



Digital India BHASHINI Division Showcases Multilingual AI Solutions for Accessible Research at ETD-2026, IIT Delhi

BHASHINI explores collaboration with higher education institutions and research repositories to expand access to theses, dissertations and scholarly content across Indian languages

Posted On: 24 OCT 2026 8:45PM by PIB Delhi

The **Digital India BHASHINI Division (DIBD)**, under **Digital India Corporation (DIC)**, **Ministry of Electronics and Information Technology (MeitY)**, Government of India, participated in the 29th International Symposium on Electronic Theses and Dissertations (ETD-2026), held at the Indian Institute of Technology (IIT) Delhi from 23 to 25 October 2026. Ms. **Baishali Chakraborty, speaking on behalf of DIBD**, conducted a session on BHASHINI's multilingual artificial intelligence capabilities and their potential to make scholarly knowledge more accessible across languages.



Titled “*Workshop on AI BHASHINI: Bridging Divides and Increasing Productivity*,” the session highlighted how BHASHINI’s **Language-First Digital Public Infrastructure (DPI)** can support higher education institutions, libraries and research repositories in addressing language barriers and expanding access to academic content.



During the session, Ms. Chakraborty **demonstrated** how **voice-first multilingual AI, speech-to-text, translation, optical character recognition (OCR) and document translation** can support multilingual access in academic and research settings. A live

demonstration of **Shrutlekh** showcased real-time speech transcription with simultaneous translation, illustrating its potential applications in lecture halls, research seminars, conferences and academic discussions.

The session also highlighted BHASHINI's broader multilingual AI ecosystem, which brings together language datasets, AI models, APIs and applications to support a range of language technology use cases. BHASHINI has recorded **over 10 billion cumulative AI inferences** and supports **36 Indian languages for text, 23 Indian languages for voice and 35 international languages**. Its website translation plugin enables multilingual access across **more than 800 government websites**.

A key focus of the session was the potential to extend **AI-powered translation to electronic theses and dissertations (ETDs)**, including lengthy and specialised academic documents. **BHASHINI Pragati**, a document translation workbench developed for the Comptroller and Auditor General of India (CAG), was presented as an example of structured document translation and review. The workbench supports the processing of large documents, including those spanning **100 to 500 pages**, with quality checks and data security considerations.

The discussion explored how similar workflows could be adapted for research repositories such as **Shodhganga, maintained by INFLIBNET**, as well as universities and other higher education institutions. Such solutions could help make theses, dissertations, research reports and other scholarly material available across Indian languages while preserving document structure, specialised terminology and meaning.

Speaking on the occasion, Harshad Sharma, Senior Manager (Finance), Digital India BHASHINI Division (DIBD), said, "Making research and scholarly knowledge accessible across Indian languages is an important step towards a more inclusive knowledge ecosystem. BHASHINI's multilingual AI capabilities, including document translation and speech technologies, can help universities, libraries and research repositories reach wider audiences. Engagement with academic institutions will be key to understanding their requirements and developing reliable solutions for multilingual access to theses, dissertations and other research content."

During the interaction, participants raised questions about translating PDF documents on the fly and the reliability of AI translation for specialised academic and scientific content. Ms. Chakraborty discussed the importance of combining domain-specific machine translation with human review and quality assurance. The discussion also explored opportunities to work with research organisations to understand their requirements and develop customised solutions suited to their repositories and academic domains.

The session further covered BHASHINI's capabilities in **OCR for extracting text from scanned documents and images, terminology management through SHABD UDYAN, and transliteration** to help preserve technical terms, names and expressions across scripts. These capabilities can support workflows that improve the accessibility and discoverability of research content while accounting for the linguistic and formatting requirements of academic documents.

Beyond translation, the session highlighted **BhashaDaan**, BHASHINI's citizen-contribution initiative, which helps build and enrich language resources through speech and text contributions, audio transcription, and text extraction from images and documents. **Granthika** was also presented as a tool for digitising and preserving books, manuscripts, literary works, folk tales and oral histories, contributing to the development of digital resources for India's multilingual knowledge ecosystem.

BHASHINI invited higher education institutions, research repositories, libraries and academic organisations to explore collaboration in identifying priority use cases, developing domain-specific language resources and adapting translation workflows to their needs. Such engagement can help address repository-specific requirements, strengthen translation quality through human validation and expand multilingual access to India's scholarly output.

BHASHINI's participation at ETD-2026 reflects its broader objective of advancing language technology as a public digital resource. Engagement with academic institutions and research repositories offers an opportunity to extend multilingual AI capabilities to scholarly communication, helping research and knowledge reach audiences across linguistic boundaries.



About Digital India BHASHINI Division (DIBD)

The Digital India BHASHINI Division (DIBD), under Digital India Corporation (DIC), Ministry of Electronics and Information Technology (MeitY), Government of India, is advancing India's Language-First Digital Public Infrastructure to enable multilingual access to digital services. BHASHINI brings together language datasets, AI models, APIs and applications to support government departments, developers, researchers and other stakeholders in building multilingual AI solutions. It powers more than 800 government websites, has enabled over 10 billion cumulative AI inferences and supports 36 Indian languages for text, 23 Indian languages for voice and 35 international languages, advancing language technology across India's digital ecosystem.

Vinod Kumar / Kanishk Sharma

Share



829 words



3 min read