

GOVERNMENT OF INDIA
MINISTRY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF SCIENCE AND TECHNOLOGY
RAJYA SABHA
STARRED QUESTION NO. 63
ANSWERED ON 05.02.2026

ROLE OF BHARATGEN AI

*63. SHRI BHUBANESWAR KALITA:

Will the Minister of SCIENCE AND TECHNOLOGY be pleased to state:

- (a) the number of Indian languages which are currently supported by BharatGen's AI models and the roadmap for full language coverage;
- (b) the specific applications of BharatGen in the healthcare, agriculture, education and governance sectors;
- (c) the number of Technology Innovation Hubs and Translational Research Parks currently active under the BharatGen network; and
- (d) the institutions which are part of the BharatGen consortium and what role it plays in its development?

ANSWER

MINISTER OF STATE (INDEPENDENT CHARGE) FOR THE
MINISTRY OF SCIENCE AND TECHNOLOGY & EARTH SCIENCES
(DR. JITENDRA SINGH)

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

STATEMENT AS REFERRED IN REPLY TO PARTS (a) TO (d) OF RAJYA SABHA
STARRED QUESTION NO. 63 FOR 05.02.2026 REGARDING “ROLE OF
BHARATGEN AI”

(a) to (d): BharatGen is the first government supported national initiative to develop a range of sovereign foundational AI models tailored to Indian languages and societal contexts. It spans multiple modalities, including text (via Large Language Models), speech (Text-to-Speech and Automatic Speech Recognition), and vision-language systems.

Currently, BharatGen’s AI models support 15 Indian languages which include Hindi, Assamese, Bengali, Gujarati, Kannada, Maithili, Malayalam, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil and Telugu. Soon, all 22 scheduled Indian languages will be covered.

BharatGen has released domain specific fine-tuned models for Ayurveda (Ayur Param), Indian agriculture (Agri Param) and Indian legal domain (Legal Param). In addition, all BharatGen models (text, speech and vision) are useful for applications across healthcare, agriculture, education and governance.

Two Technology Innovation Hubs namely TIH Foundation for IoT and IoE, IIT Bombay and IITM Pravartak Technologies Foundation, IIT Madras under National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) of Department of Science and Technology (DST) are currently active as part of BharatGen network.

The following institutions are a part of the BharatGen consortium:

Institution Name	Role in BharatGen
Indian Institute of Technology, Bombay	Lead institution, guiding research and integration across consortium partners
International Institute of Information Technology, Hyderabad	Vision-language document modeling
Indian Institute of Technology, Madras	Speech foundation model development and evaluation
Indian Institute of Technology, Kanpur	Legal AI research, domain-specific datasets, and developing tokenization strategies for multilingual models
Indian Institute of Technology, Hyderabad	Advanced tokenization and vocabulary optimization for large multilingual LLMs
Indian Institute of Technology, Mandi	Inclusive multilingual model development and research on efficient training strategies for LLMs
Indian Institute of Management, Indore	Bharat-centric evaluation and benchmarking of LLMs, multilingual and multimodal data collection
