

GOVERNMENT OF INDIA
MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY
LOK SABHA
UNSTARRED QUESTION. NO. 4215
TO BE ANSWERED ON: 18.03.2026

BHASHINI

4215. SHRI BAIJAYANT PANDA:

Will the Minister of ELECTRONICS AND INFORMATION TECHNOLOGY be pleased to state:

- (a) the number of Indian languages supported under BHASHINI and the nature of services provided;
- (b) the initiatives undertaken to scale up BHASHINI through engagement with startups, academia and State Governments; and
- (c) the steps taken by the Government to strengthen language datasets and improve accuracy and inclusiveness?

ANSWER

MINISTER OF STATE FOR ELECTRONICS AND INFORMATION TECHNOLOGY
(SHRI JITIN PRASADA)

(a) to (c): Government of India's digital policies follow Prime Minister's vision of democratising technology to empower citizens. Bhashini is Government's initiative aimed at building a national digital public infrastructure for Indian languages.

At present, the Bhashini platform supports text-to-text translation across 36 Indian languages and speech-to-speech translation across 23 Indian languages.

The platform provides various AI-enabled language technology services, including Machine Translation (MT), Automatic Speech Recognition (ASR), Text-to-Speech (TTS), Optical Character Recognition (OCR).

Government has adopted a collaborative and multi-stakeholder approach involving startups, academic institutions, and State Governments. Startups and MSMEs are encouraged to develop innovative multilingual applications and sector-specific solutions using BHASHINI's open APIs.

Academic and research institutions across the country are engaged for research, dataset preparation, linguistic validation, and development of AI models. Further, the Mission collaborates with Central Ministries, State Governments and State Language Missions to identify domain-specific datasets and enable integration of BHASHINI services into citizen-facing government platforms.

A national repository of language datasets and AI models has been created through the Unified Language Contribution API (ULCA), which enables standardised submission, validation and cataloguing of datasets, models and tools.

Further, crowdsourcing initiatives such as BhashaDaan and other community participation programmes are undertaken to collect speech and text data from diverse linguistic communities, including speakers of regional and lesser-resourced languages.
