GOVERNMENT OF INDIA
MINISTRY OF EDUCATION
DEPARTMENT OF HIGHER EDUCATION
LOK SABHA
STARRED QUESTION NO. 34
ANSWERED ON 02.02.2026

PROMOTION OF SCHEDULED INDIAN LANGUAGES THROUGH AI LANGUAGE PLATFORMS

*34. SHRI SHASHANK MANI:
SHRI MANOJ TIWARI:

Will the Minister of EDUCATION be pleased to state:

(a) the details of initiatives undertaken by the Government to support all 22 scheduled Indian languages through AI-based language platforms such as Bhashini and BharatGen;
(b) the scope of digitisation of linguistic data, development of multilingual tools and creation of open language datasets under these initiatives;
(c) the manner in which academic institutions, startups and public sector agencies across the country have been involved in the development and deployment of these language technologies; and
(d) whether the Government proposes to expand the coverage of the platforms and include the Tulu language, spoken predominantly in Dakshina Kannada and Udupi districts of Karnataka, under these Al-based language platforms for education, governance and public service delivery, and if so, the details and timelines thereof?

ANSWER

MINISTER OF EDUCATION
(SHRI DHARMENDRA PRADHAN)

(a) to (d): A Statement is laid on the table of the House.

*****

STATEMENT REFERRED TO IN REPLY TO PARTS (A) to (D) IN RESPECT OF LOK SABHA STARRED QUESTION NO. 34 ANSWERED ON 02.02.2026 REGARDING "PROMOTION OF SCHEDULED INDIAN LANGUAGES THROUGH AI LANGUAGE PLATFORMS" ASKED BY SHRI SHASHANK MANI & SHRI MANOJ TIWARI", HON'BLE MEMBERS OF PARLIAMENT

(a) to (d): The National Education Policy (NEP) 2020 highlights the importance of multilingualism and places strong emphasis on the promotion of all Indian languages. In alignment with the objectives of NEP 2020, the Government of India has undertaken several initiatives promoting education, preservation and research on Indian languages. These efforts have further been augmented through the adoption of Artificial Intelligence and Machine Learning (AI/ML) technologies, including the development of Large Language Models, translation tools and language-enabled digital services in Indian languages.

Government of India has started the BharatGen project, which is a multimodal Large Language Model Project focused on developing efficient and inclusive AI solutions that support all 22 Scheduled languages and enable the creation of a robust digital AI infrastructure for India's unique socio-cultural context and diverse sectors. BharatGen is anchored at IIT Bombay, where the core development of language technologies is undertaken. The initiative is implemented through a consortium of leading academic institutions including IIT Kanpur, IIT Madras, IIT Hyderabad, IIIT Hyderabad, IIM Indore, and IIT Mandi.

Further, to facilitate the translation of content into Indian languages, technological advancements include the development of AI-based translation tools such as ANUVADINI by the All India Council of Technical Education (AICTE) and BHASHINI, an initiative under the Digital India Programme.

Through a collaboration of over 70 research partner institutes, BHASHINI has developed state-of-the-art AI models for Indian languages. At present, BHASHINI platform hosts a repository of over three hundred fifty AI-based language models and provides more than twenty-two specialised language services. These services include Automatic Speech Recognition, Machine Translation, Text-to-Speech, Optical Character Recognition, Transliteration and several other language technology capabilities essential for enhancing multilingual digital access in the country. The National Language Translation Mission through the BHASHINI platform is digitising large volumes of text and speech data across all 22 scheduled languages. BHASHINI follows a multi-stakeholder participation model involving academic and research institutions for fundamental research, dataset preparation and development of baseline AI models. For instance, IIT Madras has leveraged Bhashini and works closely with other academic partners on research and model development, while startups and technology companies build applications and solutions on top of the Bhashini language stack.

The Indian Language Technology Stack for education initiative of the IIT Madras Bodhan AI Foundation aims to provide technology solutions for making learning more accessible, inclusive, and effective. These efforts focus on delivering educational content and services in Indian languages to improve access to quality education in mother tongue through AI enabled tools.

Further, extensive linguistic resources for scheduled Indian Languages have been developed by the Central Institute of Indian Languages (CIIL) of the Ministry of Education under the Linguistic

Data Consortium for Indian Languages (LDC-IL) scheme. Since 2019, LDC-IL has provided resources to government agencies, government-promoted initiatives, researchers, and commercial and industrial users engaged in developing language technologies.

Under the LDC-IL scheme, 76 datasets covering all Scheduled languages have been released. Apart from the datasets, LDC-IL has also developed web-based applications and tools to support research and development in Indian languages. Several of these tools, hosted at CIIL data centres, are accessible through the Medha Bhashika website https://medha.ciil.org. Expanding the availability of high-quality digital resources for additional Indian languages, LDC-IL has taken steps for digitizing the corpora of the Tulu language.

***