

GOVERNMENT OF INDIA
MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY
RAJYA SABHA
STARRED QUESTION NO. *66
TO BE ANSWERED ON: 24.07.2026

STATUS OF SAFE AND TRUSTED AI PILLAR UNDER INDIAAI MISSION

***66. SHRI SATNAM SINGH SANDHU:**

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

- (a) the objectives and current status of the Safe and Trusted AI pillar under the IndiaAI Mission;
- (b) the nature of projects selected relating to bias mitigation, explainability, privacy-preserving Artificial Intelligence (AI) and machine learning;
- (c) whether the proposed IndiaAI Safety Institute has been operationalised; and
- (d) the manner in which India's approach to responsible AI aligns with global norms while safeguarding national interests?

ANSWER

MINISTER OF ELECTRONICS AND INFORMATION TECHNOLOGY
(SHRI ASHWINI VAISHNAW)

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

STATEMENT REFERRED TO IN THE REPLY TO RAJYA SABHA STARRED QUESTION NO. *66 FOR 24.07.2026 REGARDING “STATUS OF SAFE AND TRUSTED AI PILLAR UNDER INDIAAI MISSION”

(a) to (d): **IndiaAI Mission:** India’s AI policy is based on Prime Minister Narendra Modi’s vision of democratizing technology. It is aimed at creating economic opportunities and employment for youth, while addressing risks associated with AI.

Accordingly, the Government approved the IndiaAI Mission in March 2024 with an outlay of ₹10,371 crore over five years to build a comprehensive AI ecosystem in the country.

The Mission comprises seven pillars viz. IndiaAI Compute, Foundation Models, AIKosh, IndiaAI Application Development Initiative, FutureSkills, Startup Financing, and Safe & Trusted AI.

Progress under IndiaAI mission till date:

AI models: Government of India is supporting Indian startups and institutions to develop large language models. 20 indigenous sovereign model proposals have been identified for support, comprising 12 LLMs and 8 SLMs. Notable successes so far include:

- Sarvam AI’s Sarvam 30 billion parameter and 105 billion parameter models
- Gnani.AI’s speech-to-speech model
- BharatGen’s multilingual foundation models
- Avataar AI’s video generation model

Domain-specific SLMs, include specialized models and AI frameworks addressing healthcare, language technologies and agentic AI systems are at various stages of development and deployment.

Affordable compute:

- Compute service providers empanelled: 15
- Number of projects supported: 237
- GPU hours sanctioned: 93 lakh GPU hours

Application Development:

- 12 National-level Hackathons/Innovation Challenges launched.
- Supported the development of 62 AI prototypes and 20 AI solutions deployment

Talent development:

- 686 fellowships have been awarded across undergraduate, postgraduate and doctoral programmes in 178 institutions.
- YUVA AI for All programme has been completed by over 26 lakh individuals.

IndiaAI labs:

- 27 India Data and AI Labs have been established, which has supported training for more than 2500 students
- Work ongoing on additional 188 labs.

AI Centres of Excellence:

58 Artificial Intelligence Centres of Excellence (AI-CoEs) being established across States and Union Territories in collaboration with the respective State/UT Governments and industry partners.

Safe & Trusted AI pillar

At the same time, the Government is also cognizant of the risks associated with AI. The Safe & Trusted AI pillar of the IndiaAI Mission aims to promote the responsible development, deployment and adoption of AI through indigenous governance frameworks, standards, tools and evaluation mechanisms.

13 Responsible AI projects have been approved in educational institutions across the country. These include topics such as bias mitigation, machine unlearning, privacy-preserving AI, explainability, deepfake detection and AI risk assessment.

Key initiatives include:

- Saakshya, multi-agent framework developed by IIT Jodhpur & IIT Madras for deep fake detection
- AI Vishleshak for improving Audio-Visual forgery detection system
- IIT Kharagpur's project on Real-Time Voice Deepfake Detection System

Nature & Objectives of projects related to bias mitigation, explainability, privacy preserving Artificial Intelligence (AI) and machine learning:

- Bias Mitigation: NIT Raipur is developing responsible AI algorithms to reduce bias in medical image analysis, clinical decision-making and healthcare applications
- Privacy-Preserving & Explainable AI: IIT Delhi (with IIIT Delhi and IIT Dharwad) is developing privacy-preserving machine learning models using federated learning.
 - Defence Institute of Advanced Technology (DIAT) is building explainable & privacy-preserving AI for security applications
- Machine Unlearning: IIT Jodhpur is developing techniques to remove unwanted, sensitive or outdated information/concepts from generative foundation models

Overall, these projects are aimed at making AI fairer, more transparent, privacy-preserving, secure and trustworthy.

- The AI Safety Institute has been established for AI safety, security and developing trust with diverse stakeholders, engaging academia, startups, industry and Government bodies through a hub-and-spoke model.
- The India AI Governance Guidelines lay down a risk-based governance framework addressing algorithmic bias, misinformation, deepfakes and unintended societal harm, and propose institutional mechanisms including the AI Governance and Economic Group (AIGEG), Technology and Policy Expert Committee (TPEC) and AI Safety Institute (AISI) for oversight, standards and safety research.

The IndiaAI Safety Institute was announced in January 2025 to advance science-based and indigenous research on AI safety and governance. The Institute is responsible for promoting the safe, secure, inclusive and trustworthy development, deployment and adoption of AI in India.

It also oversees the implementation of projects selected under the two rounds of Expressions of Interest (EoIs) under the Safe & Trusted AI pillar.

India's approach to responsible AI is anchored in the vision of "AI for All". It seeks to harness AI for inclusive development while ensuring that it remains safe, trusted, and aligned with India's socio-economic priorities.

- India actively engages with global processes on AI governance, including Global Partnership on Artificial Intelligence (GPAI), G20, and United Nations, to ensure that international norms reflect the perspectives and priorities of the Global South
- As Chair of the India AI Impact Summit 2026, held in New Delhi, India contributed in shaping the global agenda on responsible AI through key deliverables:
 - As part of New Delhi Frontier AI Impact Commitments, leading AI developers made voluntary commitments on the safe and responsible development of frontier AI
 - Guidance Note on AI Governance, offered practical guidance to countries on designing context-appropriate AI governance frameworks
 - Trusted AI Commons, a shared repository of open resources to democratise access to safe and trusted AI

These outcomes reflect India's commitment to shaping global norms on AI governance while protecting its strategic and developmental interests.
