

Selected Projects for Responsible AI themed Projects under Safe & Trusted AI Pillar

This initiative is a part of the Safe & Trusted AI Pillar within the IndiaAI Mission, which seeks to advance the responsible development, deployment, and adoption of AI by developing indigenous tools and frameworks, self-assessment checklists for innovators, and other guidelines and governance frameworks.

The following projects have been selected:

User ID	Name of the Theme	Name of the Applicant	Title of the Project
ID_997	Machine Unlearning	Indian Institute of Technology Jodhpur	Unlearning in Generative Foundation Models
ID_535	Synthetic Data Generation	Indian Institute of Technology Roorkee	Design and Development of Method for Generating Synthetic Data for Mitigating Bias in Datasets and Framework for Mitigating Bias in Machine Learning Pipeline for Responsible AI
ID_1259	AI Bias Mitigation Strategy	National Institute of Technology Raipur	Development of Responsible Artificial Intelligence for Bias Mitigation in Health Care Systems
ID_2053	Explainable AI Framework	Defence Institute of Advanced Technology (DIAT),	Enabling Explainable and Privacy Preserving AI for Security

		Pune in partnership with Mindgraph Technology Pvt. Ltd.	
ID_645	Privacy Enhancing Strategy	Indian Institute of Technology Delhi in partnership with Indian Institute of Technology, Dharwad, Indraprastha Institute of Information Technology Delhi (IIIT Delhi), and Telecommunication Engineering Center (TEC)	Robust Privacy-Preserving Machine Learning Models
ID_1419	AI Ethical Certification Framework	Indraprastha Institute of Information Technology Delhi (IIIT Delhi) in partnership with Telecommunication Engineering Center (TEC)	Nishpaksh: Tools for assessing fairness of AI model
ID_595	AI Algorithm Auditing Tool	Civic Data Labs	ParakhAI - An open-source framework and toolkit for Participatory Algorithmic Auditing

ID_2167	AI Governance Testing Framework	Amrita Vishwa Vidyapeetham in partnership with Telecommunication Engineering Center (TEC)	Track-LLM, Transparency, Risk Assessment, Context & Knowledge for Large Language Models
---------	------------------------------------	--	---