



Algorithmic Bias and Regulatory Oversight: Examining the Role of AI Governance in India's Public Sector

Vivek Yadav

Research Scholar

Kusuma Devi Degree College, Amanpur, Kasganj, Uttar Pradesh, IN

yadavvivek241@gmail.com

Abstract

The rapid expansion of artificial intelligence (AI) across India's public sector from welfare delivery and policing to taxation and judicial administration has intensified concerns about opacity, discrimination, and due-process deficits. This paper evaluates how India's evolving AI governance framework confronts algorithmic bias and identifies persistent regulatory gaps. Situating India within global debates on algorithmic fairness, it draws on socio-legal scholarship showing that data-driven systems can reproduce structural inequalities even without explicit discriminatory intent. India's transition from early policy statements to recent instruments such as the Digital Personal Data Protection Act 2023, the IndiaAI Mission (2024), and the India AI Governance Guidelines (2025), along with sectoral initiatives like the RBI's FREE-AI framework, reflects growing acknowledgment of fairness and accountability. However, these norms remain fragmented and largely non-binding, with crucial omissions such as the absence of safeguards against fully automated decision-making and limited mandates for explainability or impact assessments. Through case studies on digital welfare systems, biometric and predictive policing, and AI-enabled judicial administration, the paper demonstrates how such systems risk deepening exclusion and indirect discrimination. It argues that constitutional guarantees of equality and privacy provide a normative basis for redress, and proposes a rights-based AI law, mandatory impact assessments, stronger procurement transparency, and empowered oversight institutions.

Keywords: algorithmic bias, AI governance, India, public sector, DPDP Act, IndiaAI Mission, facial recognition, Aadhaar, welfare digitalisation, constitutional equality.

1. Introduction

Across the globe, governments are embedding algorithmic systems in core public functions welfare eligibility, criminal justice risk assessments, biometric identification, and automated content moderation. Foundational scholarship has demonstrated that such systems frequently reproduce or amplify existing social hierarchies, rather than neutralising them (Barocas & Selbst, 2016; Noble, 2018; O’Neil, 2016). India is no exception: digital public infrastructure such as Aadhaar, Unified Payments Interface (UPI) and DigiLocker has become a global reference point for “digital statecraft,” even as civil society actors document exclusion and discrimination in these systems’ everyday use.

AI now sits at the centre of this ecosystem. The *National Strategy for Artificial Intelligence* framed AI as a “game changer” for inclusive growth and public service delivery, with special emphasis on health, agriculture, education and smart cities. Subsequent policy documents and mission statements including NITI Aayog’s *Principles for Responsible AI*, the IndiaAI Mission (2024) and the India AI Governance Guidelines (2025) explicitly identify algorithmic bias, opacity and safety risks as major governance challenges.

At the same time, high-profile controversies such as Aadhaar-linked starvation deaths in Jharkhand, mass welfare exclusions due to e-KYC failures, and extensive reliance on facial recognition for policing the 2020–2022 protests and North East Delhi riots have sharpened questions about regulatory oversight of state algorithms. Investigations by journalists, academics and human rights organisations reveal significant barriers to transparency and external auditing, especially where public agencies procure proprietary systems from private vendors.

This paper asks: **To what extent does India’s evolving AI governance framework provide effective safeguards against algorithmic bias in the public sector, and what institutional and legal reforms are needed to close the gap between principle and practice?**

Methodologically, the paper combines:

- **Doctrinal analysis** of constitutional provisions (Articles 14–16 and 21), the DPDP Act 2023 and the DPDP Rules 2025, AI-related advisories under the Information Technology Act, and the India AI Governance Guidelines;
- **Policy analysis** of strategic documents such as the *National Strategy for AI*, *Principles for Responsible AI*, the IndiaAI Mission, and the RBI’s FREE-AI framework;
- **Case-based examination** of public-sector AI deployments in welfare digitalisation, biometric policing, and courts, drawing on Amnesty International, investigative journalism, and recent Indian legal scholarship.

The argument proceeds in six parts. Section 2 outlines conceptual frameworks for algorithmic bias and AI governance. Section 3 maps the Indian AI governance architecture. Section 4 presents empirical case studies of bias and exclusion in public-sector AI. Section 5 analyses oversight mechanisms and regulatory gaps. Section 6 advances a reform agenda grounded in constitutional principles and recent policy trajectories. Section 7 concludes.

2. Conceptualising Algorithmic Bias and AI Governance

2.1 Algorithmic Bias and Discrimination

Algorithmic bias is often defined as systematic and unfair discrimination produced by algorithmic

systems, leading to disparate outcomes for individuals or groups along protected or socially salient lines such as caste, gender, religion, class or disability (Barocas & Selbst, 2016; Bharati, 2025). Legal and socio-technical scholarship distinguishes between multiple sources of bias:

- **Data bias**, where training data encode historical discrimination (e.g., over-policing of certain communities, gendered labour patterns);
- **Design bias**, where developers' modelling choices embed assumptions about relevant features, costs and benefits;
- **Measurement and representation bias**, where proxies or labels systematically mischaracterise certain groups; and
- **Feedback loops**, in which biased outputs reinforce the very patterns they measure (e.g., predictive policing concentrating patrols in already over-policed areas).

Work in critical algorithm studies further emphasises that algorithms operate within broader structures of power. Systems designed and deployed in formally “neutral” ways can nonetheless reproduce caste, gender or racial hierarchies when they align with existing distributions of wealth, education, and institutional trust (Noble, 2018; O’Neil, 2016). In India, this interacts with deeply entrenched inequalities, including caste-based stratification, regional disparities, and religious polarisation.

The emerging Indian literature argues that algorithmic bias should be understood as a form of indirect discrimination under Article 14 (equality) and Articles 15–16 (non-discrimination and equal opportunity), regardless of the presence of explicit animus. When state agencies deploy AI tools whose outputs disproportionately burden Scheduled Castes

(SC), Scheduled Tribes (ST), Other Backward Classes (OBC), religious minorities, women or persons with disabilities, constitutional scrutiny should focus on effects, not just intent.

2.2 AI Governance and Regulatory Oversight

“AI governance” refers broadly to the laws, policies, institutions and technical standards that shape the design, deployment and oversight of AI systems. Global debates increasingly revolve around two (often complementary) approaches:

1. **Risk-based governance**, exemplified by the EU AI Act, which categorises AI systems by risk and imposes graduated obligations (e.g., conformity assessments, incident reporting, post-market monitoring) on high-risk uses;
2. **Rights-based governance**, which foregrounds fundamental rights such as privacy, equality, due process, and freedom of expression, and seeks to guarantee individual protections like the right to explanation, contestation, and human review of automated decisions.

For the public sector, key elements of effective governance include:

- **Transparency and explainability** (disclosure of AI use, documentation of logic, audit trails);
- **Algorithmic impact assessments** (ex ante evaluation of risks to rights, including discrimination);
- **Fairness and accuracy obligations**, supported by independent audits and representative datasets;
- **Institutional oversight**, with regulators empowered to investigate, sanction and mandate remediation; and

- **Participatory governance**, enabling affected communities, civil society and experts to shape norms and monitor harms.

India's AI governance project unfolds in this global context but is also shaped by its own constitutional framework, digital public infrastructure model, and developmental priorities.

3. India's Emerging AI Governance Architecture

3.1 Foundational Strategies and Ethical Principles

India's AI policy trajectory began with NITI Aayog's *National Strategy for Artificial Intelligence* (NSAI, 2018), which identified five priority sectors and highlighted the role of AI in improving the scale and quality of public service delivery. The NSAI also acknowledged the need to manage risks around privacy, security and exclusion but did so at a relatively high level.

Building on this, NITI Aayog's *Principles for Responsible AI* (2021) articulated seven principles for responsible AI management: safety and reliability; equality; inclusivity and non-discrimination; privacy and security; transparency; accountability; and protection and reinforcement of positive human values. Crucially, these principles are explicitly derived from the Indian Constitution and fundamental rights jurisprudence, positioning constitutional morality including substantive equality and protection of vulnerable groups as the normative anchor for AI governance.

However, these documents primarily provide soft-law guidance. They do not, on their own, create enforceable obligations or institutions with investigative powers.

3.2 IndiaAI Mission and the Safe & Trusted AI Pillar

In March 2024, the Union Cabinet approved the IndiaAI Mission, a flagship, multi-year programme

with an outlay of approximately ₹10,372 crore to build AI compute, data, skills, innovation and governance capacity. One of its seven pillars **Safe & Trusted AI** is explicitly dedicated to ethics, safety and responsible use, including the establishment of the IndiaAI Safety Institute (AISI).

The AISI is envisaged as a state-backed hub-and-spoke institution that will conduct AI safety research, develop testing and evaluation standards, and support regulators and ministries in assessing high-risk systems. Early commentary suggests that AISI could become a key node for algorithmic audits and certification, particularly for frontier and high-risk models used in critical infrastructure, security, and welfare administration. Nevertheless, as of late 2025, its precise statutory basis and powers vis-à-vis line ministries, the Data Protection Board, and sector regulators remain under development.

3.3 India AI Governance Guidelines 2025

In November 2025, the Ministry of Electronics and Information Technology (MeitY) released the *India AI Governance Guidelines*, a comprehensive framework intended to guide AI governance across the public and private sectors. The Guidelines are structured around:

- **Seven guiding "sutras"** Trust, People First, Innovation over Restraint, Fairness & Equity, Accountability, Understandable by Design, and Safety, Resilience & Sustainability adapted from the RBI's FREE-AI committee principles and applicable across sectors.
- **Six pillars** Infrastructure, Capacity Building, Policy & Regulation, Risk Mitigation, Accountability, and Institutions covering enablement, regulation and oversight.

- A **time-phased action plan** outlining short-, medium- and long-term steps, including establishing governance institutions (AI Governance Group and Technology & Policy Expert Committee), developing India-specific risk frameworks, piloting regulatory sandboxes, and suggesting legal amendments.

The Guidelines explicitly recognise algorithmic bias, deepfakes and discriminatory outcomes as core risks and call for measures to prevent harms to vulnerable groups. They also emphasise a “techno-legal” model that combines normative principles with tools like AI incident reporting, bias evaluation datasets, and testing infrastructure. The Centre for Responsible AI at IIT Madras, for instance, has recently released the IndiCASA dataset to evaluate bias in language models in Indian contexts, including stereotypes around caste, religion, gender and disability; this work is referenced as part of emerging evaluation ecosystems.

Yet, the Guidelines are, at present, non-binding. While they propose institutional reforms and suggest amendments to existing laws, they do not themselves create statutory duties or sanctions for non-compliance. Their effectiveness will depend on their integration into sectoral regulations and future AI-specific legislation.

3.4 Data Protection and Ancillary Legal Frameworks

The **Digital Personal Data Protection Act, 2023** is India’s primary horizontal data protection statute. It applies to digital personal data processing both within India and extraterritorially where goods or services are offered to individuals in India. The Act defines data fiduciaries and data principals, sets out lawful bases for processing, and establishes the Data Protection Board of India as the enforcement authority, with significant penalty powers.

From the perspective of algorithmic bias and automated decision-making, several features are notable:

- The Act requires that personal data used to make decisions about a data principal be “complete, accurate and consistent,” which implicitly covers data used for algorithmic profiling.
- However, unlike the GDPR, the DPDP Act does **not** enshrine an explicit right *not* to be subject to decisions based solely on automated processing, nor does it provide a right to an explanation of automated decisions.
- As several commentators observe, the Act does not mandate algorithmic transparency, impact assessments specific to AI, or bias audits, though draft DPDP Rules 2025 propose algorithm audits for Significant Data Fiduciaries using algorithmic software.

Other relevant instruments include the Information Technology Act, 2000 and its Intermediary Guidelines. MeitY’s AI-related advisories under these rules (especially the March 2024 advisory) require intermediaries and platforms to ensure that AI models do not permit unlawful content, bias or discrimination, or threaten electoral integrity; under-tested or unreliable AI tools must be labelled, and (in the initial version) certain models required prior government permission before deployment. A revised advisory later softened the permission requirement but retained obligations regarding transparency, consent and deepfake labelling.

Judicial and criminal-law frameworks also intersect with AI governance. The nine-judge bench in *Justice K.S. Puttaswamy v. Union of India* recognised privacy as a fundamental right under Article 21 and

set out a four-part test legality, legitimate aim, proportionality, and procedural safeguards for state intrusions, including digital surveillance. This proportionality test is directly relevant to AI systems used in welfare, policing and surveillance. The Aadhaar judgment likewise scrutinised the biometric ID programme's privacy and exclusion risks, referencing scholarship that documented exclusion errors and rights violations.

3.5 Sectoral and Judicial Oversight Initiatives

Sectoral regulators have started to issue AI-related guidance. The Reserve Bank of India's FREE-AI framework (2025) recommends principles and governance mechanisms for AI use in the financial sector, emphasising risk management, audit frameworks and the need to integrate AI with India's digital public infrastructure while maintaining fairness and accountability.

In the judicial domain, the Kerala High Court's 2025 policy on AI tools in the district judiciary prohibits the use of AI for decision-making or legal reasoning and urges "extreme caution" even in limited support functions, citing risks to privacy, data security and public trust. This reflects an early form of AI governance in the justice system, albeit one that errs on restraint rather than structured integration.

Indian courts have not yet explicitly recognised algorithmic bias as a freestanding violation of Article 14 or 15, but recent scholarship notes growing judicial engagement with digital rights and the potential use of Articles 32 and 226 to demand algorithmic transparency, fairness audits, and human oversight in public-sector AI.

4. Algorithmic Bias in India's Public Sector: Case Studies

4.1 Digital Identity and Welfare: Aadhaar, PDS and Samagra Vedika

Aadhaar, India's biometric ID system, underpins a range of welfare schemes, including the public distribution system (PDS), pensions, and employment guarantees. While proponents highlight reductions in "leakages," a substantial body of empirical and investigative work documents exclusion errors arising from biometric authentication failures, faulty seeding, and rigid algorithmic matching.

Activists and journalists have linked multiple starvation deaths in Jharkhand and other states to Aadhaar-related denial of rations, often due to biometric mismatches, cancelled ration cards following data mismatches, or failed e-KYC procedures. While governments sometimes dispute direct causality, these cases illustrate how Type I exclusion errors denying legitimate beneficiaries can have severe consequences in a context of poverty and food insecurity.

From an algorithmic-bias perspective, these failures are not randomly distributed. They disproportionately affect those with worn fingerprints (e.g., manual labourers, elderly persons), remote rural communities with poor connectivity, and marginalised groups with limited documentation. This constitutes a form of indirect discrimination, where ostensibly neutral authentication rules produce disparate impacts on vulnerable populations.

Telangana's **Samagra Vedika** platform illustrates a more sophisticated form of welfare algorithmics: it integrates multiple datasets (property, tax, utilities, welfare) and uses an entity resolution algorithm, EntityMatch, to detect "ghost beneficiaries" and identify fraud. Amnesty International's attempted external audit of EntityMatch exposed major accountability gaps: researchers were denied access to code and detailed documentation on grounds of

commercial confidentiality, forcing them to rely on limited API access and public presentations. The audit highlighted the risk that threshold choices in matching could lead to systematic exclusions and that such outcomes are particularly troubling when beneficiaries are poor and have limited ability to contest automated decisions.

These cases reveal structural weaknesses in regulatory oversight: there is no binding requirement for public agencies to conduct human rights impact assessments or algorithmic fairness audits before deploying such systems; procurement contracts often treat algorithms as proprietary; and affected individuals lack a clear statutory right to explanation or human review under the DPDP Act.

4.2 Biometric and Predictive Policing: Facial Recognition

Indian police forces increasingly deploy facial recognition technology (FRT) for law enforcement, crowd control, and protest monitoring. Delhi Police, for example, has acknowledged using facial recognition in over 750 cases related to the North East Delhi riots, with results presented as evidence in court. Media reports and independent analyses suggest that in many instances, arrests were based primarily on FRT matches sometimes from side or rear-view images without robust corroborating evidence; a large proportion of these cases have resulted in acquittals or discharges, raising serious questions about reliability and bias.

Amnesty International's *Ban the Scan* campaign has documented extensive FRT deployments in Hyderabad's public spaces, including random street checks and integration with city surveillance networks, arguing that such systems amplify existing patterns of discriminatory policing and violate privacy and free expression. Indian legal and policy briefs on FRT note that international studies

have found higher error rates for darker-skinned and female faces; when layered atop existing biases in police targeting, this creates a risk of caste and religion-based discrimination in stops, searches and arrests.

Despite these risks, India lacks a dedicated statutory framework governing police use of FRT. Puttaswamy's proportionality test offers a constitutional benchmark, and some policy papers urge applying it to FRT deployments, including requirements of legality, legitimate aim, necessity, and procedural safeguards such as independent oversight and time-bound data retention. Yet, existing practice is largely governed by internal police protocols and vendor contracts rather than democratically debated legislation.

The India AI Governance Guidelines acknowledge algorithmic bias and call for sector regulators and ministries of home affairs and law to work with AISI on risk mitigation frameworks, but concrete, binding standards for FRT such as bans on real-time mass surveillance, mandatory accuracy and bias audits, or heightened standards for use in criminal trials have not yet been enacted.

4.3 AI in Courts and Public Administration

India's courts and administrative agencies are experimenting with AI tools from translation and speech-to-text systems used by the Supreme Court and High Courts, to decision-support dashboards for case management and e-governance. These deployments are often framed as improving efficiency and access to justice, particularly by supporting multilingual interfaces and aiding overburdened judges.

However, the absence of clear standards for when human decision-makers may rely on AI outputs raises concerns. The Kerala High Court's 2025

advisory, which prohibits AI tools from being used for judicial reasoning or decision-making and emphasises the risks of privacy violations and erosion of public trust, constitutes an early attempt to delineate boundaries. Still, such policies are state-level and non-binding elsewhere.

Scholars foresee possible uses of AI in sentencing recommendations, bail decisions, and risk assessments, drawing analogies to controversial tools like COMPAS in the United States. They warn that such systems could import hidden biases for instance, by using proxies correlated with caste or community status and undermine the presumption of innocence. The lack of explicit legislative safeguards, such as the right to contest AI-assisted decisions, strengthens the case for pre-emptive governance rather than ex post litigation.

4.4 Public-Sector Employment, HR and Workplace Management

Indian public institutions are also exploring AI in recruitment screening, promotions, and performance evaluation, often via vendor-provided platforms. Indian legal commentary notes that hiring algorithms trained on historical data may disadvantage women, Dalits, or candidates from less prestigious institutions, while opaque workplace monitoring tools could infringe privacy and association rights.

Because such tools are frequently procured through private vendors and embedded within broader enterprise systems, they may escape the visibility usually associated with marquee projects like Aadhaar or FRT. Yet they have direct implications for Articles 14–16 (equality and equal opportunity in public employment) and for labour rights more generally. Current governance measures such as generic data protection obligations and anti-discrimination laws do not directly address

algorithmic decision-making, leaving significant room for unregulated bias.

5. Regulatory Oversight and Gaps

5.1 Fragmentation of Legal Regimes

India's AI-related legal landscape is fragmented across data protection (DPDP Act and Rules), IT and intermediary regulations, sectoral statutes, and constitutional rights. The DPDP Act governs personal data but is largely agnostic to the specific risks of algorithmic decision-making, especially where decisions rely on non-personal or aggregated data. AI advisories under the IT Rules focus on content and platform governance, with limited attention to public-sector deployments; sectoral regulators like RBI are moving ahead with their own frameworks, but these are confined to specific industries.

This patchwork complicates oversight of cross-cutting public-sector systems such as Aadhaar-linked welfare databases or law-enforcement FRT. Without a dedicated AI statute or harmonised regulations, agencies may rely on internal policies, vendor contracts, or ad hoc judicial orders, leading to inconsistent safeguards and enforcement gaps.

5.2 Absence of Enforceable Rights in Automated Decision-Making

Perhaps the most significant gap for algorithmic bias is the absence of explicit, enforceable **rights in relation to automated decision-making**. Multiple comparative analyses highlight that, unlike the GDPR, the DPDP Act does not provide data principals with a general right not to be subject to decisions based solely on automated processing, nor a right to explanation or human intervention.

Draft DPDP Rules 2025 propose that Significant Data Fiduciaries conducting algorithmic processing

undertake “algorithm audits,” but they do not specify minimum standards for fairness, transparency or contestability, nor do they directly empower individuals to challenge algorithmic decisions in high-stakes contexts like welfare, credit, or policing.

Consequently, individuals harmed by biased public-sector AI systems must typically resort to constitutional litigation (Articles 32 or 226) or sector-specific complaints (e.g., banking ombudsman), which are resource-intensive and often ill-suited for systemic algorithmic harms.

5.3 Transparency, Procurement and Public–Private Partnerships

Another recurring theme in case studies is opacity stemming from procurement structures. When public agencies procure AI systems from private vendors, claims of trade secrecy and commercial confidentiality frequently impede disclosure of code, training data or detailed documentation, limiting the feasibility of independent audits by academics or civil society.

Freedom of Information (Right to Information) requests have occasionally yielded high-level documentation but rarely the “holy trinity” of data, documentation and code needed for rigorous algorithmic accountability. This undermines both external oversight and the ability of courts or regulators to assess whether systems comply with constitutional and statutory standards.

India AI Governance Guidelines recognise these challenges and call for clearer allocation of accountability across the AI value chain and enhanced transparency about actors’ roles and compliance. However, without procurement rules that condition public tenders on disclosure, bias

testing, and auditability requirements, these aspirations may not translate into practice.

5.4 Institutional Capacity and Mandates

The establishment of the AI Safety Institute (AISI) and proposed AI Governance Group (AIGG) suggests a shift towards more centralised oversight. But as of late 2025, these bodies are largely advisory. They lack clear statutory enforcement powers comparable to those of the Data Protection Board, which itself is focused on data protection rather than AI-specific harms.

Sector regulators (RBI, SEBI, TRAI, etc.) are expected to adapt the Guidelines within their domains, yet many lack established methodologies or technical capacity for AI audits, fairness evaluation or incident investigation. Technical initiatives like IIT Madras’s IndiCASA dataset and evaluation tools are promising, but their integration into formal regulatory processes is nascent.

5.5 Constitutional Remedies: Strengths and Limits

On the positive side, India’s constitutional framework is unusually rich in tools that can be mobilised against algorithmic bias:

- **Articles 14–16** ground challenges to discriminatory outcomes in public employment, welfare and policing.
- **Article 21**, read with *Puttaswamy*, protects privacy, dignity and informational self-determination against disproportionate digital surveillance and data practices.
- **Articles 32 and 226** empower the Supreme Court and High Courts to issue writs enforcing these rights, potentially including orders for algorithmic transparency, fairness audits, and human oversight.

Recent scholarship argues that courts could in principle require state agencies to disclose algorithmic logic or risk scores affecting individuals, mandate independent audits, or suspend biased systems. However, litigation is reactive, slow, and often case specific. It cannot substitute for ex ante governance mechanisms, particularly when harms are diffuse, probabilistic and embedded in complex digital infrastructures.

6. Towards a Robust AI Governance Framework for India's Public Sector

Drawing on the above analysis, this section outlines key elements of a more coherent and effective AI governance regime to address algorithmic bias in India's public sector.

6.1 Anchor Governance in Constitutional Equality and Due Process

First, AI governance in India must explicitly frame algorithmic bias as a **constitutional** problem, not merely a technical or managerial one. The India AI Governance Guidelines already root their principles in fairness, equity, accountability and "People First" design; these should be interpreted through the lens of Articles 14–16 and 21, as elaborated in *Puttaswamy* and the Aadhaar judgment.

This implies that:

- High-stakes public-sector AI deployments welfare eligibility, policing, criminal justice, public employment, credit guarantees must satisfy a **strict proportionality test**, including demonstration of necessity and absence of less restrictive means;
- Indirect discrimination arising from algorithmic systems (e.g., differential exclusion rates by caste or gender) should be

recognised as violating equality guarantees, regardless of intent;

- Procedural fairness requires not only notice and access to data, but also intelligible explanations and accessible routes to contestation for those affected by algorithmic decisions.

6.2 Enact a Rights-Based AI Law Complementing the DPDP Act

Second, India should move towards a dedicated, rights-based AI statute that complements the DPDP Act. Building on India AI Governance Guidelines and global models such as the EU AI Act and the U.S. AI Bill of Rights blueprint, such a law could:

- Define categories of AI risk, with **strict or prohibitive regimes** for certain public-sector uses (e.g., real-time biometric mass surveillance, social scoring, and opaque welfare risk scoring);
- Codify **individual rights** in relation to automated decision-making, including:
 - the right not to be subject to decisions based solely on automated processing in contexts with significant legal or socio-economic effects;
 - the right to human review and meaningful explanation;
 - the right to contest and seek rectification of algorithmic decisions;
- Mandate **algorithmic impact assessments (AIAs)** and **human-rights impact assessments** for high-risk public-sector systems, with ex ante publication and consultation, especially where marginalised communities are likely to be affected.

Such a statute need not simply transplant GDPR or EU AI Act standards but can adapt them to India's socio-legal context, including affirmative action obligations, linguistic diversity, and the centrality of digital public infrastructure.

6.3 Institutionalise Algorithmic Audits and Evaluation Ecosystems

Third, India should institutionalise **algorithmic fairness audits** and **evaluation ecosystems** for public-sector AI. Building on initiatives like IndiCASA and the AI Safety Institute, reforms could include:

- Requiring public agencies to conduct periodic **bias audits** for high-risk systems, with metrics disaggregated by caste, gender, religion, region, and disability status where lawful, and to publish summary findings;
- Empowering AISI to develop **standardised testing protocols** and **benchmark datasets** for common tasks (face recognition, risk scoring, language models in Indian languages), and to accredit independent auditors;
- Integrating algorithm audits envisaged for Significant Data Fiduciaries under DPDP Rules into a broader AI governance scheme, with explicit fairness criteria and sanctions for non-compliance.

Crucially, audits must be designed to avoid becoming mere box-ticking exercises. They should involve meaningful engagement with affected communities and civil society, and their results should inform procurement, funding and deployment decisions.

6.4 Reform Public-Procurement Rules for AI Systems

Fourth, procurement law and practice must be updated to embed algorithmic accountability. Standard contract terms for public-sector AI projects could require:

- Transparency clauses mandating disclosure of system documentation, data schemas, and model cards to regulators and (with appropriate safeguards) to independent auditors and courts;
- Bias and impact testing obligations prior to deployment and at regular intervals;
- Provisions guaranteeing that vendors cooperate with investigations into harms, including providing technical access necessary for audits;
- Sunset or review clauses that require re-authorisation of high-risk systems based on evidence of performance and fairness.

These reforms would operationalise the India AI Governance Guidelines' focus on accountability across the AI value chain and prevent vendors' IP claims from entirely eclipsing public-interest scrutiny.

6.5 Create Public Registers and Incident Reporting for Public-Sector AI

Fifth, transparency to the public could be enhanced through **AI system registers** and **incident reporting** mechanisms. The Guidelines already mention AI incident systems; India could:

- Require all public bodies to maintain and publish inventories of AI systems in use, with descriptions of purpose, data sources, risk assessments, and oversight mechanisms;
- Implement mandatory **incident reporting** for significant harms or near-misses, with anonymised public summaries and follow-up actions;
- Encourage whistleblower protections for civil servants or contractors who disclose harmful AI practices.

Such measures would not only support academic and civil society scrutiny but also build public trust, aligning with the “Trust is the Foundation” sutra.

6.6 Strengthen Sectoral Regulators and Judicial Capacity

Finally, sectoral regulators and the judiciary will need targeted capacity-building to engage with complex AI systems. This includes technical training, access to expert panels, and integration with AISI’s research.

Courts, in turn, could develop **model guidelines** for evaluating algorithmic evidence in criminal and civil cases (e.g., FRT matches, risk scores), including standards for admissibility, cross-examination of expert witnesses, and the weight assigned to AI outputs relative to other evidence. The Kerala High Court’s cautious policy could serve as a starting point, supplemented by jurisprudential development that recognises algorithmic bias as a justiciable equality concern.

7. Conclusion

India’s public sector sits at the forefront of AI-enabled governance in the Global South. Digital identity, payments, welfare systems, and law-enforcement technologies are being intertwined with machine learning systems that promise efficiency and fraud detection but simultaneously risk entrenching old inequalities in new, opaque ways. The case studies of Aadhaar-linked welfare, Samagra Vedika, and facial recognition in policing show that algorithmic bias in India is not a hypothetical concern; it manifests as exclusion from food rations, intensified surveillance of already marginalised communities, and questionable criminal prosecutions.

India’s AI governance architecture has evolved rapidly. From early strategy documents and ethical

principles, the state has moved towards a more comprehensive ecosystem: a national AI mission with a Safe & Trusted AI pillar, a dedicated AI Safety Institute, detailed AI Governance Guidelines, and a data protection regime with increasing operational detail. This trajectory reflects a conscious attempt to balance innovation with risk mitigation and global competitiveness with domestic rights commitments.

Yet this architecture remains incomplete. Fragmented laws, the absence of explicit rights in automated decision-making, weak procurement safeguards, and nascent institutional capacity mean that algorithmic bias in public-sector AI is still largely addressed *ex post*, if at all. The burden of proof often falls on those least able to bear it poor, marginalised communities whose everyday interactions with AI-enabled bureaucracies are shaped by power asymmetries and informational opacity.

To move from principles to practice, India must treat AI governance as an extension of its constitutional project: securing equality, dignity and due process in the digital age. A rights-based AI law, robust algorithmic audit infrastructure, reformed procurement rules, public registers and incident systems, and empowered institutions like AISI and sector regulators would together form a coherent oversight regime. Such a regime would not only curb algorithmic bias in India’s public sector but also position India as a norm-setter in the Global South demonstrating how large, diverse democracies can govern AI in ways that are innovative, context-sensitive and fundamentally committed to justice.

References

- Barocas, S., & Selbst, A. D. (2016). Big data’s disparate impact. *California Law Review*, 104(3), 671–732.

- Bharati, D. R. (2025). Bias and fairness in AI algorithms: Legal standards and governance. *SSRN Working Paper*.
- Dixon, P. (2017). A failure to “do no harm”: India’s Aadhaar biometric ID program and its inability to protect privacy. *Health and Technology*, 7(4), 539–567.
- IndiaAI. (2025). *India AI Governance Guidelines: Empowering ethical and responsible AI*. Ministry of Electronics and Information Technology.
- IndiaAI. (2024). *IndiaAI Mission and Safe & Trusted AI Pillar*. Government of India.
- International Institute of SDGs & Public Policy Research. (2025). *AI and data protection: Challenges in automated decision-making in India*.
- Johari, A. (2018). A year after Jharkhand girl died of starvation, Aadhaar tragedies are on the rise. *Scroll.in*.
- Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.
- Kerala High Court. (2025). *Policy regarding use of artificial intelligence tools in district judiciary*.
- Mayer Brown. (2023). *India passes privacy law: Digital Personal Data Protection Act 2023 compared with GDPR*.
- MeitY. (2024). *Advisory on use of AI models / LLMs / generative AI under IT Rules, 2021 (1 March 2024)*.
- MeitY. (2025). *Digital Personal Data Protection Rules, 2025*. Government of India.
- Mitigating Bias in AI through Indian Laws. (2025). *International Journal of Business and Management Invention*, 14(8), 6–10.
- Muralidharan, K., et al. (2020). Identity verification standards in welfare programs. *NBER Working Paper 26744*.
- NITI Aayog. (2018). *National Strategy for Artificial Intelligence (#AIforAll)*. Government of India.
- NITI Aayog. (2021). *Principles for responsible management of AI in India (Responsible AI for All)*. Government of India.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Office of the Principal Scientific Adviser. (2024). *Safe & Trusted AI pillar: Concept note for IndiaAI Safety Institute*.
- Pandey, A., & Shukla, U. (2025). Algorithms of oppression: Addressing AI bias in India. *NLIU Cyber Law & Technology Blog*.
- Rebelo, K. (2025). *How to hold algorithms accountable in India*. Reuters Institute for the Study of Journalism.
- Reserve Bank of India. (2025). *Framework for Responsible and Ethical Enablement of AI (FREE-AI)*. RBI Committee Report.
- Tapadia, P., & Rathore, R. (2025). AI and data protection: Challenges in automated decision-making. *IISPPR Blog*.
- Tiwari, S. (2025). Towards a rights-based AI framework in India: Bridging global models and constitutional duties. *LiveLaw*.

Amnesty International. (2024). *Entity resolution in India's welfare digitalisation: The case of Samagra Vedika*.

Amnesty International. (2022). *Ban the scan: Hyderabad – The dangers of facial recognition*.

Vipra, J. (2025). Towards AI sovereignty? The good, the bad, and the ugly of AI policy in India. *African Journal of Information and Communication*.

Access Partnership. (2023). The key policy frameworks governing AI in India.

International Comparative Legal Guides. (2025). Data protection laws and regulations – India.

Centre for Responsible AI, IIT Madras. (2025). IndiCASA: An Indian-context bias evaluation dataset and toolset for language models.