

GOVERNMENT OF INDIA
MINISTRY OF ELECTRONICS AND INFORMATION TECHNOLOGY
LOK SABHA
UNSTARRED QUESTION NO. 2988
TO BE ANSWERED ON: 05.08.2026

INSTITUTIONAL FRAMEWORK FOR AUDITING ARTIFICIAL INTELLIGENCE

2988. DR. C N MANJUNATH:

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

- (a) whether the Government has proposed to establish an institutional framework for auditing Artificial Intelligence (AI) algorithms deployed in governance processes and citizen-facing digital services;
- (b) the details of the existing guidelines, protocols, standards or evaluation mechanisms, if any, for assessing the accuracy, bias, fairness, transparency, reliability and accountability of AI systems used by Government Ministries, Departments and agencies;
- (c) whether the Government is developing any certification, validation, testing or compliance mechanism for AI models procured, developed or deployed for public sector applications and if so, the details thereof; and
- (d) whether the Government proposes to collaborate with academic and research institutions, technology experts, standards development organisations and ethics committees to formulate guidelines, benchmarks and best practices for ensuring algorithmic transparency, accountability and the responsible use of AI in governance and if so, the details thereof?

ANSWER

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

(a) to (d): 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.

India's AI strategy builds on India's vibrant IT sector with annual revenue of 300 billion dollars and 60 Lakh strong workforce. India AI Mission has seven pillars including development of AI models, applications, datasets, talent development and affordable compute.

The Government has put in place institutional mechanisms for the governance and oversight of AI systems, including those deployed in governance processes and citizen-facing digital services.

IndiaAI Governance Guidelines:

The IndiaAI Governance Guidelines were released on 5th November 2025, providing a comprehensive national framework for the safe, responsible and inclusive development and deployment of Artificial Intelligence in the country, anchored around seven core principles (“Sutras”): Trust, People First, Innovation over Restraint, Fairness & Equity, Accountability, Understandable by Design, and Safety, Resilience & Sustainability.

The IndiaAI Governance Guidelines address accuracy, bias, fairness, transparency, reliability and accountability through their Fairness & Equity, Accountability and Understandable by Design Sutras, laying down a risk-based governance approach with safeguards against algorithmic bias.

Safe & Trusted AI Pillar:

Under this pillar, 13 projects addressing bias mitigation, machine unlearning, privacy-preserving AI, algorithm auditing, explainability, deepfake detection have been selected, in partnership with leading academic and research institutions. These include:

- Development of Responsible Artificial Intelligence for Bias Mitigation in Health Care Systems, implemented by NIT Raipur.
- Robust Privacy Preserving ML Models, implemented by IIT Delhi in collaboration with IIIT Delhi and IIT Dharwad.
- Enabling Explainable & Privacy Preserving AI for Security, implemented by the Defence Institute of Advanced Technology (DIAT) Pune.
- Machine Unlearning in Generative Foundation Models, implemented by IIT Jodhpur.

Collaboration:

The Government has initiated collaboration with a wide range of stakeholders including academic and research institutions, technology experts, industry, startups and government bodies to develop guidelines, benchmarks and best practices for responsible AI:

- Safe & Trusted AI projects are implemented in direct partnership with academic and research institutions, including NIT Raipur, IIT Delhi, IIIT Delhi, IIT Dharwad, IIT Jodhpur and the Defence Institute of Advanced Technology.
- India was the founding Chair of the Global Partnership on Artificial Intelligence (GPAI), which has produced an Advancement Report, “Towards Substantive Equality in Artificial Intelligence,” on gender and diversity in AI policy.
- India is among the 194 member States that adopted UNESCO's Recommendation on the Ethics of Artificial Intelligence (November 2021).
- At the India-AI Impact Summit 2026, the New Delhi Frontier AI Impact Commitments were signed by 13 leading global and Indian frontier model developers, and the Trusted AI Commons was announced with 23 partner countries to create an open repository of AI safety resources and tools.
- A Benchmarking Committee has been constituted to recommend suitable benchmarks and evaluation protocols for India's foundational AI models.
- Sectoral collaborations and hackathons have been organised with institutions such as I4C, Geological Survey of India, Ministry of Ayush, Ministry of MSME, NFRA and National Cancer Grid to promote applied AI solutions and best practices;

Benchmarking and Certification:

Mechanisms supporting certification, validation and testing of AI models for public sector applications include the following:

- The Benchmarking Committee's evaluation protocols apply to foundation models supported under the IndiaAI Mission.
- The bias mitigation, explainability and algorithm auditing projects under the Safe & Trusted AI pillar contribute tools and methodologies for testing and validating AI systems prior to and during deployment.

- The IndiaAI Governance Guidelines provide the overarching risk-based framework within which compliance expectations for AI systems, including those procured or deployed for public sector applications, are situated.
- National Accreditation Board for Certification Bodies (NABCB) provides accreditation to the Certification Bodies for Artificial Intelligence Management System (AIMS) as per ISO/IEC 17021-1:2015 and ISO/IEC 42006:2025 for certifying the AI Developers / AI Producers / AI Users as per ISO/IEC 42001:2023.
- NABCB has developed accreditation framework for Artificial Intelligence Management System (AIMS) certification as per ISO/IEC 42001:2023 (Information technology - Artificial intelligence - Management system).
