AI-Driven Transformation of India's Statistical and Data Ecosystem

March 20, 2026

Key Takeaways

- **AI tools for NIC coding, unlocking legacy data, intelligent document search, AI enabled chatbot on website have been introduced**

- **e-Sankhyiki has integrated MCP and semantic search,** for ease of access of official statistical datasets

- **NDAP is expanding cross-sectoral analytics** through visualisation tools and AI-based search capability

- **Institutional initiatives such as the Data Innovation Lab are strengthening AI readiness across public data functions**

- Bharat-VISTAAR, meteorological forecasting, and Aadhaar authentication illustrate **wider sectoral AI deployment**

- Unique Identification Authority of India (UIDAI) has introduced **advanced AI-based biometric deduplication and document verification platform**

Introduction

Artificial Intelligence (AI) has emerged as one of the most transformative technologies of the 21st century, influencing the way economies, institutions, and public systems function across the world. As noted by the OECD, the **adoption of AI in government opens new possibilities**; for improving public service delivery, decision-making, and administrative efficiency.

India has similarly recognised AI as a critical enabler for improving productivity, accelerating innovation, and strengthening governance through better use of data and digital tools. This technological shift is increasingly visible across India's public data systems, where AI, machine learning, and advanced data analytics are being integrated into classification, access, and decision-support processes, supporting a more efficient and responsive digital ecosystem.

AI-Enabled Statistical Access and Public Data Platforms

India's official statistical platforms are increasingly moving toward AI-enabled data access systems that improve how users interact with public datasets. Recent initiatives by the NITI Aayog and Ministry of Statistics and Programme Implementation (MoSPI) reflect a broader shift to **AI-based interfaces that support direct querying, natural language search, and context-based information accessibility**.

MCP Integration with e-Sankhyiki Platform

In February 2026, the National Statistics Office (NSO) introduced the beta version of Model Context Protocol (MCP) server on the e-Sankhyiki portal, India's national platform for official statistics. This initiative is part of NSO's broader effort to improve access to official statistics for citizens, researchers, and businesses. **MCP is designed to enable direct interaction with statistical datasets through users' own AI-based tools and applications.**

> **e-Sankhyiki as India's Official Statistics Data Platform**
>
> e-Sankhyiki portal was **launched in 2024,** with the objective to **establish a comprehensive data management and sharing system** for ease of dissemination of official statistics in the country.
>
> **As of date, it has 21 statistical products with more than 136 million records** for better discovery and data management.

The server currently provides access to **21 statistical products**, available on e-Sankhyiki. It enables users to **query data directly without downloading large files**, connect official datasets with their own analytical systems, **automate statistical reporting**, and access multiple datasets through a **unified interface**. This is expected to reduce time spent on data retrieval and improve efficiency in analysis and decision-making.

Steps to connect AI agent to official statistics can be accessed at https://datainnovation.mospi.gov.in/mospi-mcp#connect.

Semantic Search for e-Sankhyiki Datasets

A beta version of semantic search feature is developed to allow users to explore datasets on e-Sankhyiki through natural language prompts. This will improve the usability of the e-Sankhyiki dashboard through a better user experience. As per the context provided in the natural language prompt, the system can direct users to the most appropriate product page on the portal, thereby simplifying access to required statistical information.

Semantic Search for e-Sankhyiki Datasets can be accessed at https://esankhyiki.mospi.gov.in/.

AI Chatbot for Statistical Information Access

An AI-powered chatbot has also been introduced to make the **MoSPI website** more interactive and user-friendly. The chatbot allows users to search **datasets, reports, and statistical publications** through **simple natural language queries**.

It is designed to provide **context-based responses** while maintaining **conversational continuity** for smoother user interaction. In addition to answering queries, the chatbot also
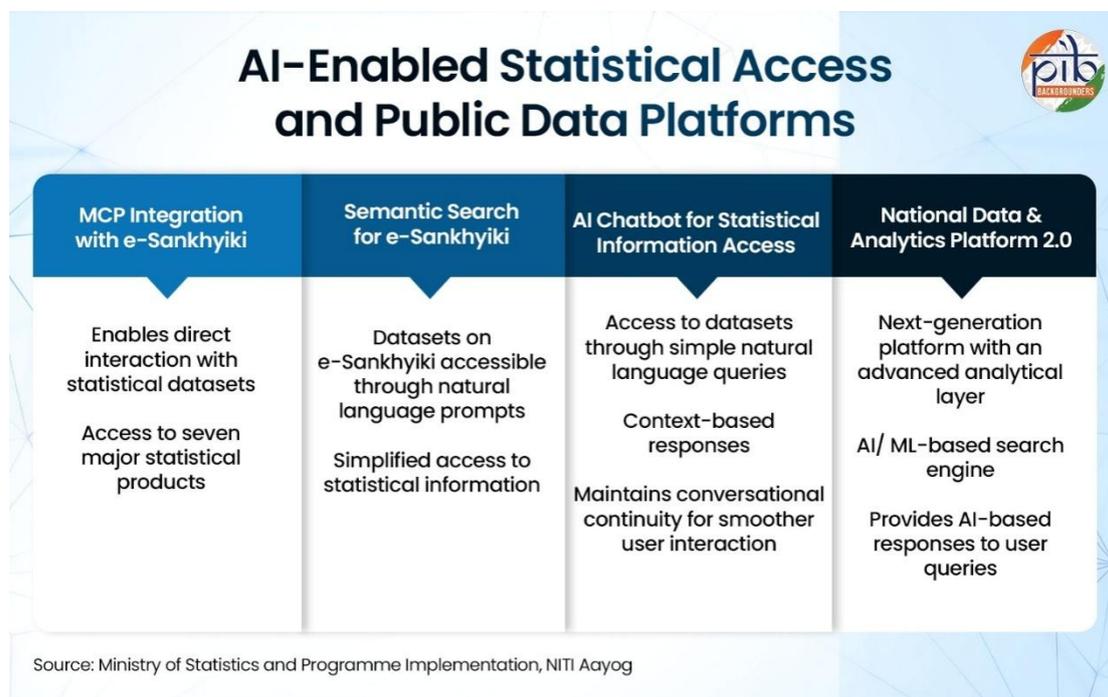
directs users to **relevant sections of the website through embedded links**, helping them access required statistical information more quickly and with minimal effort.

The AI Chatbot for statistical information can be accessed at https://www.mospi.gov.in.

Source: Ministry of Statistics and Programme Implementation, NITI Aayog

National Data & Analytics Platform (NDAP)

Launched in 2022, NDAP hosts datasets from multiple government agencies, presents them in a coherent format, and provides tools for **analytics and visualisation**. At present, the **data available on NDAP is categorised across 52 ministries and 31 sectors.**

Under **NDAP 2.0, a next-generation platform with an advanced analytical layer** is envisaged to **improve data discoverability, utility, cross-sectoral analysis and support data-driven decision-making processes**. The platform aims to enhance usability, strengthen user engagement and improve data discoverability to enable a more intuitive and efficient data exploration experience. This includes **pre-curated insights through visualisations and charts, domain-specific analytical modules, harmonisation of micro-level data, improved visualisation through UI/ UX enhancements, and an AI/ ML-based search engine capable of providing AI-based responses to user queries**.

These initiatives together, are strengthening the transition from data availability to intelligent data usability across India's public statistical systems.

AI in Statistical Classification and Survey Operations

AI is being integrated into statistical classification and survey operations **to improve accuracy, reduce manual effort, and support faster field-level decision-making**. Through classification tools and search solutions, official survey processes are being made **more efficient, consistent, and responsive** to the growing complexity of statistical data collection.

AI tool for Industrial Classification

An AI/ ML-based classification tool has been introduced to ease the use of the National Industrial Classification (NIC) in the production of official statistics. The tool applies **natural language processing to allow stakeholders to enter text queries and generate the three-most relevant NIC code suggestions**. This initiative helps reduce manual effort in classification, improves enumerator productivity, and supports greater accuracy in statistical data collection, thereby strengthening the quality of evidence available for planning and policymaking.

AI-powered NIC Code Semantic Search Tool can be accessed at https://nicfinder.mospi.gov.in/

MoSPI StatsDoc AI Assistant: AI-Enabled Intelligent Search Solution for Documents

An **AI-powered Intelligent Search Solution for Documents, which allows users to search uploaded documents in natural language**, keeping in mind the different stakeholders including field investigators, who need to refer to manuals, reports, publications, etc. mostly available in PDFs or images. **This tool has a knowledge base of the all the latest documents that are published by the Ministry (from April 2025), including the Instruction Manuals for different surveys.**

The chatbot can be accessed under the AI Pilots of Offerings Section of the MoSPI website as well as through https://statsdoc.ai.mospi.gov.in/.

AI-Based Legacy Data Extraction and Processing Tool

Legacy data of NSO India are stored in formats such as PDFs, CSVs, Excel files (with merged cells, Hindi text, etc.), and images. Extracting this data requires detailed coding. Without necessary coding knowledge or resources, the data's usability is significantly hindered, affecting efficient analysis and decision making, necessitating the need to rejuvenate legacy data for ease of access and deriving useful insights in an efficient manner. **The solution is an AI-based tool that can extract legacy data from various documents and store it in a database for further processing and analysis.**
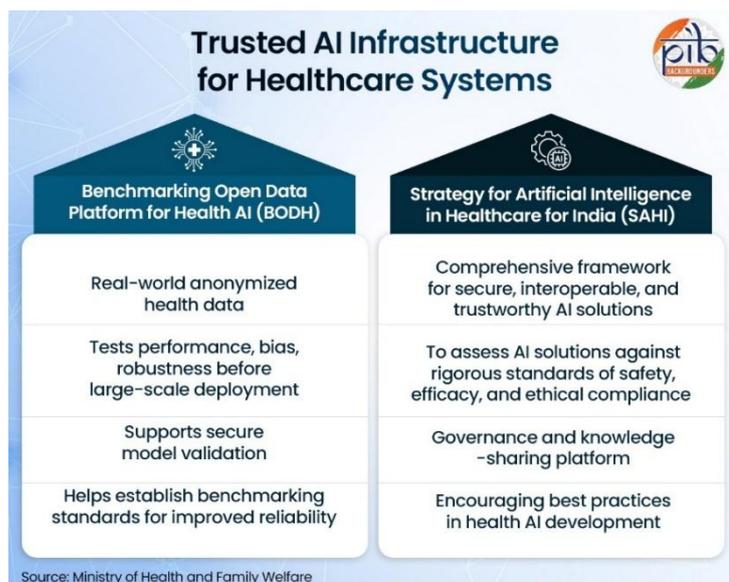
The tool can be accessed under the AI Pilots of Offerings Section of the MoSPI website as well as through https://legacydata.ai.mospi.gov.in/.

Illustrative Sectoral Use Cases of AI Across Public Data Domains

India is increasingly applying AI across critical public service sectors where reliability, safety, and real-time decision support are essential. The following initiatives are certain **examples of how public datasets, digital infrastructure, and policy frameworks are being used to support this broader transition**, are strengthening support in areas directly linked to public welfare.

Trusted AI for Healthcare Systems

a) **Benchmarking Open Data Platform for Health AI (BODH):** BODH was launched in February 2026 to enable **systematic evaluation of AI models using diverse and anonymized real-world health datasets.** The platform **assesses AI systems for performance, robustness, bias, and generalizability before large-scale deployment**, helping establish benchmarking standards that

Trusted AI Infrastructure for Healthcare Systems

Benchmarking Open Data Platform for Health AI (BODH)	Strategy for Artificial Intelligence in Healthcare for India (SAHI)
Real-world anonymized health data	Comprehensive framework for secure, interoperable, and trustworthy AI solutions
Tests performance, bias, robustness before large-scale deployment	To assess AI solutions against rigorous standards of safety, efficacy, and ethical compliance
Supports secure model validation	Governance and knowledge-sharing platform
Helps establish benchmarking standards for improved reliability	Encouraging best practices in health AI development

Source: Ministry of Health and Family Welfare

improve reliability and clinical relevance in line with national public health priorities. It **strategically resolves the "AI Quality Testing Trilemma"** — the traditional trade-off between reliability, openness, and coverage.

Built on the Ayushman Bharat Digital Mission (ABDM) framework, BODH uses nationwide digital health data to support secure model testing and validation while maintaining privacy safeguards. It provides an environment where developers can train and evaluate AI systems on diverse datasets, and regulators can undertake structured third-party assessments with stronger statistical confidence. The platform is intended to build a **trusted ecosystem for benchmarking health AI models** and improving their consistency across healthcare settings.

> **ABDM: Building Digital Health Infrastructure in India**
>
> The Ayushman Bharat Digital Mission (ABDM) aims to **build the foundational digital infrastructure necessary to support the healthcare system in India**. It seeks to connect different stakeholders across the healthcare ecosystem through digital highways.

b) **Strategy for Artificial Intelligence in Healthcare for India (SAHI) :** Launched alongside BODH, **SAHI has been conceptualized as a comprehensive framework to support the development of secure, coordinated, and trustworthy AI solutions in healthcare.** It aims to facilitate collaboration among healthcare institutions, technology developers, researchers, and policymakers so that AI solutions are assessed against rigorous standards of safety, efficacy, and ethical compliance before large-scale adoption.

It will also **function as a governance and knowledge-sharing platform,** encouraging best practices in health AI development and implementation while emphasizing patient data protection, responsible use of algorithms, and accountability across the healthcare system.
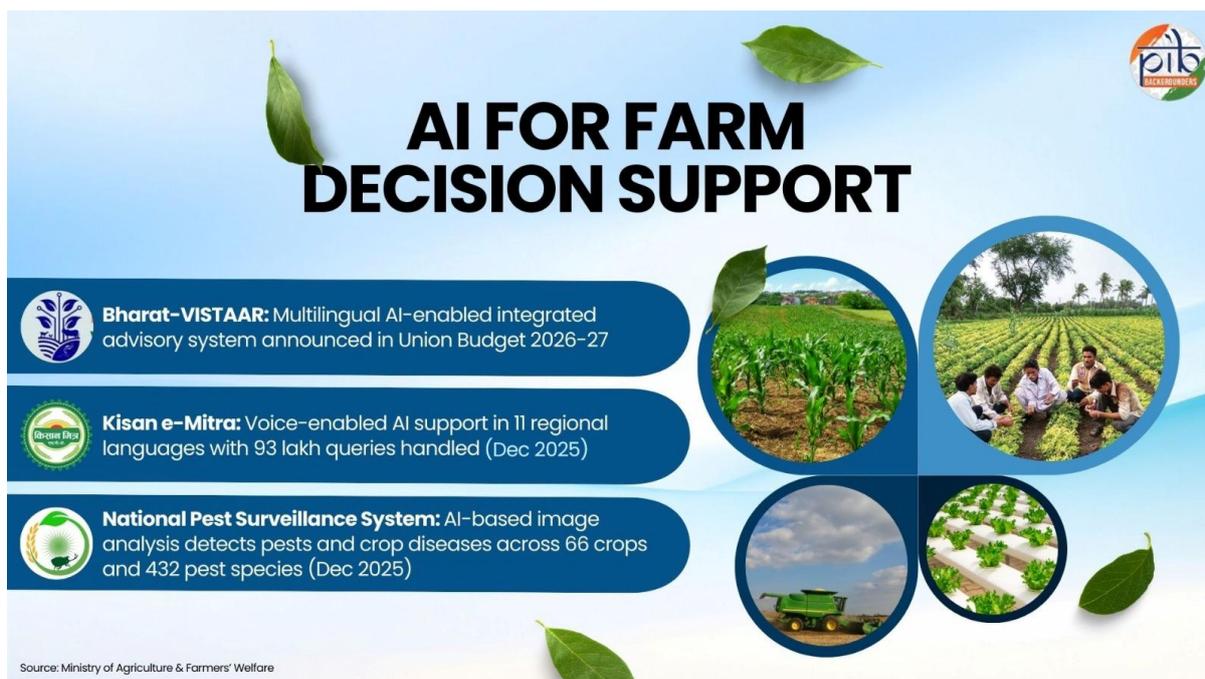
Meteorological Data and AI-Based Forecasting

India Meteorological Department and other institutions under the Ministry of Earth Sciences are increasingly applying AI-based tools for experimental weather and climate forecasting. These include the Advanced Dvorak Technique for cyclone intensity estimation, along with AI and machine learning–based foundation models and hybrid systems that combine artificial intelligence with dynamical forecasting methods for weather prediction.

AI-related research is also being applied across several forecasting areas, including short-range global forecasts, precipitation downscaling, fire location prediction, fog forecasting, lightning and thunderstorm alerts, and improved precipitation estimation through deep learning within numerical weather prediction systems.

AI for Farm Decision Support

a) **Bharat-VISTAAR (Virtually Integrated System to Access Agricultural Resources)**: The **Union Budget 2026-27 has proposed Bharat-VISTAAR**, **a multilingual AI-enabled system that shall integrate AgriStack portals and Indian Council of Agricultural Research advisory resources**. This will enhance farm productivity, improve decision making and reduce risk through customized advisory support for the farmers.

b) **Kisan e-Mitra**: **Launched in 2023, the platform provides voice-enabled AI support to farmers** by answering queries on major government schemes in 11 regional languages and had responded to over 93 lakh queries as of December 2025.

c) **National Pest Surveillance System (NPSS): Further strengthening digital agricultural support, the NPSS uses AI and machine learning to detect pest attacks and crop diseases through image-based analysis**, supporting over 10,000 extension workers across 66 crops and 432 pest species as of December 2025.

AI FOR FARM DECISION SUPPORT

Bharat-VISTAAR: Multilingual AI-enabled integrated advisory system announced in Union Budget 2026-27

Kisan e-Mitra: Voice-enabled AI support in 11 regional languages with 93 lakh queries handled (Dec 2025)

National Pest Surveillance System: AI-based image analysis detects pests and crop diseases across 66 crops and 432 pest species (Dec 2025)

Source: Ministry of Agriculture & Farmers' Welfare

Together, they reflect **a growing institutional emphasis on using data-driven tools** to improve reliability, strengthen service delivery, and support more informed public decision-making.

Institutional Architecture Supporting AI Integration

India's data ecosystem is increasingly being strengthened through institutional platforms that support innovation, technology adoption, and capacity building across public systems. Recent initiatives indicate how dedicated data infrastructure, collaborative research frameworks, and digital public platforms are being used to integrate artificial intelligence into official statistics.

Alongside technological deployment, **emphasis is also being placed on partnerships and trainings**, to ensure that emerging tools improve efficiency, expand accessibility, and support reliable decision-making across sectors.
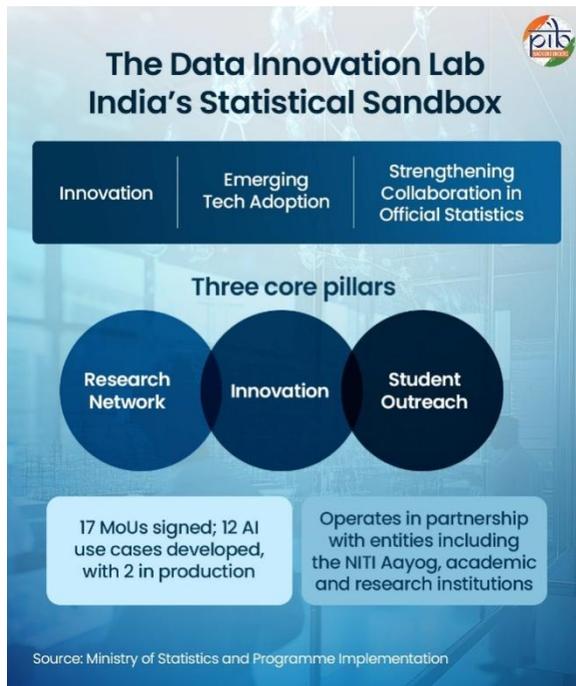
The Data Innovation Lab: India's Statistical Sandbox

The establishment of the Data Innovation Lab (DIL) under Data Informatics and Innovation Division (DIID) represents a cornerstone in modernising India's National Statistical System. It functions as a **strategic platform for promoting innovation, supporting the adoption of emerging technologies, and strengthening collaboration in the field of official statistics.**

In line with the growing importance of data-driven systems, the **DI Lab focuses on improving the efficiency, accuracy and accessibility of statistical processes through the use of advanced technologies such as AI, Big Data Analytics and Cloud Computing.** The objective is to build a stronger and more transparent statistical ecosystem that supports evidence-based policymaking and data-centric governance.

The initiative is structured around three core pillars, **Research Network, Innovation and Student Outreach**, which together support the development of the country's data ecosystem

through the **DI Lab Portal**, designed as a collaborative digital platform for research, experimentation and engagement.

The Data Innovation Lab
India's Statistical Sandbox

| Innovation | Emerging Tech Adoption | Strengthening Collaboration in Official Statistics |

Three core pillars

Research Network — Innovation — Student Outreach

17 MoUs signed; 12 AI use cases developed, with 2 in production

Operates in partnership with entities including the NITI Aayog, academic and research institutions

Source: Ministry of Statistics and Programme Implementation

As of January 2026, MoSPI has signed 17MoUs with various institutions. It has also developed a repository of 12 new AI use cases under the Data Innovation Lab with two use cases in production.

Collaborative Ecosystem and Partnerships: The success of AI integration in official statistics depends on effective collaboration across government agencies, academic institutions, and international organizations. **The DI Lab operates in partnership with entities including the NITI Aayog, academic and research institutions**. This collaborative approach ensures that India's statistical modernisation continues to benefit with wider knowledge-sharing across key institutions.

Training and Capacity Building for AI-Driven Statistical Systems

Aligned with the national philosophy of **Sabka Saath, Sabka Vikas, Sabka Vishwas and Sabka Prayas**, capacity building is being positioned to move beyond traditional methods through the adoption of technology-driven systems, including **Big Data, Artificial Intelligence and Machine Learning**, to improve the accuracy, timeliness and global comparability of official statistics. Engagement with State statistical systems is being encouraged alongside expanded partnerships with universities and specialised institutes, reaffirming the commitment under **Mission Karmayogi** to strengthen training programmes and build a future-ready statistical workforce.

With the growing shift toward **Agentic AI**, where AI systems increasingly function with greater autonomy, the scope of training at the **National Statistical System Training Academy (NSSTA)** is also expected to expand further. A key emphasis continues to remain on the role of human intervention in providing context to raw data and addressing the risks of embedded bias in AI systems, recognising that continuous capacity building is essential to ensure that statistical officers remain equipped to produce high-quality and unbiased data in support of governance.

BharatGen Building India's Multilingual AI Foundation

Launched in June 2025, BharatGen is India's first government-funded, sovereign, multilingual, and multimodal Large Language Model (LLM). It has been developed under the National Mission on Interdisciplinary Cyber-Physical Systems and advanced through the IndiaAI Mission. **The model supports 22 Indian languages and integrates text, speech, and document-vision capabilities.** Built on India-centric datasets and led by a consortium of

academic institutions, BharatGen establishes a domestically developed AI stack for public and developmental applications.

BharatGen products can be accessed at https://bharatgen.com/

Advancing Aadhaar through Emerging Technologies

a) Invisible Shield Securing the Identities of over a Billion Indians: The Unique Identification Authority of India (UIDAI) has introduced **advanced AI-based biometric deduplication and document verification platform in February 2026**, as part of India's evolving digital security architecture. The new platform is designed to **improve enrolment and update accuracy** by **strengthening biometric matching across fingerprint, face, and iris modalities.** With the latest AI inference technologies, UIDAI has already **completed enhanced deduplication rollouts for several states and is on track to complete this exercise nationwide** in the coming months.

b) Aadhaar Vision 2032 : Recognising the rapidly evolving technological and regulatory environment, the UIDAI has initiated a **comprehensive strategic and technological review to guide the future development of Aadhaar under the Aadhaar Vision 2032 framework.** It aims to **integrate advanced technologies such as artificial intelligence, blockchain, quantum computing, advanced encryption, and next-generation data security systems** to strengthen Aadhaar's resilience, improve scalability, and ensure secure adaptation to future digital requirements.

AI-Based Aadhaar Authentication Solution

The AI and machine learning-based Aadhaar Face Authentication solution developed by UIDAI has seen rapid adoption. It is currently being used across numerous government services to enable smoother delivery of benefits to targeted beneficiaries.

Several flagship schemes, including PM Awas (Urban), PM E-Drive, PM-JAY, PM Ujjwala, PM Kisan, PM Internship, have integrated this system for improved service delivery.

Together, these initiatives reflect a **broader effort to build long-term institutional capacity for AI-enabled data systems in India.** They also highlight the importance of combining innovation, partnerships, and human capability to strengthen trusted digital public infrastructure.

Conclusion

India's experience demonstrates how AI can be integrated at scale to strengthen public digital capabilities. Beyond technology, the evolving use of AI reflects India's larger effort to make data more meaningful, accessible, and responsive to the everyday needs of its people. The aim is to strengthen both the quality of data and its practical use in responsive governance and sector-specific decision-making.

Many of the solutions being developed are shaped by India's unique context, including linguistic diversity and varying levels of socio-economic development. There is a clear emphasis on ensuring that **technological progress translates into inclusive and widely accessible solutions**. These initiatives indicate that **AI is increasingly emerging as an enabling layer across India's public data landscape**, improving the way information is classified, accessed, interpreted, and applied.

References

Ministry of Statistics and Programme Implementation

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2224472&reg=3&lang=2

https://datainnovation.mospi.gov.in/ai-pilots

https://nicfinder.mospi.gov.in/

https://www.pib.gov.in/Pressreleaseshare.aspx?PRID=2119641&reg=3&lang=2

https://datainnovation.mospi.gov.in/mospi-ai-use-cases#

https://www.mospi.gov.in/sites/default/files/announcements/Revised_DI_Lab_Guidelines_0_0.pdf

https://datainnovation.mospi.gov.in/aboutus

https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2208162&reg=3&lang=2

https://datainnovation.mospi.gov.in/governing-council

https://datainnovation.mospi.gov.in/selection-committee

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2227794&reg=3&lang=1

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2076859&reg=3&lang=2

https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2208162&reg=3&lang=2

Ministry of Finance

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2221432&reg=3&lang=2

Ministry of Health and Family Welfare

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2229226&reg=3&lang=1

https://abdm.gov.in/abdm

Ministry of Earth Sciences

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2158416&reg=3&lang=2

Ministry of Agriculture & Farmers' Welfare

https://static.pib.gov.in/WriteReadData/specificdocs/documents/2026/feb/doc2026214789901.pdf

Ministry of Electronics and Information Technology

https://www.pib.gov.in/PressReleasePage.aspx?PRID=2184639&reg=3&lang=2

https://uidai.gov.in/images/FaceAuthTransation.pdf

https://uidai.gov.in/images/The_Invisible_Shield_Press_release.pdf

https://www.pib.gov.in/PressNoteDetails.aspx?ModuleId=3&NoteId=157528&reg=3&lang=2

NITI Aayog

https://niti.gov.in/sites/default/files/2025-02/Annual%20Report%202024-25%20English_FINAL_LOW%20RES_0.pdf

Other Domestic Institutions

https://www.iitk.ac.in/bodh-at-india-ai-impact-summit-2026

International Institutions

https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report/how-artificial-intelligence-is-accelerating-the-digital-government-journey_d9552dc7.html

PIB Research