp-content/uploads/2025/04/ijrsml_July-2016_Language-Preservation-and-Its-Socio-Cultural-Impact-in-Tribal-Communities-of-India-10-20.pdf

[7]. Singh, R. (2023). *Integrating India's traditional wisdom: Exploring the role of NEP-2020 in the Indian Knowledge System*. *Education Quarterly*, 15(3). https://ndpublisher.in/admin/issues/EQv15n3d.pdf

[8]. Wikipedia. (n.d.). *National Education Policy 2020*. https://en.wikipedia.org/wiki/National_Education_Policy_2020

[9]. Institute for Industrial Strategy, Policy, and Public Research. (2023). *Decoding NEP 2020: Analyzing the prospects and challenges*. https://iisppr.org.in/decoding-nep-2020-analyzing-the-prospects-and-challenges/

[10]. Sharma, S., & Singh, D. (2024). *Preservation and revitalisation of Indian languages through digital technologies*. *DESIDOC Journal of Library & Information Technology*. https://publications.drdo.gov.in/ojs/in

dex.php/djlit/article/download/20138/8496/88553

[11]. Press Information Bureau. (2024). *22 languages, digitally reimagined*. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2182427

[12]. The Economic Times. (2024). *India's multilingual AI model for the world*. https://m.economictimes.com/news/india/indias-multilingual-ai-model-for-the-world/articleshow/124331987.cms

[13]. McCarty, T. L. (2018). *Indigenous language revitalization*. *Oxford Research Encyclopedia of Education*. https://oxfordre.com/education/display/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-559

[14]. Mohanty, A. K. (2019). *The effectiveness of mother tongue–based multilingual education (MTB-MLE)* (Master's thesis). Syracuse University. https://surface.syr.edu/cgi/viewcontent.cgi?article=1150&context=eli

[15]. Mishra, S. (2022). *Mother tongue–based multilingual education (MTB-MLE) programme: A panacea for tribal learners*. *South Eastern European Journal of Public Health*. https://www.seejph.com/index.php/seejph/article/download/5714/3787/8700

[16]. American Indian Language Development Institute. (n.d.). *What we offer*. https://www.aildi.arizona.edu/nalrc-resources

[17]. UNESCO Associated Schools Network. (2023). *Teaching and learning Indigenous languages: Inspiring practices*. York University. https://www.yorku.ca/unsdgs/toolkit/wp-content/uploads/sites/617/2023/02/Teaching-and-Learning-Indigenous-Languages-Inspiring-Practices.pdf

[18]. Singh, P., & Das, M. (2024). *Indigenous knowledge systems and sustainable learning: A comprehensive regional analysis of Northeast India*. ResearchGate. https://www.researchgate.net/publication/394759329_Indigenous_Knowledge_Systems_And_Sustainable_Learning_A_Comprehensive_Regional_Analysis_Of_Northeast_India

[19]. Reyhner, J. (n.d.). *Revitalizing Indigenous languages: Introduction*. Northern Arizona University. https://jan.ucc.nau.edu/~jar/RIL_Intro.html

[20]. Graham-McLay, C. (2022, September 12). *As Māori language use grows in New Zealand, the challenge is to match deeds to words*. *The Guardian*. https://www.theguardian.com/world/2022/sep/12/as-maori-language-use-grows-in-new-zealand-the-challenge-is-to-match-deeds-to-words

[21]. Dossou, B., & others. (2023). *Natural language processing applications for low-resource languages*. *Natural Language Processing*. Cambridge University Press. https://www.cambridge.org/core/journals/natural-language-processing/article/natural-language-processing-applications-for-lowresource-languages/7D3DA31DB6C01B13C6B1F698D4495951

[22]. Gupta, R., & Rao, A. (2025). *Exploring NLP benchmarks in an extremely low-resource setting*. arXiv. https://arxiv.org/html/2509.03962v1

[23]. Sarvam AI. (2024). *India's sovereign LLM*. https://www.sarvam.ai/blogs/indias-sovereign-llm

[24]. TSS Review. (2024). *The extinction of tribal languages in India: A socio-linguistic and cultural loss*. https://tssreview.in/wp-content/uploads/2024/10/27-1.pdf

[25]. Travel.Earth. (n.d.). *Experience Indian handicrafts with 29 traditional crafts of India*. https://travel.earth/traditional-crafts-of-india-you-need-to-experience/amp/