

**GOVERNMENT OF INDIA
MINISTRY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF SCIENCE AND TECHNOLOGY
LOK SABHA**

STARRED QUESTION No. *257

ANSWERED ON 06/08/2025

BHARATGEN AI MODELS

***257. SHRI JANARDAN SINGH SIGRIWAL:**

SHRI JANARDAN MISHRA:

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

- (a) the details of the number of Indian languages currently supported by BharatGen's AI model and the road map for achieving full language coverage;**
- (b) the specific applications of BharatGen in the healthcare, agriculture, education, and governance sectors, State and District-wise including Dewas-Shajapur, Madhya Pradesh;**
- (c) the number of Technology Innovation Hubs and Translational Research Parks currently active under the BharatGen network and their thematic focus, State-wise;**
- (d) the details of the institutions involved in the BharatGen consortium and their role in its development and deployment;**
- (e) whether the Government proposes to partner with academic or research institutions in Dakshina Kannada to promote BharatGen's research, language development or pilot applications in the region and if so, the details thereof;**
- (f) whether the Government has launched or is in the process of launching BharatGen AI for public and institutional use in the Dewas-Shajapur, Lok Sabha Constituency of Madhya Pradesh; and**
- (g) if so, the details thereof including the objectives, features and intended user groups of BharatGen AI and the benefits to the rural and semi-urban areas of Dewas-Shajapur?**

ANSWER

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

विज्ञान और प्रौद्योगिकी तथा पृथ्वी विज्ञान मंत्रालय के राज्य मंत्री (स्वतंत्र प्रभार)
(डॉ. जितेंद्र सिंह)

(a) to (g): A statement is laid on the Table of the House.

STATEMENT AS REFERRED IN REPLY TO PARTS (a) to (g) OF LOK SABHA STARRED QUESTION NO. 257 FOR 06.08.2025 REGARDING “BHARATGEN AI MODELS”

(a) to (g): 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 models cover 9 Indian languages which include Hindi, Marathi, Tamil, Malayalam, Bengali, Punjabi, Gujarati, Telugu, and Kannada.

Roadmap of language coverage across these models includes the following milestones:

- By December 2025, a total of 15 Indian Languages (including Assamese, Bengali, Gujarati, Hindi, Kannada, Maithili, Malayalam, Marathi, Nepali, Odia, Punjabi, Sanskrit, Sindhi, Tamil and Telugu) will be covered.**
- By June 2026, all 22 scheduled Indian languages will be covered.**

BharatGen has developed applications in the sectors of agriculture, governance and defence, wherein pilots have been carried out. Once deployed fully, it is planned to make these applications available across all states and districts.

BharatGen is implemented as a project under the National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) of Department of Science and Technology (DST). As part of the BharatGen network, following Technology Innovation Hubs (TIHs) are currently active:

1. TIH Foundation for IoT and IoE, IIT Bombay (Maharashtra)

This TIH is hosting the BharatGen project and serves as the central program implementation and coordination hub. It is responsible for end-to-end execution, including managing the national academic consortium, overseeing the development of sovereign foundational AI models across text, speech, and vision, and driving ecosystem partnerships for compute, data, and talent. The TIH also drives governance and strategic planning for the BharatGen project, ensuring cohesive progress across all stakeholders.

2. IITM Pravartak Technologies Foundation, IIT Madras (Tamil Nadu)

This TIH serves as an implementation partner, focusing on solutioning and real-world deployment of BharatGen AI technologies. Thematic focus areas include governance, security, and media-related use cases.

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

While BharatGen is working with the above consortia members, it may explore partnerships with research institutions in Karnataka.

Since BharatGen AI is currently under pilot deployment phase, it has not been released for public and institutional use. Once fully deployed, there is a plan to extend its applicability across all states and districts.
