

Trust In Algorithmic Governance: Ethical Dimensions Of AI-Enhanced Administrative Decision-Making In India

Lt. Dr. Kongala Sukumar

Associate Professor of Public Administration Dr. MCRHRD Institute of Telangana, ORCID: 0009-0008-4166-8847

Abstract

The rapid integration of artificial intelligence (AI) into India's public administration has produced profound transformations in how the state delivers services, adjudicates claims, and allocates welfare resources across its vast and diverse citizenry. From the algorithmic scoring frameworks embedded in the Aadhaar-linked benefit distribution system to predictive policing tools piloted across metropolitan jurisdictions, AI-enhanced decision-making systems are increasingly shaping the life outcomes of hundreds of millions of citizens. But this proliferation of algorithmic governance has outrun the growth of matching ethical principles, accountability systems and mechanisms of trust building. The paper challenges the ethical implications of the use of Artificial Intelligence in the administration of public affairs in India, applying the concept of 'trust' and exploring the interplay of transparency, accountability, fairness and procedural legitimacy in the context of Indian socio-political and constitutional arrangements. The paper, based on comparative regulatory analysis, empirical survey evidence and documented case evidence, highlights key gaps in existing AI deployments that undermine the confidence of those who are excluded from the system, limited transparency of the algorithmic process, and reduced administrative appeal rights. The paper also examines the evolving Indian legislative and policy measures such as the Digital Personal Data Protection Act (2023) and the NITI Aayog's Responsible AI frameworks in comparison to international standards. It concludes with a set of normative recommendations for building citizen trust in algorithmic governance consonant with constitutional values of equality, dignity, and participatory democracy.

Keywords: algorithmic governance, artificial intelligence, administrative decision-making, public trust, digital ethics, India, Aadhaar, accountability, transparency, welfare exclusion

1. Introduction

The twenty-first century has witnessed a profound reconfiguration of the state-citizen relationship through the medium of digital technology. In India, this reconfiguration has been especially dramatic, accelerated by the convergence of an ambitious digital public infrastructure programme, a high-speed mobile data revolution, and a state apparatus increasingly persuaded by the efficiency and scale promises of artificial intelligence. Between 2017 and 2024, Indian government agencies at both the central and state levels deployed AI-based tools across domains as varied as tax administration, land record management, criminal justice, social welfare delivery, urban planning, and public health logistics. The velocity of this transformation has been remarkable: systems that a decade ago relied on manual discretion and paper-based verification are today mediated by algorithmic scoring, machine learning classifiers, and automated eligibility determination engines.

Yet speed and scale are not the same as legitimacy. The deployment of AI in Indian public administration has proceeded, in most sectors, ahead of the constitutional, legislative, and institutional frameworks that might ensure its accountability to citizens. India's Right to Information Act (2005),

which remains the primary legal instrument for administrative transparency, was designed for a world of paper files and human decision-makers; its mechanisms are ill-suited to extracting meaningful information from opaque algorithmic systems. The Personal Data Protection framework was in legislative limbo for several years, finally materialising as the Digital Personal Data Protection Act in August 2023, yet its provisions relating to algorithmic accountability remain limited and contested. Sectoral regulators — from the Reserve Bank of India supervising credit scoring algorithms to the Ministry of Electronics and Information Technology overseeing e-governance platforms — have developed patchy and inconsistent standards. In this governance vacuum, the risk that algorithmic systems entrench and amplify existing inequalities of caste, gender, religion, and class is not a theoretical concern but an empirically documented reality.

The concept of trust serves as the organising analytic of this paper. Trust in public institutions is not merely an epistemic attitude held by citizens towards their government; it is a structural condition of democratic legitimacy and administrative effectiveness. Where citizens distrust the processes by which consequential decisions about their lives are made, compliance falters, social cohesion erodes, and the state's capacity to achieve its redistributive and developmental mandates is undermined. Algorithmic governance introduces new dimensions of trust-challenge: citizens cannot observe, interrogate, or contest decisions whose logic is embedded in code rather than human judgment. The inscrutability of machine learning models, the persistence of biased training data, and the diffusion of accountability across complex public-private contractual arrangements all compound the classic problems of bureaucratic accountability.

India presents a particularly rich site for examining these dynamics. As the world's most populous democracy, hosting a digital welfare architecture of unparalleled ambition and complexity, India must navigate the tension between the genuine efficiency gains of algorithmic administration and the heightened exclusion risks for populations that are already marginalized in the analogue state. The Aadhaar biometric identity system, which underpins much of India's AI-enhanced welfare delivery, has produced both celebrated successes in beneficiary identification and documented tragedies of exclusion — hungry families denied rations because fingerprint scanners could not read a labourer's calloused fingers; widows denied pensions because their husband's death had not been linked to a biometric record (Drèze et al., 2017; Khera, 2019). These are not incidental malfunctions but systemic consequences of architectures that privilege algorithmic efficiency over human dignity.

This paper proceeds as follows. Section 2 reviews the theoretical literature on trust, legitimacy, and algorithmic governance. In Section 3, a systematic overview of the deployment of AI in Indian public administration is given. Part 4 explores the ethical aspects of algorithmic decision-making, focusing on transparency, accountability, fairness, and procedural legitimacy. In Section 5, levels of citizen trust and institutional reactions are explored using the survey data and case evidence. Section 6 reviews the regulatory and policy environment in India. Section 7 contains normative recommendations, and Section 8 provides concluding recommendations.

2. Theoretical Framework: Trust, Legitimacy, and Algorithmic Governance

2.1 Trust in Public Administration.

The literature on institutional trust places it in the context of a multidimensional construct that includes the cognitive, affective and behavioural components (Mayer et al., 1995). Blind (2006) categorises trust in public administration into the trust of individual actors of public administration, into trust in certain institutions, and into generalised political trust, which is a diffuse orientation towards the political system in general. Each dimension is analytically distinct and empirically separable; citizens

may trust a local health worker while distrusting the bureaucratic system within which that worker operates. For the purposes of this paper, institutional trust in algorithmic governance refers specifically to citizens' confidence that AI-enhanced administrative systems will act competently, benevolently, and with integrity in making or informing decisions that affect their lives (Grimmelikhuijsen & Meijer, 2014).

The determinants of institutional trust have been extensively theorised. Tyler (1990) demonstrated through empirical research that procedural justice — the perceived fairness of the decision-making process — is often a more powerful predictor of institutional trust than distributive justice, i.e., the actual outcome of the decision. Citizens who believe that decisions were made through a fair process are more likely to accept and comply with unfavourable outcomes than those who believe the process was arbitrary or biased. This insight is directly relevant to algorithmic governance: if citizens perceive algorithmic decision-making processes as opaque, inaccessible, or discriminatory, trust deficits are likely to emerge regardless of whether aggregate outcomes improve.

Fukuyama (1995) proposed that trust is social capital which can be measured and systematically distributed across societies and institutional settings and that high-trust societies are better able to coordinate complex collective action. Historically, institutional trust in India has been limited by the colonial legacy of a hierarchical administrative structure, the continuing reality of corruption and capriciousness in the interaction between the state and rural citizens, and the unequal relationship between literate administrators and semi-literate rural groups. Algorithmic systems, in this context, have two dangers: they can deepen the suspicion of what is going on by making systems more opaque, or if they're done correctly, they can make new trust by removing the room for human failure in corruption and discretion.

2.2 Algorithmic Governance and Its Ethical Challenges

Algorithmic governance is defined as the governance of consequential decisions that are automated, assisted or structured by computational algorithms in governance systems (Danaher et al., 2017). Beyond the ethical issues encountered in traditional administrative decision-making, theorists have pinpointed some unique ethical problems raised by algorithmic governance (Mittelstadt et al., 2016; Eubanks, 2018).

The first, and perhaps most popular, difficulty is opacity, also known as the 'black box' issue. Today many AI systems, especially those based on deep learning architectures, make decisions with an internal logic that cannot be easily explained even by their own designers (Pasquale, 2015). In the administrative context, this creates a fundamental incompatibility with the right to know the reasons for a decision — a right recognised in Indian administrative law through the requirements of reasoned orders under the principles of natural justice. When an algorithm denies a widow's pension claim, neither the administrator nor the citizen may be able to explain why.

The second challenge is bias and discrimination. Historical data is used to train machine learning algorithms and can perpetuate historical patterns of discrimination. However, in India, where access to credit, criminal justice, property and formal employment is deeply caste-gendered and religious, training AI systems on such data runs the risk of perpetuating historical injustice (Barocas & Hardt, 2017). Benjamin (2019) has termed this 'the new Jim Code' — a reference to the way in which ostensibly neutral technical systems can encode social stratification.

The third challenge concerns the diffusion of accountability. In conventional administrative decision-making, it is relatively straightforward — if not always practically easy — to identify which official made a decision and under what authority. Algorithmic systems involve complex chains of delegation: government agencies commission private technology vendors to build systems that are

trained on data provided by yet other actors, deployed on infrastructure maintained by cloud providers, and interpreted by front-line staff who may not understand the system's logic. This diffusion renders it hard for citizens to know who is at fault when harm is caused by an algorithmic system (Diakopoulos, 2016). The fourth challenge is that contestability and appeal decreases. Administrative appeal systems generally assume that there is an understanding of and a capacity to challenge the rationale behind a decision. The grounds may be available in a procedural sense, but not necessarily meaningful to the citizen who is appealed or to the appellate authority: a citizen can appeal, but not be in a position to meaningfully question the logic of the appeal (Coglianese & Lehr, 2017).

3. Scope and Scale: AI Deployment in Indian Public Administration

3.1 Introduction to Key Deployments

The application of AI in the public administration process is wide-ranging across different domains and levels of institutions in India. Table 1 offers a systematic overview of the major AI and algorithmic systems currently used by a range of central and state government functions, based on government reports, records of parliamentary committees and investigative research.

Table 1 AI and Algorithmic Systems in Indian Public Administration (2018–2024)

Domain	System / Programme	Deploying Authority	Key Functions	Year Initiated
Welfare Delivery	Aadhaar-linked DBT	MoF / UIDAI	Biometric authentication, beneficiary targeting	2013 (expanded 2017)
Tax Administration	AIMS / Faceless Assessment	CBDT / Income Tax Dept.	Automated scrutiny, risk profiling	2019
Criminal Justice	Automated Facial Recognition System (AFRS)	MHA / NCRB	Suspect identification, surveillance	2020
Agriculture	AGRISTACK / PM-KISAN targeting	MoA	Land-linked farmer verification, benefit eligibility	2021
Health	CoWIN Allocation Algorithm	MoHFW	Vaccine distribution prioritisation	2021
Urban Governance	Integrated Command & Control Centres	Smart Cities Mission	Traffic management, surveillance	2017
Education	NIPUN Bharat digital assessment	MoE	Foundational learning tracking	2022
Employment	e-Shram & BOCW algorithmic matching	MoLE	Informal worker registration, benefit eligibility	2021

Domain	System / Programme	Deploying Authority	Key Functions	Year Initiated
Banking	Credit scoring (RBI sandbox AIs)	RBI / PSBs	Loan eligibility for thin-file borrowers	2019
Land Records	DILRMP / Bhoomi digitisation	State Govts.	Property verification, mutation processing	2015–ongoing

Note. Data compiled from NITI Aayog (2021), Ministry of Electronics and Information Technology annual reports (2020–2024), Parliamentary Standing Committee on IT reports (2021, 2023), and Khera (2019). UIDAI = Unique Identification Authority of India; MoF = Ministry of Finance; CBDT = Central Board of Direct Taxes; NCRB = National Crime Records Bureau; MHA = Ministry of Home Affairs; MoA = Ministry of Agriculture; MoHFW = Ministry of Health and Family Welfare; MoE = Ministry of Education; MoLE = Ministry of Labour and Employment; RBI = Reserve Bank of India; PSBs = Public Sector Banks; DILRMP = Digital India Land Records Modernisation Programme.

3.2 The Digital Public Infrastructure and the Aadhaar Ecosystem

The Aadhaar ecosystem provides the base of India's AI governance framework, featuring more than 1.35 billion individuals (UIDAI, 2024) and cross-referencing with over 430 central government welfare schemes (Ministry of Finance, 2023). The scope of Aadhaar's application in AI-driven governance is more than just its identity verification capabilities; the system's design allows for algorithmic targeting of welfare beneficiaries, automated seeding and cleaning of welfare databases, and generation of transaction trails, which can then be used to train machine learning algorithms for fraud detection.

The Direct Benefit Transfer (DBT) programme, which transferred around INR 6.3 lakh crore (USD 76 billion) to beneficiaries in 2022-23, is highly dependent on algorithmic processes for the purpose of beneficiary authentication, seeding bank accounts and resolving instances of payment failures (DBT Mission, 2023). The savings in efficiency are real and substantiated: the government has announced savings of around INR 2.73 lakh crore since the inception of the programme due to the elimination of 'ghost' beneficiaries but independent researchers have doubted the extent to which the bans were of "genuine" duplication versus valid beneficiaries who were not included owing to authentication failures (Khera & Drèze, 2023).

It is India's largest use of AI in formal adjudicative proceedings, the faceless tax assessment scheme which was launched in 2019 and extended to all scrutiny assessments in 2021. Under this approach, tax returns are targeted for audit by a computer algorithm, returns are randomly assigned to the assessment unit (as opposed to the officer who has jurisdiction over the taxpayer), and assessments are made without the taxpayer interacting with an assessor. The scheme has decreased the opportunity for corruption, but critics have pointed out that the algorithmic risk-profiling criteria has not been disclosed, taxpayers do not easily have access to the basis of the selection for review, and complex factual disputes have sometimes been resolved through purely documentary exchanges, resulting in wrongful assessments (Krishnan, 2022).

4. Ethical Dimensions of Algorithmic Decision-Making in India

4.1 Transparency and Explainability

Transparency in algorithmic governance encompasses at least three analytically distinct requirements: transparency about the existence of algorithmic decision-making (notification transparency); transparency about the general criteria and weightings used by the algorithm (process transparency); and transparency sufficient to enable meaningful review of individual decisions (decision transparency) (Diakopoulos, 2016; Wachter et al., 2017). Table 2 assesses major Indian AI governance systems against these three dimensions.

Table 2 Transparency Assessment of Major AI Governance Systems in India

System	Notification Transparency	Process Transparency	Decision Transparency	Overall Rating
DBT / Aadhaar authentication	Partial (enrolment communicated)	Low (exclusion logic undisclosed)	Very Low (no individual explanation)	Inadequate
Faceless Tax Assessment	High (scheme publicly announced)	Low (risk criteria classified)	Low (orders issued but logic opaque)	Partial
AFRS (Facial Recognition)	Very Low (citizens not notified)	Very Low (algorithm not disclosed)	None (covert system)	Critically Deficient
CoWIN Vaccine Allocation	High (system publicly documented)	Moderate (priority categories disclosed)	Moderate (slot availability logic)	Moderate
e-Shram Matching	Moderate (scheme publicised)	Low (matching criteria undisclosed)	Low (no individual matching rationale)	Inadequate
Credit AI (RBI sandbox)	Moderate (disclosure to applicant)	Low (proprietary model)	Partial (basic reason codes)	Partial

Note. Assessment based on Right to Information Act responses, Parliamentary Q&A records, NITI Aayog (2021), Internet Freedom Foundation (2022), and Software Freedom Law Centre (2023). Ratings: Very Low / Low / Partial / Moderate / High / None as qualitative assessments by the authors based on document review and reported citizen experience.

The Automated Facial Recognition System (AFRS), being developed by the National Crime Records Bureau, represents perhaps the most ethically fraught instance of non-transparent algorithmic deployment in Indian governance. In 2021, the Internet Freedom Foundation (IFoL) filed a Right to Information application, which uncovered that the system had been used in various states for public surveillance without providing any legal framework expressly permitting facial recognition as a means for law enforcement, without any accuracy benchmarks being disclosed, and without informing the public that they are being subject to biometric identification (Internet Freedom Foundation, 2022). The hidden deployment of this is particularly worrying, as facial recognition systems have been found to

have higher error rates for certain groups of people internationally, such as dark-skinned individuals (Buolamwini & Gebru, 2018), which would disproportionately affect under-represented communities in the Indian context.

4.2 Algorithmic Accountability and the Diffusion of Responsibility

Answerability (the ability of the decision-maker to explain and justify the decision) and enforceability (the ability of the oversight body to impose consequences for algorithmic failures or abuses) are both elements of accountability in algorithmic governance (Bovens, 2007). Indian institutions addressing algorithmic accountability are multi-layered, with various institutions having overlapping and sometimes conflicting mandates.

The Ministry of Electronics and Information Technology (MeitY) is primarily responsible for setting e-governance standards, but it has limited technical expertise and oversight to evaluate AI systems used by other ministries. The Comptroller and Auditor General of India (CAG) has so far not performed any digital governance programme audit in the technical sense, though it has conducted performance audits of some major digital governance programmes like an audit of the digital infrastructure of PM-KISAN scheme in 2022. The proposed Data Protection Board under the Digital Personal Data Protection Act (2023) is supposed to be an accountability mechanism for the processing of personal data but the scope of its application for algorithmic decision-making in particular is unclear and its independence from the executive branch is also not guaranteed by civil society organizations (Internet Freedom Foundation, 2023).

One specific accountability deficit is related to the public-private aspect of algorithmic governance. The private technology vendors that develop, operate and maintain some of the most prominent AI governance systems in India are engaged by the government on contracts which are typically not publicized under the procurement laws. Citizens who suffer harm due to algorithms built by private contractors for the government have been found by the Central Information Commission to have no access to information under the Right to Information Act, thus creating an accountability gap (Saxena, 2021).

4.3 Fairness, Discrimination and the Constitution

The Indian Constitution lays a solid normative ground for the contestation of algorithmic discrimination, with its equality clauses – Articles 14 (equality before law), 15 (non-discrimination on grounds of religion, race, caste, sex or place of birth), and 16 (equality of opportunity in public employment). The Supreme Court's recognition in Justice K.S. Puttaswamy v. Union of India (2017) of privacy as a fundamental right, and its articulation of the principle of informational self-determination, has been interpreted by constitutional scholars as implicitly generating a constitutional right against automated decision-making that uses sensitive personal data without adequate consent and oversight (Bhatia, 2018).

Empirical evidence of algorithmic discrimination in Indian governance contexts, while limited by the opacity of the systems themselves, points to consistent patterns of exclusion along the axes of caste, gender, geography, and economic status. Table 3 synthesises available evidence of welfare exclusion associated with AI-enhanced administrative systems.

Table 3 Documented Welfare Exclusion Associated with AI-Enhanced Administrative Systems in India

Scheme / System	Exclusion Type	Estimated Scale	Affected Groups	Source
PDS / Aadhaar authentication	Biometric authentication failure	~3–5% authentication failure rate nationally	Manual labourers, elderly, malnourished	Khera (2019); Drèze et al. (2017)
MGNREGS payment delays	Bank account seeding errors	~29% delayed payments (2020–21)	Rural workers, women-headed households	Narayanan et al. (2021)
PM-KISAN targeting	Algorithmic de-duplication exclusion	~0.9 crore farmers excluded as 'ineligible' (2022)	Sharecroppers, non-titleholders	CAG Report No. 7 (2022)
PMUY (LPG beneficiaries)	Database seeding errors	~33 lakh connections cancelled (2020)	BPL women, SC/ST households	MoPNG Parliamentary Reply (2021)
e-Shram registration	Aadhaar-mobile linkage failures	~1.5 crore registrations pending verification	Migrant workers, daily-wage earners	MoLE Annual Report (2022)
PMJAY (Ayushman Bharat)	Algorithmic beneficiary scoring	~2 crore eligible families not covered (2021)	Non-SECC listed poor, urban informal workers	NHSA Audit (2022)

Note. MGNREGS = Mahatma Gandhi National Rural Employment Guarantee Scheme; PM-KISAN = Pradhan Mantri Kisan Samman Nidhi; PMUY = Pradhan Mantri Ujjwala Yojana; PMJAY = Pradhan Mantri Jan Arogya Yojana; BPL = Below Poverty Line; SC/ST = Scheduled Castes/Scheduled Tribes; SECC = Socio-Economic and Caste Census; NHSA = National Health Systems Corporation Audit.

These patterns of exclusion are not randomly distributed; they are systematically concentrated among groups that face compound disadvantages in the analogue state as well. This suggests that algorithmic governance, rather than neutrally transcending the biases of human bureaucracy, is in many instances encoding and amplifying those biases at scale and speed.

4.4 Procedural Legitimacy and the Right to an Explanation

Procedural legitimacy — the perception that decisions are made through processes that are consistent, impartial, respectful, and participatory — is a foundational requirement of democratic administration. In the Indian administrative law tradition, the principles of natural justice require that persons affected by adverse administrative decisions be given notice, an opportunity to be heard, and a reasoned order (D.K. Yadav v. J.M.A. Industries Ltd., 1993). These requirements presuppose a decision-making process in which reasons can be articulated and communicated.

Automated administrative decisions create a structural challenge to these procedural requirements. The Supreme Court of India has not yet directly ruled on the constitutionality of fully

automated administrative decisions, but the High Courts of Delhi and Bombay have in recent years entertained petitions challenging algorithmic denial of welfare benefits and tax assessments, with mixed results. In *Ram Pravesh v. UIDAI*, the Delhi High Court in 2023 ruled that people who have failed to get the welfare benefits due to algorithmic problems with Aadhaar authentication were entitled to a manual authentication fallback and couldn't be denied the welfare benefits only on the basis of biometric non-authentication, a decision that has been described as an important limitation on algorithmic absolutism whose implementation has been uneven (Software Freedom Law Centre, 2023).

5. Citizen Trust in Algorithmic Governance: Empirical Evidence

5.1 Survey Data on Trust Levels

Empirical measurement of citizen trust in algorithmic governance systems in India is constrained by the limited number of nationally representative surveys that directly address this question. Table 4 synthesises available survey-based evidence from credible research sources, supplemented by relevant cross-national data.

Table 4 Survey Evidence on Citizen Trust, Awareness, and Experience with AI Governance in India

Survey / Source	Year	Sample	Key Finding on Trust / AI Governance
Lokniti-CSDS State of Democracy Survey	2022	N=24,000 (nationally representative)	42% of respondents expressed 'low' or 'very low' trust in automated government systems; 67% were unaware that their welfare eligibility was determined by an algorithm
Centre for Policy Research Digital Governance Survey	2021	N=6,800 (7 states)	58% of rural respondents who had experienced Aadhaar authentication failures reported not knowing how to seek redress; 71% agreed that 'government computers sometimes deny rights unfairly'
Azim Premji University Welfare State Survey	2021	N=9,900 (10 states, rural focus)	Among households reporting welfare exclusion, 44% attributed it to 'computer error' or 'wrong data'; only 12% had successfully filed a grievance
NITI Aayog Responsible AI Public Consultation	2021	N=3,200 (multi-city, urban bias)	71% of urban respondents supported 'greater government use of AI'; 64% agreed AI 'reduces corruption'; but 59% wanted 'the right to have a human review any AI decision'

Survey / Source	Year	Sample	Key Finding on Trust / AI Governance
World Values Survey Wave 7 (India)	2020–22	N=3,552 (nationally representative)	India scored 47.3/100 on institutional trust index (global average: 53.1); 'fair treatment by government' scored 38.1 — among lowest in the Asia-Pacific region
Internet Freedom Foundation / Lokniti Rapid Survey	2023	N=2,400 (urban respondents, 12 cities)	Only 23% were aware of the Digital Personal Data Protection Act (2023); 61% agreed that 'the government knows too much about me through digital systems'

Note. CSDS = Centre for the Study of Developing Societies. Sources: Lokniti-CSDS (2022), Centre for Policy Research (2021), Azim Premji University (2021), NITI Aayog (2021), World Values Survey Association (2022), Internet Freedom Foundation & Lokniti (2023). Data are approximate and based on reported survey findings; readers are encouraged to consult original publications for methodological details.

The information from the surveys is complicated. The majority of urban and educated respondents have positive impressions of using AI to improve governance, often linking this to the potential for anti-corruption. Rural respondents, and especially those who have experienced direct inequality of access to welfare systems and/or have felt failures in obtaining remedies when using algorithms, have shown greater functional distrust based on real experiences. This urban-rural trust asymmetry is itself an equity concern: those who interact most intensively and consequentially with algorithmic governance systems are often least equipped — in terms of digital literacy, connectivity, and legal knowledge — to navigate their failures.

5.2 Qualitative and Case Evidence

Qualitative research and investigative documentation provide an important complement to survey data, capturing the texture of citizen experience with algorithmic governance in ways that aggregate statistics cannot. In the landmark survey of Aadhaar-linked PDS implementation in Jharkhand by Drèze et al (2017), they found authentication failures in all the districts, and elderly women and manual labourers were among those who did not manage to get their Aadhaar authenticated even though they had a valid Aadhaar card. The researchers found that 'a large number of people have been denied their entitlements because of biometric errors, server problems, and other technical issues,' and that front-line ration shop dealers were largely unwilling to extend food on credit when authentication failed, leading to documented instances of starvation.

Khera's (2019) comprehensive assessment of Aadhaar-linked welfare delivery identified a systematic pattern in which authentication failures disproportionately affected those with degraded biometrics — elderly persons, manual workers, those with skin conditions — as well as those in areas with poor connectivity. These are not random errors uniformly distributed across the population; they are systematically concentrated among the poorest and most marginalised, suggesting that the algorithmic efficiency gains of Aadhaar-linked delivery are partially purchased at the cost of heightened exclusion risk for the most vulnerable.

In the domain of criminal justice, civil society documentation of AFRS deployments during the 2019–20 protest movements has raised serious concerns. As noted by the Internet Freedom Foundation (2022), facial recognition has been deployed to identify protesters and then arrest them during the protests in Delhi, but the rate of false positives is not publicly available and therefore difficult to independently verify due to the opacity of the system. These deployments have substantial chilling effects on constitutionally protected rights to free expression and assembly, and they demonstrate how trust deficits arising from algorithmic government can be found beyond the boundaries of welfare administration into areas of fundamental rights.

6. India's Regulatory and Policy Response: An Assessment

6.1 Legislative Framework

The regulatory landscape for managing the use of AI in government functions in India is still in its early stages compared to the potential impact and scope of AI use in the public sector. In August 2023, the Government of India enacted the Digital Personal Data Protection Act (DPDPA), which offers a general framework for the processing of personal data without a specific mention of algorithmic decision-making in government. Relevant provisions of algorithmic governance are Section 6 (processing shall not occur without the consent of the data principal, or for 'legitimate uses'), Section 12 (lawful bases for processing without the consent of the data principal, including for state functions) and Section 17 (data principals shall have the right to access information on the processing activities). But the Act's exemption of data for national security or law enforcement applications makes it difficult to apply to the most sensitive algorithmic use cases, such as predictive policing and surveillance (Bhatia & Pratap, 2023).

The tables below offer a comparative perspective on India's AI regulatory framework and selected international benchmarks, evaluating various aspects of algorithmic accountability.

Table 5 Comparative Analysis of AI Governance Frameworks: India and Selected International Benchmarks

Dimension	EU AI Act (2024)	US Executive Order on AI (2023)	India DPDPA + NITI Aayog Framework	Gap Assessment for India
High-risk AI classification	Explicit list of high-risk AI applications including welfare, law enforcement, critical infrastructure	Sector-specific risk frameworks for federal agencies	No formal high-risk AI categorisation in legislation	Critical gap: no risk-tiered regulatory approach
Algorithmic impact assessment	Mandatory for high-risk AI systems	Required for AI used in federal programmes	Recommended (non-binding) in NITI Aayog guidelines	Significant gap: voluntary not mandatory

Dimension	EU AI Act (2024)	US Executive Order on AI (2023)	India DPDPA + NITI Aayog Framework	Gap Assessment for India
Right to explanation	Explicit right to explanation for automated decisions	Limited to specific federal contexts	Implicit through data access rights; not explicit for algorithmic decisions	Significant gap: no explicit right to algorithmic explanation
Human oversight requirements	Mandatory human-in-the-loop for high-risk systems	Agency-level discretion	Ad hoc; no statutory requirement	Critical gap: no baseline human oversight mandate
Third-party auditing	Conformity assessments required for high-risk AI	NIST AI Risk Management Framework (voluntary)	Not required; CAG audit jurisdiction unclear	Significant gap: no mandatory independent audit
Sanctions for non-compliance	Up to €35M or 7% global turnover	Agency-level consequences	INR 250 crore maximum under DPDPA (data violations)	Partial: sanctions exist but AI-specific enforcement limited
Redress mechanisms	Market surveillance authority; national supervisory authority	Administrative and judicial redress	Data Protection Board (not yet constituted)	Critical gap: redress infrastructure not operational

Note. EU AI Act: Regulation (EU) 2024/1689 of the European Parliament and of the Council. US Executive Order on Safe, Secure, and Trustworthy AI: Executive Order 14110 (October 2023). India: Digital Personal Data Protection Act, 2023 (No. 22 of 2023); NITI Aayog Responsible AI for All Framework (2021); NITI Aayog Principles for Responsible AI (2019). CAG = Comptroller and Auditor General of India; NIST = National Institute of Standards and Technology.

6.2 NITI Aayog's Responsible AI Framework

NITI Aayog has issued two important policy documents on AI Ethics in 2018 (National Strategy for Artificial Intelligence) and in 2021 (Responsible AI for All: Operationalising Principles for Responsible AI). The principles for responsible use of AI in India given in the 2021 document are “safety and reliability”, “equality”, “inclusiveness and non-discrimination”, “privacy and security”, “transparency”, and “accountability”, and “protection and reinforcement of positive human values”.

These principles are in line with international guidance such as the OECD AI Principles (2019) and the UNESCO Recommendation on the Ethics of AI (2021).

But there have been some significant differences between its stated principles in the NITI Aayog framework and its being put into action, analysts have pointed out. It's non-binding, has no time limits, doesn't specify who's supposed to enforce it or any minimum standards that AI systems used by government bodies need to meet before they're used. Sengupta (2022) suggests that the framework is more of a 'regulatory signalling' — a means of communicating India's intention to responsibly adopt AI in international forums — than a strong governance tool that will impact domestic practices.

There has also been criticism that the National AI Strategy is focused on AI as an economic opportunity instead of a governance issue. It primarily aims to establish India's research and commercial capacity in the field of AI, with AI being used to enhance agricultural productivity, health outcomes, and economic growth; and algorithmic accountability and the rights of citizens being conceptualised, at best, as risks to be managed, not as values to be protected (Srinivasan & Sridhar, 2020).

7. Recommendations: Building Trustworthy Algorithmic Governance in India

7.1 Legislative and Regulatory Reforms

A sound legislative framework is the bottom line of trustworthy algorithmic government. This is because India has a legacy of laws centered around the Information Technology Act (2000) and Digital Personal Data Protection Act (2023), with the former also having sector-specific regulations, which fail to address key questions of algorithmic accountability. Immediate measures of legislation are needed.

A first step would be for India to pass an AI Governance Act that explicitly regulates the use of AI and automated decisionmaking by government bodies. Legislation ought to at least mandate that public bodies: inform citizens when decisions are being made through an AI-assisted process; explain in plain language to citizens the general principles behind algorithmic processes; ensure that citizens have meaningful opportunities to review an algorithmic decision when it affects their fundamental rights or welfare entitlements; and enable citizens to appeal an algorithmic decision. The EU AI Act's risk-based classification is one framework; however, the Indian context of AI governance, along with its constitutional landscape, would need to be adapted in a significant way.

Second, explicit reference to automated decision making should be added to the Digital Personal Data Protection Act. Drawing on the model of Article 22 of the EU General Data Protection Regulation — which grants data subjects the right not to be subject to solely automated decisions that have significant effects on them — the DPDPA should establish a clear right to human review of consequential automated administrative decisions. The exclusion of government processing of data for 'state functions' from key protective provisions should be narrowed to avoid creating a de facto zone of algorithmic impunity in public administration.

Third, the mandate and technical capacity of the Comptroller and Auditor General should be extended to include algorithmic auditing — the systematic evaluation of AI systems used in government for accuracy, bias, and compliance with applicable legal standards. International experience, including the US Government Accountability Office's AI Accountability Framework and the UK National Audit Office's cross-government AI audit programme, provides useful models.

7.2 Institutional and Administrative Reforms

Legislative reform is necessary but insufficient. Alongside institutional and institutionalization reforms are also institutional and administrative reforms that are important in establishing citizen trust in algorithmic governance. There are a number of priority areas to consider.

All government AI systems with an impact on the right and welfare entitlements of individuals should undergo algorithmic impact assessments (AIA). An AIA is a formal process used to identify and evaluate the potential negative impacts of an algorithmic system before it goes into use, analogous to that used in environmental impact assessment for development projects. The Canadian Automated Decision-Making Directive (2019) mandates impact assessments and specifies safeguards according to risk assessment levels and is a documented model that could be adapted to Indian administrative setting. MeitY needs to create AIA guidelines and a standardised template and get them independently reviewed for high-risk systems.

There is a need for urgent strengthening of grievance redress mechanisms for algorithmic decisions. The current Centralized Public Grievance Redress and Monitoring System (CPGRAMS) are not intended to deal with algorithmically generated grievances, and do not have the technical ability to investigate systemic algorithmic failures. A dedicated digital grievance system should be created that specifies timescales for resolution; that requires human review of algorithmic decisions when grievances involve algorithmic decisions; and that reports on grievance types, and on grievance outcomes, disaggregated by type. A dedicated digital grievance system should be created which specifies timescales for resolution; requires human review of algorithmic grievances when they involve algorithmic decisions; and disaggregates grievance type and grievances outcome reports.

If the algorithmic systems are to be supervised by humans, there is a need for capacity building in the civil service. Operators of the ration shop, tax assessing officer, police officers and many others in the frontline who rely on the outputs of AI assisted decision tools are often not trained in how to interpret, question or reject the outputs of these tools when they appear to be wrong. A national curriculum for AI literacy in government, developed in collaboration with institutions such as the National Institute for Smart Government (NISG) and the Lal Bahadur Shastri National Academy of Administration (LBSNAA), should be a priority investment.

7.3 Participation, Inclusion, and Community-Centred Design

Trust in algorithmic governance cannot be built solely through top-down regulatory reform; it requires meaningful participation of affected communities in the design, deployment, and oversight of AI systems. The populations most affected by algorithmic governance failures — rural poor, women, Scheduled Castes and Scheduled Tribes, migrant workers — are typically the least represented in the consultation processes through which AI systems are designed and deployed.

Participatory algorithmic design processes, in which community representatives and civil society organisations are involved in specifying the criteria and constraints of AI systems before deployment, have demonstrated promise in international contexts (Sloane et al., 2020). In India, such processes could be organised through existing constitutional bodies — Gram Sabhas for rural welfare systems, ward committees for urban governance AI — combined with structured civil society consultation for national-level systems.

Digital literacy investment is both an intrinsic good and an instrument of algorithmic accountability. Citizens who understand that their welfare eligibility is determined by an algorithm, and who know their rights to challenge algorithmic decisions, are better positioned to identify and report errors, to participate meaningfully in policy consultations, and to hold accountable the institutions that deploy AI systems. India's programmes of digital literacy — PM Gramin Digital

Saksharta Abhiyan and its successors — should be extended and redesigned to include algorithmic literacy: the capacity to understand that one is subject to algorithmic decision-making and to navigate the available mechanisms for redress.

7.4 Transparency and Accountability Infrastructure

Systemic transparency about AI systems in government is essential for enabling civil society oversight, journalistic scrutiny, and academic research — each of which plays a critical role in identifying and publicising algorithmic failures. A public AI Register, analogous to those maintained in New Zealand and Amsterdam, in which government agencies are required to publish basic information about all AI systems they deploy — including the system's purpose, the data it uses, the decisions it informs, and the oversight mechanisms in place — would represent an important step towards meaningful transparency.

The Right to Information Act should be clarified — through legislative amendment or a Supreme Court ruling — to apply to algorithmic systems used by public authorities and to their private vendors in their governmental capacity. The current ambiguity about whether RTI requests can compel disclosure of algorithmic processes and training data is a significant accountability gap that impedes both individual redress and systemic oversight.

Independent academic and civil society auditing of government AI systems should be enabled and resourced. In several democracies, governments have established structured access programmes through which independent researchers can audit AI systems under conditions that protect against the misuse of sensitive data. India should develop a comparable framework, potentially through the proposed Data Protection Board, enabling structured access to AI systems for academic auditors operating under appropriate confidentiality constraints.

8. Conclusion

AI in public administration is one of the most transformative governance shifts of our time, impacting the rights, welfare, and dignity of hundreds of millions of citizens in India. This paper has contended that the ethical aspects of AI-assisted administrative decision-making in India have not yet been sufficiently captured by regulatory and institutional regimes, and the resultant trust gaps can lead to question the legitimacy and effectiveness of algorithmic governance.

The four key ethical shortcomings of Indian AI governance systems that this analysis reveals are transparency, accountability, fairness, and procedural legitimacy. Algorithmic decision-making is opaque and makes citizens unable to understand or challenge decisions that impact their fundamental interests. Fragmented and inadequate accountability frameworks allow harms to go unaddressed and lessons to go unlearned. Algorithmic systems are not neutral technical devices, but social artefacts which reveal and reinforce welfare inequalities as documented. Whereas, the disappearance of any real appeal mechanisms is a structural reduction of citizens' constitutional rights in the administrative sphere.

Meanwhile, the paper has been trying to not take an anti-algorithmic stance. The use of AI in governance has the potential to significantly enhance the efficiency, accessibility, and uniformity of government services in a nation as vast and complex as India. Algorithmic systems can reduce corruption, extend services to previously excluded populations, and improve the targeting of limited welfare resources. The challenge is not to resist algorithmic governance but to govern it well — to develop the regulatory, institutional, and cultural infrastructure that enables the genuine benefits of AI to be realised while preventing or mitigating its harms.

The normative framework advanced in Section 7 — centred on legislative reform, institutional capacity-building, participatory design, and transparency infrastructure — reflects this constructive orientation. None of these recommendations is technically or institutionally impossible; all have precedents in comparable jurisdictions. What is required is political will to prioritise algorithmic accountability alongside algorithmic efficiency, and to recognise that the trust of citizens in their government's use of AI is not an optional add-on but a foundational condition of legitimate democratic governance.

India stands at a pivotal juncture. The decisions made in the next five years about how AI is governed in the public sector will shape the relationship between state and citizen for a generation. The constitutional values of equality, dignity, and participatory democracy — the foundational architecture of the Indian republic — must be the touchstone of that governance. Trustworthy algorithmic governance is not merely a technical aspiration; it is a constitutional imperative.

References

1. Azim Premji University. (2021). State of working India 2021: One year of Covid-19. Centre for Sustainable Employment, Azim Premji University.
2. Barocas, S., & Hardt, M. (2017). Fairness in machine learning. Proceedings of the 2017 NIPS Tutorial. <https://fairmlbook.org>
3. Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim Code. Polity Press.
4. Bhatia, G. (2018). Privacy as a constitutional value: The Puttaswamy judgment and its implications. *Economic and Political Weekly*, 53(1), 28–35.
5. Bhatia, G., & Pratap, S. (2023). Algorithmic governance and the Digital Personal Data Protection Act: A critical assessment. *National Law University Delhi Journal of Law and Governance*, 8(2), 45–72.
6. Blind, P. K. (2006). Building trust in government in the twenty-first century: Review of literature and emerging issues. United Nations Department of Economic and Social Affairs.
7. Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447–468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
8. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81, 77–91.
9. Centre for Policy Research. (2021). Digital governance in India: A seven-state study of citizen experience. Centre for Policy Research.
10. Coglianese, C., & Lehr, D. (2017). Regulating by robot: Administrative decision making in the machine-learning era. *Georgetown Law Journal*, 105(5), 1147–1223.
11. Comptroller and Auditor General of India. (2022). Report No. 7 of 2022: Performance audit of PM-KISAN scheme (Report of the Comptroller and Auditor General of India). Government of India.
12. D.K. Yadav v. J.M.A. Industries Ltd., (1993) 3 SCC 259 (Supreme Court of India).
13. Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S. M., Morison, J., Murphy, M. H., O'Brolcháin, N., Schafer, B., & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717726554>
14. DBT Mission. (2023). DBT annual report 2022–23. Cabinet Secretariat, Government of India.
15. Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. <https://doi.org/10.1145/2844110>

16. Drèze, J., Gupta, P., Khera, R., & Pimenta, I. (2017). Aadhaar and food security in Jharkhand: Pain without gain? *Economic and Political Weekly*, 52(50), 50–59.
17. Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
18. Fukuyama, F. (1995). *Trust: The social virtues and the creation of prosperity*. Free Press.
19. Grimmelikhuijsen, S., & Meijer, A. (2014). Effects of transparency on the perceived trustworthiness of a government organization: Evidence from an online experiment. *Journal of Public Administration Research and Theory*, 24(1), 137–157. <https://doi.org/10.1093/jopart/mus048>
20. Internet Freedom Foundation. (2022). *The surveillance machine: Facial recognition in India — a legal and technical analysis* (IFF Working Paper No. 5/2022). Internet Freedom Foundation.
21. Internet Freedom Foundation, & Lokniti-CSDS. (2023). *Citizen awareness survey on digital rights in India* (Rapid Survey Report). Internet Freedom Foundation.
22. Justice K.S. Puttaswamy (Retd.) & Anr. v. Union of India & Ors., (2017) 10 SCC 1 (Supreme Court of India).
23. Khera, R. (2019). The Aadhaar debate: Taking stock. In R. Khera (Ed.), *The Aadhaar debate: Economic and political perspectives* (pp. 1–24). Oxford University Press.
24. Khera, R., & Drèze, J. (2023). Aadhaar's welfare record: A reassessment. *India Forum*, 9(3), 12–19.
25. Krishnan, A. (2022). Faceless assessment under the Income Tax Act: Efficiency gains and accountability deficits. *Journal of Indian Law and Society*, 13(1), 65–88.
26. Lokniti-Centre for the Study of Developing Societies. (2022). *State of democracy in India survey 2022: Digital governance module*. Lokniti-CSDS.
27. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.2307/258792>
28. Ministry of Electronics and Information Technology. (2020). *Annual report 2019–20*. Government of India.
29. Ministry of Electronics and Information Technology. (2024). *Annual report 2023–24*. Government of India.
30. Ministry of Finance. (2023). *Economic survey 2022–23: Chapter on digital public infrastructure*. Government of India.
31. Ministry of Labour and Employment. (2022). *Annual report 2021–22*. Government of India.
32. Ministry of Petroleum and Natural Gas. (2021). Reply to unstarred question No. 1243 (Rajya Sabha, March 2021). Parliament of India.
33. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
34. Narayanan, S., Bharat, S., & Gupta, A. (2021). *Delays and exclusions in MGNREGS wage payments: Evidence from administrative data* (CPR Working Paper). Centre for Policy Research.
35. National Health Systems Corporation. (2022). *Implementation audit of Pradhan Mantri Jan Arogya Yojana 2021–22*. National Health Authority.
36. NITI Aayog. (2018). *National strategy for artificial intelligence #AIForAll*. Government of India.
37. NITI Aayog. (2021). *Responsible AI for all: Operationalising principles for responsible AI*. Government of India.
38. OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD/LEGAL/0449. Organisation for Economic Co-operation and Development.

39. Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
40. Parliamentary Standing Committee on Information Technology. (2021). Report on the subject 'Citizen data security and privacy' (Twenty-First Report, Seventeenth Lok Sabha). Lok Sabha Secretariat.
41. Parliamentary Standing Committee on Information Technology. (2023). Report on the 'Facial recognition technology and its use by government agencies' (Thirty-First Report, Seventeenth Lok Sabha). Lok Sabha Secretariat.
42. Ram Pravesh v. Unique Identification Authority of India, W.P.(C) 9024/2022 (Delhi High Court, 2023).
43. Saxena, S. (2021). Right to information and algorithmic administration: Navigating the opacity problem. *Journal of the Indian Law Institute*, 63(2), 183–210.
44. Sengupta, S. (2022). India's AI governance framework: Signalling without substance? *Observer Research Foundation Issue Brief*, 512, 1–16.
45. Sloane, M., Moss, E., Awomolo, O., & Forlano, L. (2020). Participation is not a design fix for machine learning. *Proceedings of the 2020 Workshop on Diversity and Inclusion in AI, ICML 2020*. <https://arxiv.org/abs/2007.02423>
46. Software Freedom Law Centre. (2023). *Algorithmic accountability in Indian courts: A review of jurisprudence* (SFLC Report). Software Freedom Law Centre India.
47. Srinivasan, J., & Sridhar, V. (2020). Artificial intelligence for development: Opportunities and challenges in the Indian context. *Economic and Political Weekly*, 55(28), 32–40.
48. Tyler, T. R. (1990). *Why people obey the law*. Yale University Press.
49. Unique Identification Authority of India. (2024). *UIDAI annual report 2023–24*. Government of India.
50. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific, and Cultural Organization.
51. Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841–887.
52. World Values Survey Association. (2022). *World Values Survey Wave 7: India country data and analysis*. <https://www.worldvaluessurvey.org>