



22 Languages, Digitally Reimagined

Unlocking India's Linguistic Future Through Technology

25 October, 2025

"A language is simply not just a mode of communication, it is the soul of a civilization, it's culture, it's heritage."
-Prime Minister Narendra Modi

Key Takeaways

- Support for all 22 scheduled languages through AI platforms like Bhashini and BharatGen.
- Digitized language data from SPPEL (Scheme for Protection and Preservation of Endangered Languages) and Sanchika enrich AI model training for multilingual solutions.
- Tech-driven initiatives position India as a global leader in multilingual digital transformation.

Introduction

India's linguistic landscape is among the most diverse in the world, with 22 Scheduled Languages and¹ hundreds of tribal and regional dialects spoken across its vast geography. As digital transformation accelerates, the need to embed this linguistic diversity into digital infrastructure has become critical. Technology is no longer just a medium of communication; it is the backbone of inclusion.

The Government of India is leveraging advanced technologies such as Artificial Intelligence (AI), Natural Language Processing (NLP), machine learning, and speech recognition to build intelligent, scalable language solutions. These initiatives aim to democratize access to digital services by enabling seamless communication, real-time translation, voice-enabled interfaces, and localized content delivery. By building a robust technological ecosystem that respects linguistic diversity, India is setting the foundation for an inclusive digital future where every citizen, regardless of their mother tongue, can participate fully in the digital economy and governance.



¹ <https://rajbhasha.gov.in/en/languages-included-eighth-schedule-indian-constitution>

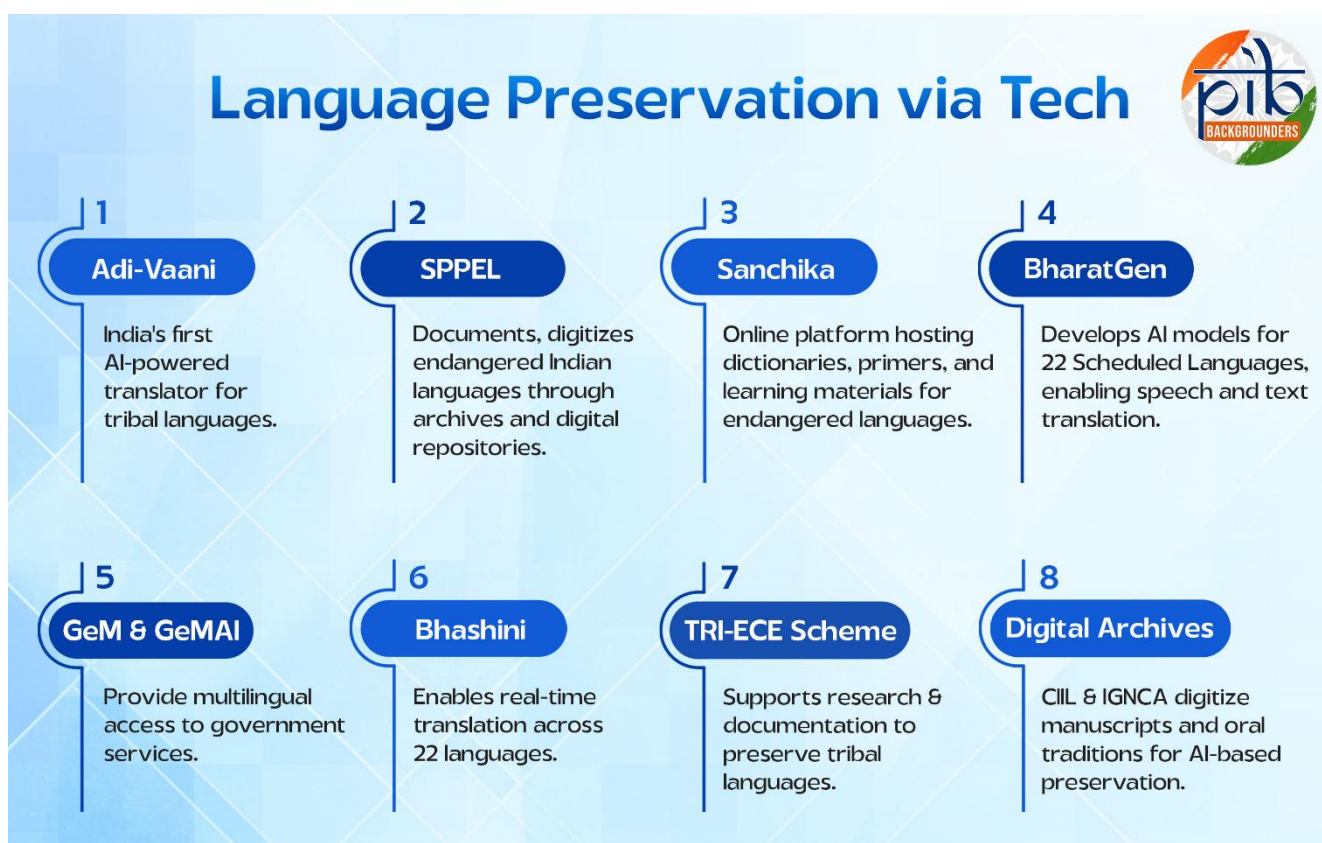
Key platforms driving linguistic inclusion

AI-driven language platforms and expansive digital repositories are reimagining how India's languages are preserved, used, and evolved. Platforms like Bhashini and BharatGen offer multilingual support across governance, healthcare, and education. Initiatives like Adi-Vaani bring tribal languages into the digital fold. This integration ensures that India's linguistic heritage is not only preserved but made functional and dynamic in the digital era.

Over the past decade, advances in artificial intelligence, natural language processing, and digital infrastructure have accelerated efforts to document, digitize, and revitalize India's linguistic diversity. These technologies have enabled large-scale language data collection, automated translation, and speech recognition across hundreds of languages and dialects, many of which were previously underserved. This technological momentum has helped bridge communication gaps, promote inclusive governance, and empower communities by making digital content accessible in their native languages.

Adi-Vaani: AI for Tribal Language Inclusion

Founded in 2024², Adi-Vaani is India's first AI-driven platform dedicated to the real-time translation and preservation of tribal languages. Designed to revolutionize communication through cutting-edge language technologies, Adi-Vaani combines the precision of artificial intelligence with human linguistic expertise to deliver seamless multilingual experiences.



At its core, Adi-Vaani leverages advanced speech recognition and natural language processing (NLP) to support languages such as Santali, Bhili, Mundari, and Gondi—many of which have traditionally relied on oral transmission and lacked sufficient digital representation. By enabling real-time translation between tribal languages and major Indian languages, the platform not only preserves

² <https://adivaani.tribal.gov.in/>

these rich linguistic traditions but also makes them accessible for education, governance, and cultural documentation.

Scheme for Protection and Preservation of Endangered Languages (SPPEL)

The **Scheme for Protection and Preservation of Endangered Languages (SPPEL)**, launched in **2013** by the **Ministry of Education** and implemented by the **Central Institute of Indian Languages (CIIL), Mysuru**,³ focuses on documenting and digitally archiving endangered Indian languages—particularly those spoken by fewer than **10,000 people**.

It generates rich **text, audio, and video datasets** that serve both preservation and innovation, providing critical resources for **AI and Natural Language Processing (NLP)** systems. Platforms like **Sanchika**, CIIL's digital repository, support **AI model training**, machine translation, and the development of culturally rooted language technologies.

Sanchika: Digital Repository for Indian Languages

Managed by Central Institute of Indian Languages, **Sanchika** aggregates dictionaries⁴, primers, storybooks, and multimedia resources for Scheduled and tribal languages. This centralized digital archive is a vital data source for training language models, developing translation systems, and preserving cultural narratives.

The platform offers linguistically categorized digital resources including text, audio, and visual materials—that aid academic research, language education, and cultural documentation. These rich and diverse collections provide foundational datasets for emerging AI and natural language processing applications, enabling more inclusive and effective digital tools for low-resource tribal and regional languages.

BharatGen: AI Models for Indian Languages

BharatGen develops advanced text-to-text and text-to-speech translation models for all 22 Scheduled Languages. It leverages data from **SPPEL** and **Sanchika** to create multilingual AI systems that power applications in governance, education, and healthcare — ensuring that digital content is accessible in every major Indian language.

BharatGen's multilingual AI systems are designed to enhance digital accessibility and inclusivity across sectors such as **governance, education, and healthcare**, enabling seamless communication and content delivery in India's diverse linguistic landscape.

GeM and GeMAI: AI-Powered Multilingual Assistant for Government e-Marketplace

The Government e-Marketplace (GeM) is India's digital platform⁵ for public procurement, launched on **August 9, 2016**, by the Ministry of Commerce and Industry. GeM streamlines the purchasing process for government departments and public sector entities, ensuring transparency and efficiency.

To enhance user accessibility and inclusivity, GeM has integrated **GeMAI**, an AI-powered multilingual assistant. GeMAI leverages advanced natural language processing (NLP) and machine learning to provide voice and text-based support across multiple Indian languages. This enables users to search, navigate, and complete transactions on the platform more easily, helping to overcome language barriers in government procurement.

Bhashini: AI-Driven Multilingual Translation for Inclusive India

³ <https://www.sppel.org/about-us.aspx>

⁴ <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2150806>

⁵ <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2129647>

Bhashini, under the National Language Translation Mission (NLTM), is a pioneering AI platform enabling real-time translation for 22 Scheduled Languages and tribal languages. It facilitates access to government services, digital content, and promotes digital inclusion through machine translation, speech recognition, and natural language understanding.

Key achievements:

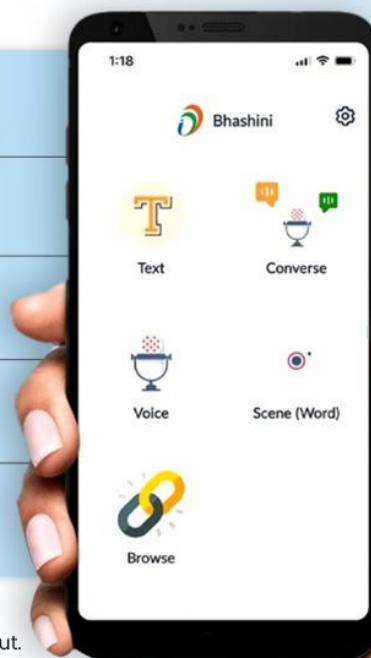
- Integration with Tripura CM's Helpline, eVidhan, Kisan Sahayata App for local language interaction.
- Tribal language models for Mizoram's Mizo, Hmar, Chakma languages.
- Real-time multilingual announcements at Maha Kumbh 2025.
- Sansad Bhashini for AI-powered parliamentary debate translations and citizen engagement.

(Data as on 14-10-2025)



Bhashini's Milestones at a Glance

1 Million+	Mobile App Downloads
350+	Total AI Models
3 Billion+	Total Inferencing*
295 Million	Translation
20+	Languages Supported



*Inferencing: Process where an AI model makes real-time predictions based on new input.

Source: bhashini.gov.in

Tribal Research, Information, Education, Communication and Events (TRI-ECE) Scheme

The **Tribal Research, Information, Education, Communication and Events (TRI-ECE)** scheme, under the **Ministry of Tribal Affairs**, supports innovative research and ⁶documentation projects aimed at preserving tribal languages and cultures. As part of this initiative, the Ministry has backed the development of **AI-based language translation tools** capable of converting **English/Hindi text and speech** into tribal languages and vice versa.

These tools integrate **machine learning, speech recognition, and natural language processing (NLP)** to support real-time translation and digital preservation of endangered tribal languages. The project

⁶ <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2159087>

also emphasizes community involvement through collaboration with Tribal Research Institutes and language experts, ensuring linguistic accuracy and cultural sensitivity.

Digital Archives and Academic Efforts

Institutions like the Central Institute of Indian Languages (CIIL) and the Indira Gandhi National Centre for the Arts (IGNCA) collaborate with Bhashini by digitizing ancient manuscripts, folk literature, and oral traditions. These digital archives enrich AI and Natural Language Processing (NLP) systems, supporting both preservation and state-of-the-art translation solutions- reinforcing the link between cultural heritage and modern technology.

Empowering Education Through AI-Driven Multilingual Platforms

Artificial Intelligence is transforming India's education landscape by making learning more inclusive, accessible, and linguistically diverse. The integration of AI-based language technologies is driving the vision of the National Education Policy (NEP) 2020, which emphasizes instruction in the learner's home language, mother tongue, or regional language—at least up to Grade 5 and preferably till Grade 8 and beyond.

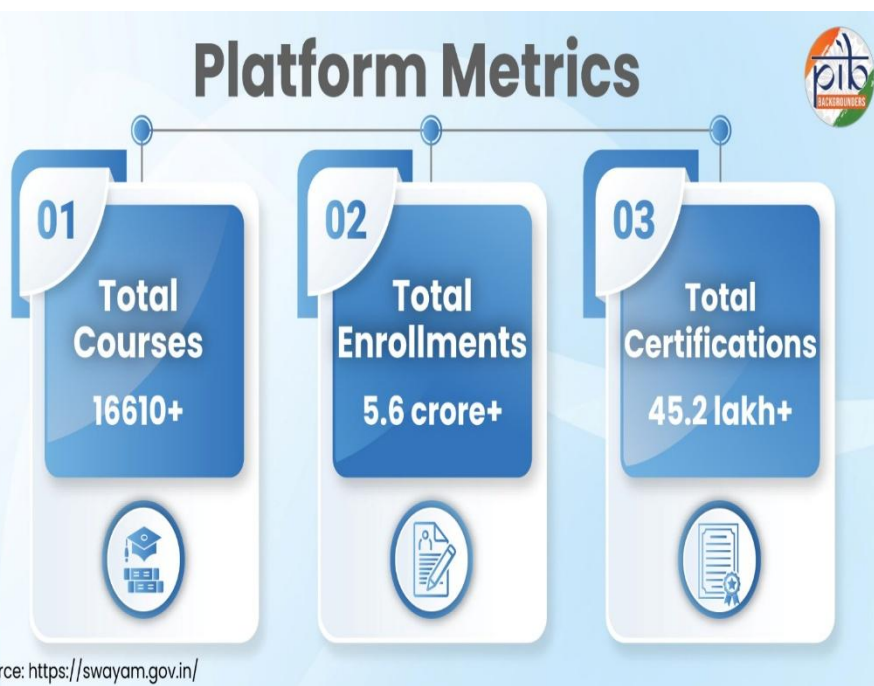
At the institutional level, the **AICTE's Anuvadini app**, an indigenous AI-based multilingual translation tool, enables rapid translation of engineering, medical, law, undergraduate, postgraduate, and skill-development books into Indian languages.⁷ The translated content is hosted on the **e-KUMBH portal**, expanding access to technical knowledge in native tongues.

What is e-Kumbh portal?

e-KUMBH portal is an AICTE platform that provides free access to technical books and study materials in multiple Indian languages, supporting NEP 2020's vision of education in the mother tongue.

Complementing these AI-driven initiatives are long-standing national efforts such as the **National Translation Mission (NTM)**—⁸which facilitates the translation of knowledge texts into Indian languages—and the **National Mission on Manuscripts (NMM)**, which preserves and digitizes India's ancient scholarly works. Together, they build a continuum between India's linguistic heritage and its future-ready, AI-enabled education ecosystem.

Meanwhile, platforms such as **SWAYAM** provide the digital backbone for multilingual content delivery. As of mid-2025, over **5 crore learners** have enrolled on SWAYAM⁹, while the government has directed that **all school and higher-education textbooks and study materials** be made digitally available in



⁷ <https://www.pib.gov.in/PressRelea>

⁸ <https://ntm.org.in/languages/engli> Source: <https://swayam.gov.in/>

⁹ <https://swayam.gov.in/>

Indian languages within the next **three years**.¹⁰

Together with language-AI platforms such as **Bhashini**, these initiatives enable schools, ed-tech firms, and higher-education institutions to deliver localized learning materials, interactive tools, and teacher-aids in native languages—bridging linguistic divides, improving comprehension, and empowering every learner to access digital education in their mother tongue.

This emerging **multilingual digital education framework** not only strengthens educational inclusion but also reinforces India's linguistic diversity—ensuring that the nation's many languages remain **living, functional media of instruction, knowledge, and innovation**, rather than mere cultural relics.

Technology behind transformation

India's multilingual digital ecosystem is powered by advanced AI and computational linguistics technologies designed specifically for its linguistic diversity. By harnessing cutting-edge innovations, these technologies not only preserve linguistic heritage but also enable seamless, real-time communication across diverse languages, fostering digital inclusion at scale.

Key components of this ecosystem include:

- **Automatic Speech Recognition (ASR):** Converts diverse spoken Indian languages into accurate text¹¹, enabling voice-based applications, command interfaces, and real-time transcription services.
- **Text-to-Speech (TTS):** Synthesizes natural, intelligible speech outputs in native languages, enhancing accessibility in digital assistants, educational tools, and government services.
- **Neural Machine Translation (NMT):** Employs deep learning models to provide context-aware, real-time translations between multiple Indian languages¹², overcoming syntactic and semantic complexities.
- **Natural Language Understanding (NLU):** Facilitates AI systems in interpreting user intent, sentiment, and context within native languages, improving conversational agents and user interaction.
- **Transformer-based Architectures (IndicBERT, mBART):** These state-of-the-art models are pre-trained on massive multilingual Indian language corpora, enabling higher accuracy in language modeling, translation, and understanding tasks.¹³
- **Corpus Development and Data Curation:** Extensive datasets are compiled from digitized manuscripts, folklore, oral traditions, government records, and educational content, providing rich, representative data to train and fine-tune AI models for India's varied linguistic landscape.

This technological backbone drives platforms like Bhashini, BharatGen, and Adi-Vaani, ensuring scalable, accurate, and inclusive language technologies tailored for India's unique multilingual context.

Conclusion

¹⁰ <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1997882>

¹¹ <https://dic.gov.in/bhashini>

¹² <https://digital.gov/event/2019/10/15/using-neural-machine-translation-for-multilingual-communication>

¹³ https://aikosh.indiaai.gov.in/home/models/details/ai4bharat_indicbert_multilingual_language_representation_model.html

India's future in language preservation is powered by cutting-edge technology, integrating AI and digital archives to keep its rich linguistic heritage vibrant and accessible. Platforms like Bhashini, BharatGen, and Adi-Vaani, along with targeted initiatives such as SPPEL and TRI-ECE, empower citizens nationwide to engage with services in their native languages. This comprehensive approach not only safeguards India's cultural diversity but also drives inclusive digital growth, positioning the country as a global leader in multilingual innovation.

References

Press Information Bureau

- <https://www.pib.gov.in/PressReleaseDetailm.aspx?PRID=2065892>
- <https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2082387>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2129647>
- <https://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/aug/doc2025812605201.pdf>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1997882>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2153187>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2159087>
- <https://www.pib.gov.in/FeaturesDeatils.aspx?ModuleId=2&NotelId=155013>
- <https://www.pib.gov.in/PressReleaselframePage.aspx?PRID=1777333>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=1577804>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2159096>
- <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2162278>
- <https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=1883839>

Digital.gov

- <https://digital.gov/event/2019/10/15/using-neural-machine-translation-for-multilingual-communication>

Ministry Of Home Affairs

- <https://rajbhasha.gov.in/en/languages-included-eighth-schedule-indian-constitution>

Ministry of Electronics & Information Technology

<https://dic.gov.in/bhashini>

https://aikosh.indiaai.gov.in/home/models/details/ai4bharat_indicbert_multilingual_language_representation_model.html

Ministry Of Tribal Affairs

<https://adivaani.tribal.gov.in/>

Ministry Of Education

<https://swayam.gov.in/>

SK/SA