International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Algorithmic Trust Deficits: Public Perceptions of AI-Enabled Administrative Decision-Making in India

Lt. Kongala Sukumar

Assistant Professor of Public Administration
MALD Government Degree College,
GADWAL, Pin-509126, Telangana

Abstract:
As Indian public administration increasingly deploys artificial intelligence (AI) to automate welfare targeting, law enforcement, tax assessment, and service delivery, questions of public trust have become central to the legitimacy of digital governance. This paper examines the phenomenon of "algorithmic trust deficits" – the gap between the technical promise of AI-enabled administrative systems and citizens' willingness to accept their outcomes – in the Indian context. Drawing on a synthesis of policy documents, scholarly literature, and secondary survey evidence, the paper develops a conceptual framework identifying five interacting drivers of trust deficits: opacity of decision logic, weak statutory accountability, algorithmic bias against marginalised groups, inadequate data protection, and limited grievance redressal. Using case illustrations from Aadhaar-linked welfare authentication, the Crime and Criminal Tracking Network and Systems (CCTNS), and automated toll-collection and surveillance systems, the paper argues that India's AI governance architecture – anchored in NITI Aayog's non-binding "AI for All" strategy – has outpaced the legal and institutional safeguards needed to sustain public confidence. The paper concludes with policy recommendations centred on algorithmic auditing, participatory design, and statutory accountability, arguing that trust, rather than technical performance alone, will determine the long-term legitimacy of algorithmic governance in India.

Keywords: algorithmic accountability; artificial intelligence governance; public trust; administrative law; digital governance; India.

1. Introduction

Over the past decade, the Indian state has embraced artificial intelligence (AI) and algorithmic tools as instruments of administrative modernisation. From biometric authentication for welfare disbursal under Aadhaar, to predictive policing and facial recognition under the Crime and Criminal Tracking Network and Systems (CCTNS), to automated toll collection through FASTag and algorithmic case triage in the judiciary through the Supreme Court Portal for Assistance in Court Efficiency (SUPACE), algorithmic systems now mediate a growing share of citizens' encounters with the state (NITI Aayog, 2018). These systems are typically justified in the language of efficiency, objectivity, and the reduction of corruption and discretion-based leakage. Yet the same qualities that make algorithmic decision-making attractive to

administrators – speed, scale, and apparent neutrality – are precisely what make it opaque and difficult to contest for the citizens subject to it (Pasquale, 2015; Yeung, 2018).

This paper is concerned with a specific and under-examined dimension of this shift: the erosion, or non-formation, of public trust in AI-enabled administrative decision-making. Trust is not a peripheral or merely reputational concern for algorithmic governance; it is constitutive of administrative legitimacy. Classical administrative law doctrine in India – rooted in principles of natural justice, reasoned decision-making, and the right to be heard – assumes a human decision-maker who can be questioned, whose reasoning can be probed, and who can be held accountable through appeal or judicial review (Centre for Law and Policy Research [CLPR], 2022). When decisions are delegated, wholly or partly, to statistical models and machine-learning classifiers, these assumptions come under strain. Citizens frequently cannot ascertain why a welfare claim was rejected, why a name was flagged in a watchlist, or why a subsidy was withheld after an Aadhaar authentication failure (Eubanks, 2018).

The concept of a "trust deficit" is used here to capture the widening gap between the technocratic promise of AI-enabled governance and the lived experience of citizens who interact with it, particularly those from marginalised and low-income groups who are disproportionately dependent on state welfare systems and therefore disproportionately exposed to algorithmic error (Eubanks, 2018; O'Neil, 2016). This trust deficit is distinct from, though related to, general scepticism about digital technology; it is specifically about the perceived fairness, accountability, and contestability of decisions once they are mediated by an algorithm rather than a visible human official.

The paper addresses three research questions. First, what are the principal drivers of algorithmic trust deficits in Indian public administration, as identified in existing scholarship and policy documents? Second, how do India's existing AI governance instruments – principally NITI Aayog's National Strategy for Artificial Intelligence and its Responsible AI approach papers – address, or fail to address, these drivers? Third, what institutional and legal reforms would be required to narrow the trust deficit without foreclosing the efficiency gains that motivate AI adoption in the first place?

The paper proceeds as follows. Section 2 reviews the international and Indian literature on algorithmic accountability, administrative discretion, and public trust in automated systems. Section 3 develops a conceptual "trust deficit" framework comprising five interacting dimensions. Section 4 outlines the paper's methodology, which relies on document analysis and synthesis of secondary survey evidence rather than primary data collection. Section 5 presents the analysis, organised around case illustrations and comparative tables. Section 6 discusses the findings in relation to India's constitutional administrative law tradition. Section 7 offers policy recommendations, and Section 8 concludes.

The stakes of this inquiry extend beyond individual instances of administrative error. Public trust in the institutions of the state is a cumulative resource, built and depleted over repeated interactions, and algorithmic systems – precisely because of their scale – have the capacity to affect that resource unusually quickly. A single Aadhaar authentication failure denying a family its monthly ration is, from the perspective of the implementing department, a negligible error rate within an otherwise highly efficient system; from the perspective of the affected family, it may be the defining experience of their relationship with the digital state (Eubanks, 2018). Understanding how such experiences aggregate into broader patterns of trust or distrust is therefore essential not only to the welfare of individual citizens but to the sustainability of the broader digital governance project that successive Indian governments have pursued since the launch of the Digital India programme.

It is also worth clarifying at the outset what this paper does not claim. It does not argue that algorithmic tools are inherently illegitimate instruments of governance, nor does it call for a moratorium on their use. The efficiency, consistency, and anti-corruption benefits associated with automation are real and, in a country of India's scale and administrative capacity constraints, often considerable (NITI Aayog, 2018). The argument instead is a narrower and, the author suggests, more actionable one: that the legitimacy and durability of these benefits depend on institutional investments – in transparency, accountability, fairness monitoring, data protection, and redressal – that have thus far lagged behind the pace of technological deployment. The paper's conceptual framework and policy recommendations are offered in that constructive spirit, as a diagnostic tool for policymakers and administrators seeking to sustain, rather than undermine, public confidence in digital governance.

2. Literature Review

2.1 Algorithmic Accountability: The Global Debate

The scholarly literature on algorithmic accountability emerged largely from American and European administrative and technology law, and has since diversified into a genuinely global field. Citron (2008) coined the phrase "technological due process" to describe the erosion of procedural safeguards when automated systems replace human adjudicators in benefits administration, arguing that opaque scoring systems deny claimants a meaningful opportunity to be heard. Citron and Pasquale (2014) extended this argument to what they termed "the scored society," in which algorithmic scores increasingly determine access to credit, employment, and public services without corresponding rights of explanation or correction. Pasquale (2015) generalised this critique through the metaphor of the "black box," describing how proprietary and technically complex systems resist scrutiny by the very citizens and regulators they affect.

A parallel strand of scholarship has focused on defining accountability itself. Bovens (2007) offers an influential conceptual framework in which accountability requires an actor to explain and justify conduct to a forum that can pose questions and render judgement, with potential consequences for the actor. Applied to algorithmic systems, this framework exposes a structural problem: the "actor" is often distributed across a vendor, a data set, a model, and an implementing agency, making it difficult to identify who, precisely, must explain and justify a given automated decision (Kroll et al., 2017; Diakopoulos, 2016). Diakopoulos (2016) further distinguishes between accountability for the design of an algorithm and accountability for its situated use, a distinction that is particularly salient in government contexts where off-the-shelf or vendor-supplied models are deployed within pre-existing bureaucratic structures.

Yeung (2018) situates algorithmic decision-making within a broader typology of "algorithmic regulation," distinguishing systems that merely inform human decision-makers from those that directly determine outcomes with minimal human intervention. She argues that the latter – which increasingly characterises high-volume administrative contexts such as tax assessment, welfare eligibility, and traffic enforcement – raises the most acute accountability concerns because the human "in the loop" becomes a formality rather than a genuine check. Selbst and Barocas (2018) add a further complication: even where an algorithm's logic can technically be disclosed, the underlying statistical reasoning may not be intuitively explicable to affected citizens, administrators, or even courts, producing what they term the gap between explainability and intuitive understanding.

O'Neil (2016) and Eubanks (2018) shift the focus from procedural design to distributive impact, documenting how algorithmic systems in policing, education, and welfare administration in the United

States systematically disadvantage low-income and minority populations, often while claiming objectivity. Eubanks' case studies of automated eligibility systems are particularly relevant to the Indian context because they describe a similar dynamic to that documented around Aadhaar-linked welfare authentication: technical failures (biometric mismatches, connectivity failures, database errors) are experienced by claimants as arbitrary denials of entitlement, with no accessible mechanism for explanation or appeal. Zuboff (2019) situates these dynamics within a broader critique of "surveillance capitalism," arguing that the asymmetry of information between data-holding institutions and data subjects is itself a driver of diminished trust, independent of any specific instance of algorithmic error.

On the question of trust specifically, Floridi et al. (2018) propose an ethical framework for AI organised around the principles of beneficence, non-maleficence, autonomy, justice, and explicability, arguing that public trust in AI systems is contingent on their being auditable against these principles by an independent body. Engin and Treleaven (2019) examine "algorithmic government" as an emerging paradigm of public administration and note that its legitimacy depends on civil servants retaining sufficient understanding of the systems they operate to defend decisions to the public – a condition they find frequently unmet in practice.

2.2 The Indian Context: Digital Governance and Its Discontents

India's engagement with algorithmic governance has been shaped by the scale and ambition of its digital public infrastructure. The Aadhaar biometric identification system, the largest such programme in the world, has become the backbone for authenticating access to subsidised food grain, cooking gas, pensions, and a wide range of welfare schemes. Scholarship on Aadhaar-linked welfare delivery has documented recurring instances of authentication failure – arising from worn fingerprints, poor network connectivity in rural areas, or database mismatches – resulting in the denial of entitlements to intended beneficiaries, a phenomenon that predates but is structurally analogous to algorithmic denial of service in other jurisdictions (Eubanks, 2018; CLPR, 2022).

At the policy level, NITI Aayog's National Strategy for Artificial Intelligence (2018), branded "AI for All," identifies healthcare, agriculture, education, smart cities and infrastructure, and smart mobility as priority sectors, and explicitly names "privacy and security" and the absence of "formal regulation around data anonymisation" as barriers to responsible adoption (NITI Aayog, 2018). This was followed by the Responsible AI approach papers (NITI Aayog, 2021), which articulate system-level principles – including safety, equality, inclusivity, and accountability – alongside societal-level concerns about the impact of automation on employment. Civil society responses to these documents, notably the Centre for Internet and Society's discussion paper (Abraham et al., 2018), welcomed the strategy's ambition while cautioning that it lacked binding legal force, enforceable audit requirements, or a designated regulator, leaving implementation to the discretion of individual ministries and departments.

Legal scholarship on administrative discretion in India has begun to grapple explicitly with algorithmic decision-making's implications for constitutional administrative law. Analyses of automated systems such as CCTNS and FASTag argue that India's administrative law framework – built around principles of reasoned decision-making, proportionality, and natural justice – lacks the doctrinal vocabulary to address decisions generated by statistical models rather than identifiable officials exercising discretion (CLPR, 2022). This body of work concludes that a dedicated legal framework, or at minimum statutory amendments establishing algorithmic accountability standards and avenues for judicial review, is required to reconcile automation with constitutional norms of fair administrative action.

A smaller but growing literature addresses public perception directly. Multi-country survey research on trust in AI has consistently found that respondents in emerging economies, including India, report comparatively high levels of general optimism about AI's benefits alongside significant unease about specific applications such as facial recognition, employment screening, and law enforcement (Smith, 2017; Edelman, 2022). The Edelman Trust Barometer (2022), a long-running cross-national survey of institutional trust, has documented a broader erosion of public confidence in government's capacity to manage the societal consequences of emerging technology, a finding that provides useful context for, though is not specific to, AI-enabled administration in India. Chatterjee (2020) surveys adoption challenges facing India's AI strategy and identifies public awareness, data governance capacity, and inter-departmental coordination as recurring constraints, while a wider review by Dwivedi et al. (2021) situates such trust and adoption challenges within the global literature on AI's multidisciplinary policy implications.

Taken together, this literature suggests that algorithmic trust deficits are not merely a matter of technical system design but arise from the interaction of technical opacity with pre-existing institutional weaknesses – limited grievance redressal capacity, uneven digital literacy, and an administrative law tradition not yet adapted to automated decision-making. It is this interaction that the conceptual framework in Section 3 seeks to capture.

2.3 From Discretion to Automation: A Shifting Administrative Law Problem

A further strand of relevant literature concerns the changing nature of administrative discretion itself. Classical Indian administrative law jurisprudence developed extensive doctrine around the exercise, and abuse, of discretionary power by identifiable officials – doctrines requiring reasoned orders, prohibiting arbitrariness, and guaranteeing an opportunity to be heard before an adverse decision is finalised (CLPR, 2022). Algorithmic systems do not eliminate discretion so much as relocate it: decisions about which variables to include in a model, how to weight them, and what threshold constitutes a positive or negative outcome are themselves discretionary choices, but they are made once, upstream, by data scientists and system designers, rather than repeatedly, downstream, by front-line officials (Yeung, 2018; Kroll et al., 2017). This relocation has two consequences of direct relevance to the trust-deficit framework developed in this paper. First, it removes discretion from the point of contact between citizen and state, where historically it was at least visible and, in principle, contestable in the moment. Second, it concentrates the consequences of design choices across potentially millions of subsequent transactions, such that a single biased or poorly calibrated parameter can produce disparate outcomes at a scale no individual official's discretion ever could (O'Neil, 2016).

This relocation of discretion has prompted calls, both internationally and within the Indian legal academy, for administrative law doctrine to evolve – for instance, by treating the design specifications and training data of a deployed algorithmic system as themselves subject to disclosure and review, analogous to the file notings and internal deliberations that have historically been subject to scrutiny under right-to-information regimes (CLPR, 2022; Kroll et al., 2017). Whether and how Indian courts will extend existing administrative law doctrine to this new terrain remains, at the time of writing, a largely open question, and one that bears directly on whether the accountability dimension of the trust-deficit framework can be addressed through litigation or will instead require dedicated legislative intervention.

3. Conceptual Framework: The Algorithmic Trust Deficit Model

Building on Bovens' (2007) accountability framework and the algorithmic accountability literature reviewed above, this paper proposes a five-dimensional model of the algorithmic trust deficit in public administration. The model treats trust as a function of citizens' assessment of a system along five interacting dimensions: transparency, accountability, fairness, data protection, and redressal. A deficit along any single dimension can undermine overall trust, and deficits tend to compound: opacity makes bias harder to detect, weak accountability makes redressal harder to obtain, and inadequate data protection amplifies the perceived risk of both.

Transparency refers to the extent to which the existence, purpose, and general logic of an algorithmic system are disclosed to the public and to affected individuals. This is distinct from full technical explainability, which may be neither feasible nor necessary; what matters for trust is that citizens know an algorithm is involved, understand in general terms what it does, and can access this information without specialised expertise (Selbst & Barocas, 2018).

Accountability refers to the existence of an identifiable actor – whether a government department, a statutory authority, or a specific official – who can be required to explain and justify an algorithmic decision, and who bears consequences for erroneous or harmful outcomes (Bovens, 2007; Diakopoulos, 2016). In the Indian context, the diffusion of responsibility across ministries, implementing agencies, and private technology vendors often obscures this line of accountability (Abraham et al., 2018).

Fairness concerns whether an algorithmic system produces systematically disparate outcomes for particular social groups – by caste, gender, religion, disability, region, or income – and whether such disparities are actively monitored and corrected (O'Neil, 2016; Eubanks, 2018). Given India's pronounced social stratification, fairness concerns are especially salient for systems governing access to welfare, employment, and criminal justice.

Data protection concerns the safeguards governing the collection, storage, sharing, and security of the personal data on which algorithmic systems depend. In the absence of a comprehensive statutory data protection framework at the time of writing, concerns about data protection in India are closely linked to concerns about function creep – the use of data collected for one purpose (such as welfare authentication) for another (such as surveillance) (Zuboff, 2019; NITI Aayog, 2018).

Redressal refers to the practical availability of a mechanism through which a citizen who believes an algorithmic decision to be wrong can seek review, correction, or compensation. Formal rights of appeal are of limited value if the redressal process itself requires digital literacy, documentation, or travel that is inaccessible to the affected population – a concern repeatedly raised in relation to Aadhaar-linked welfare denials (Eubanks, 2018; CLPR, 2022).

Table 1 summarises this framework, mapping each dimension to illustrative Indian administrative contexts and the corresponding scholarly or policy source from which the concern is drawn.

Table 1. Five-Dimensional Framework of the Algorithmic Trust Deficit

Dimension	Core Concern	Illustrative Indian Context	Key Source(s)
Transparency	Citizens are unaware an algorithm is involved,	Aadhaar authentication failures with no disclosed reason code	Pasquale (2015); Selbst & Barocas (2018)

	or cannot access its general logic		
Accountability	No identifiable, answerable actor for an automated decision	Vendor-supplied facial recognition tools used by state police	Bovens (2007); Abraham et al. (2018)
Fairness	Disparate impact on marginalised groups, uncorrected over time	Watchlist and predictive-policing flags under CCTNS	O'Neil (2016); Eubanks (2018)
Data protection	Absence of statutory limits on data collection, sharing, and reuse	Cross-linkage of Aadhaar data across welfare and surveillance systems	Zuboff (2019); NITI Aayog (2018)
Redressal	Formal appeal rights exist but are practically inaccessible	Denial of subsidised rations after biometric mismatch	CLPR (2022); Eubanks (2018)

Source: Author's synthesis, based on the cited literature and NITI Aayog (2018, 2021) policy documents.

4. Methodology

This paper adopts a qualitative, document-based research design combining doctrinal and policy analysis with a synthesis of secondary survey evidence. Primary data collection through original public opinion surveys was outside the scope of this study; instead, the paper triangulates three categories of source material.

First, primary policy documents were analysed, principally NITI Aayog's National Strategy for Artificial Intelligence (2018) and its Responsible AI approach papers (2021), together with associated civil society commentary (Abraham et al., 2018). These documents were examined for their explicit and implicit treatment of the five trust-deficit dimensions identified in Section 3.

Second, the paper draws on the growing body of Indian legal and public administration scholarship addressing algorithmic decision-making, including analyses of Aadhaar-linked service delivery, predictive policing, and administrative law doctrine (CLPR, 2022; Eubanks, 2018, applied comparatively). Case illustrations of specific systems – Aadhaar authentication, CCTNS, FASTag, and SUPACE – were compiled from publicly available government and secondary sources to ground the conceptual framework in concrete administrative practice.

Third, secondary survey evidence on public attitudes toward AI and institutional trust was reviewed, including the Edelman Trust Barometer (2022) and Pew Research Center survey work on public attitudes toward automation (Smith, 2017). This evidence is used cautiously and comparatively: cross-national surveys of this kind are not designed to isolate attitudes toward AI-enabled public administration specifically, and their findings are treated in this paper as indicative context rather than as precise measurements of Indian administrative trust. This limitation is discussed further in Section 6.

The analysis in Section 5 proceeds thematically rather than case-by-case, organising evidence under each of the five trust-deficit dimensions and using the identified administrative systems as recurring

illustrations. This approach is consistent with comparable conceptual and policy-analytic studies of algorithmic governance (Yeung, 2018; Engin & Treleaven, 2019) and is appropriate given the paper's aim of synthesising a fragmented literature into a coherent framework applicable to the Indian context, rather than testing a specific causal hypothesis.

A limitation of this design, acknowledged at the outset, is that it cannot directly measure the magnitude of public trust deficits in India; it can only identify their structural drivers as documented in existing scholarship and policy debate, and situate them within available comparative survey evidence. Section 7's recommendation for systematic, India-specific survey research on algorithmic trust follows directly from this limitation.

It should also be noted that document-based and synthesis-oriented methodologies of this kind are well established in the algorithmic governance literature, where the rapid pace of technological deployment frequently outstrips the availability of large-scale, methodologically rigorous primary survey data, particularly outside North America and Europe (Yeung, 2018; Engin & Treleaven, 2019). The approach adopted here is therefore consistent with the current state of the field, while explicitly flagging – rather than concealing – the resulting evidentiary constraints, in line with standard practice in policy-analytic and doctrinal legal scholarship.

5. Analysis: Mapping the Trust Deficit Across Indian AI-Enabled Systems

5.1 Transparency: The Disclosure Gap

Across the systems examined, the most consistently documented driver of trust deficit is the absence of accessible disclosure. Aadhaar-linked authentication for welfare schemes typically returns a binary success or failure result to the operator at the point of service, with no explanation legible to the beneficiary of why a transaction failed – whether due to a biometric mismatch, a server timeout, or a data entry error at enrolment (Eubanks, 2018, applied to the Indian case by CLPR, 2022). This is compounded by the fact that many beneficiaries interact with the system through intermediaries – fair price shop dealers or banking correspondents – who themselves often lack the technical information to explain a failure. NITI Aayog's own strategy document acknowledges "low awareness for adoption of AI" as a barrier (NITI Aayog, 2018), but frames this primarily as an obstacle to AI uptake rather than as a right of the citizen subject to automated decisions.

Facial recognition and predictive policing tools integrated into CCTNS present a related but distinct transparency problem: their deployment is frequently not proactively disclosed to the public at all, such that awareness of their existence depends on investigative journalism, right-to-information requests, or parliamentary questions rather than routine administrative disclosure (Abraham et al., 2018).

5.2 Accountability: Diffused Responsibility

The accountability dimension is complicated in India, as elsewhere, by the layered involvement of government departments, statutory authorities, and private technology vendors in designing and operating algorithmic systems. Facial recognition and analytics tools used by state police forces are frequently procured from private vendors under contracts that are not always subject to public disclosure, making it difficult for citizens, researchers, or even legislators to identify the entity responsible for a given system's design choices (Abraham et al., 2018). Bovens' (2007) requirement that an identifiable actor be able to explain and justify conduct to a forum is difficult to satisfy when no single actor has full visibility into how a system was trained, validated, and deployed.

NITI Aayog's Responsible AI documents (2021) articulate accountability as a design principle but stop short of proposing a statutory mechanism – such as a designated regulator with audit and enforcement powers – through which accountability could be operationalised. This gap between principle and enforceable mechanism is a recurring theme in civil society critique of India's AI policy architecture (Abraham et al., 2018; CLPR, 2022).

5.3 Fairness: Disparate Impact on Vulnerable Groups

Fairness concerns in the Indian context are shaped by the country's social structure, in which access to welfare, employment, and legal protection is already mediated by caste, gender, religion, disability, and regional inequality. Algorithmic systems built on historical administrative data risk encoding these existing inequalities rather than correcting them (O'Neil, 2016). Biometric authentication failure rates, for instance, are not evenly distributed: manual labourers, elderly persons, and persons with disabilities are documented to experience disproportionately higher rates of fingerprint mismatch, meaning that a nominally neutral authentication requirement can function as a disproportionate barrier for precisely the populations most dependent on the welfare schemes it gates (Eubanks, 2018, applied comparatively; CLPR, 2022).

Predictive and watchlist-based policing tools raise a parallel concern: where historical arrest or surveillance data reflects prior policing patterns that were themselves unevenly distributed across communities, a system trained on that data risks reproducing and formalising the same unevenness under the appearance of statistical objectivity (O'Neil, 2016; Yeung, 2018).

5.4 Data Protection: Function Creep and Cross-Linkage

A recurring concern in Indian public discourse around Aadhaar and related systems is function creep: data collected for one stated purpose, such as welfare authentication, being subsequently linked or repurposed for another, such as tax administration, telecom verification, or law enforcement (Zuboff, 2019, applied to the Indian case). NITI Aayog's own strategy document identifies the "absence of enabling data ecosystems" alongside "privacy and security" concerns and the lack of "formal regulations around anonymisation of data" as structural barriers to responsible AI adoption (NITI Aayog, 2018). In the absence of a comprehensive, dedicated statutory data protection framework at the time of this paper's writing, cross-linkage of large administrative databases proceeds largely on the basis of departmental notifications and memoranda of understanding rather than a uniform, independently enforced legal standard (Abraham et al., 2018).

5.5 Redressal: The Accessibility Gap

Even where formal grievance redressal mechanisms exist – helplines, online grievance portals, or appellate procedures – their practical accessibility to the populations most affected by algorithmic denial is frequently limited by digital literacy, language, documentation requirements, and geographic distance from a functioning grievance office (Eubanks, 2018; CLPR, 2022). This produces a paradox noted in the broader accountability literature: the existence of a formal right of appeal can itself become a source of legitimacy for a system, even where the right is rarely or unevenly exercised in practice (Citron, 2008; Citron & Pasquale, 2014).

Table 2 consolidates illustrative evidence across four widely discussed AI-enabled administrative systems in India, mapped against the five trust-deficit dimensions, drawing on the sources discussed above.

Table 2. Illustrative Trust-Deficit Profile of Selected AI-Enabled Administrative Systems in India

System	Transparency	Accountability	Fairness Concern Documented	Redressal Mechanism
Aadhaar-linked welfare authentication	Low – failure reasons rarely disclosed to beneficiary	Diffused across UIDAI, issuing department, and banking correspondents	Yes – biometric mismatch among manual labourers, elderly, and disabled persons	Formal grievance portal exists; limited accessibility in rural areas
CCTNS / predictive policing tools	Low – deployment often not proactively disclosed	Diffused across state police and private vendors	Yes – risk of encoding historical policing patterns	Largely absent for algorithm-specific grievances
FASTag automated toll collection	Moderate – purpose is publicly known	Relatively clear – National Highways Authority of India and banking partners	Limited evidence documented	Standard banking/toll dispute channels
SUPACE case-assistance tool (judiciary)	Moderate – pilot status publicly acknowledged	Clear – Supreme Court e-Committee	Limited evidence documented; tool is advisory, not determinative	Judicial oversight retained; human judge remains decision-maker

Source: Author's synthesis, based on NITI Aayog (2018, 2021); Abraham et al. (2018); CLPR (2022); and Eubanks (2018), applied comparatively to the Indian context. Ratings are qualitative assessments derived from the cited literature, not primary survey measurements.

Comparative survey evidence, while not specific to Indian public administration, provides useful contextual grounding for the trust-deficit framework. The Edelman Trust Barometer (2022) has documented a persistent global pattern in which trust in government's capacity to manage the societal risks of emerging technology lags behind trust in the technology sector's capacity to innovate – a gap with direct relevance to state deployment of AI, where government is simultaneously the innovator and the

guarantor of citizens' rights. Pew Research Center survey work on public attitudes toward automated decision-making similarly finds that support for automation is strongly conditional on the perceived fairness and reviewability of the resulting decisions, rather than on generalised attitudes toward technology (Smith, 2017). Table 3 summarises the broad thematic pattern emerging from this comparative survey literature, alongside its relevance to the Indian case.

Table 3. Thematic Pattern in Comparative Survey Literature on Trust in Automated Decision-Making

Thematic Finding	Source	Relevance to Indian AI-Enabled Administration
Government lags behind the technology sector in perceived competence to manage technological risk	Edelman (2022)	Relevant to citizen confidence in ministries deploying AI without a dedicated regulator
Public support for automated decisions is conditional on perceived fairness and reviewability, not on general attitudes to technology	Smith (2017)	Suggests redressal accessibility, not technical sophistication, is the more decisive trust driver
Emerging-economy respondents often express relatively high generalised optimism about AI's benefits alongside specific application-level concerns	Cross-national trust-in-AI survey literature (see Section 2.2)	Consistent with a pattern of broad support for digital governance ambitions alongside specific unease about surveillance-adjacent uses

Source: Author's synthesis of comparative survey literature. Figures are not India-specific unless otherwise noted; see Section 6 for a discussion of this limitation.

5.6 Comparative Regulatory Snapshot

Placing India's approach in comparative perspective clarifies the distinctiveness of its current regulatory posture. Table 4 compares India's AI governance architecture, as of this paper's writing, with the regulatory approaches then under discussion in the European Union and the United States, drawing on the accountability and regulatory design literature reviewed in Section 2 (Yeung, 2018; Floridi et al., 2018).

Table 4. Comparative Snapshot of AI Governance Approaches

Jurisdiction	Primary Instrument	Legal Status	Accountability Mechanism
India	NITI Aayog National Strategy for AI (2018) and Responsible AI approach papers (2021)	Non-binding policy guidance	No dedicated statutory AI regulator; sectoral ministries retain discretion
European Union	Proposed risk-tiered AI regulation then under negotiation	Moving toward binding regulation	Proposed conformity assessment and market-surveillance authorities
United States	Sectoral executive and agency guidance	Largely non-binding at federal level, with sectoral exceptions	Existing sectoral regulators (e.g., consumer protection, civil rights bodies) applying general statutes

Source: Author's synthesis, based on NITI Aayog (2018, 2021); Yeung (2018); and Floridi et al. (2018). Status reflects the regulatory landscape as understood at the time of writing and is illustrative rather than exhaustive.

5.7 Sectoral Variation: Why Some Systems Attract More Trust Than Others

A notable pattern across the case illustrations in Table 2 is that trust deficits are not uniform across sectors but appear to track two underlying variables: the visibility of the affected right, and the degree to which a human decision-maker remains meaningfully in the loop. FASTag, for instance, automates a comparatively low-stakes transaction – toll payment – in which the consequences of error are financially bounded and readily reviewable through existing banking dispute mechanisms; it is correspondingly the least contested of the four systems examined in this paper. SUPACE, similarly, is explicitly designed as a research and case-management aid to a human judge rather than a determinative decision-maker, preserving judicial accountability and, with it, a channel of recourse that predates the introduction of any algorithmic tool (Engin & Treleaven, 2019).

By contrast, Aadhaar-linked welfare authentication and CCTNS-integrated policing tools both involve high-stakes, rights-affecting decisions – access to subsistence entitlements and exposure to law-enforcement scrutiny, respectively – in which the human official's role has in practice been substantially narrowed to accepting or rejecting a system-generated output, with limited practical capacity to override it (Eubanks, 2018, applied comparatively; Abraham et al., 2018). This pattern suggests a broader design principle for policymakers: the acceptability of automation is not a fixed property of a technology but varies with the stakes of the decision it mediates and the degree of meaningful human oversight retained. Systems that automate low-stakes, easily reviewable transactions can reasonably operate with lighter-touch governance; systems that mediate access to subsistence entitlements, liberty, or due process

protections warrant the fuller suite of transparency, accountability, fairness, data protection, and redressal safeguards proposed in Section 7 (Yeung, 2018; Floridi et al., 2018).

6. Discussion

The analysis in Section 5 suggests that India's algorithmic trust deficit is best understood not as a single failure but as the cumulative effect of gaps across all five dimensions of the proposed framework, reinforced by the country's existing administrative law tradition being poorly adapted to automated decision-making (CLPR, 2022). This has two important implications.

First, the trust deficit is unlikely to be resolved by improvements in technical accuracy alone. Even a highly accurate algorithmic system can generate a trust deficit if its operation is opaque, its accountability diffuse, and its redressal mechanisms inaccessible – conditions under which errors, however rare, become disproportionately salient and difficult to correct (Selbst & Barocas, 2018; Diakopoulos, 2016). This suggests that policy attention focused narrowly on improving model performance, while necessary, is insufficient without corresponding institutional investment in disclosure, accountability, and redressal infrastructure.

Second, the distributive dimension of the trust deficit deserves particular emphasis in the Indian context. Because AI-enabled systems are most heavily deployed in welfare administration and law enforcement – domains in which the poorest and most socially marginalised citizens have the most frequent and highest-stakes interactions with the state – failures of transparency, fairness, and redressal are likely to fall disproportionately on populations with the least capacity to contest them (Eubanks, 2018; O'Neil, 2016). This has the effect of concentrating the costs of algorithmic error among citizens least able to absorb them, a pattern with clear implications for constitutional guarantees of equality and non-arbitrariness in state action (CLPR, 2022).

The comparative evidence discussed in Table 4 further suggests that India's current reliance on non-binding strategic guidance, while allowing flexibility and rapid adoption, leaves the accountability dimension of the trust deficit structurally unaddressed relative to jurisdictions moving toward binding, risk-tiered regulation. This is consistent with the civil society critique that non-binding principles, however well-articulated, cannot substitute for enforceable audit and redressal obligations (Abraham et al., 2018).

A caveat is warranted regarding the survey evidence discussed in Table 3. Because this evidence is drawn from cross-national studies not designed specifically around Indian public administration, it should be read as indicative context for the conceptual framework rather than as a precise measurement of trust in any specific Indian system. This is itself a finding of the present study: the absence of systematic, India-specific survey research on public trust in AI-enabled administration is a significant evidentiary gap, and one that constrains the precision with which policymakers can currently target interventions.

Finally, the sectoral variation identified in Section 5.7 suggests that a single, uniform governance standard applied indiscriminately across all AI-enabled administrative systems would be both administratively burdensome and analytically imprecise. A risk-tiered approach – calibrating the intensity of transparency, accountability, and redressal obligations to the stakes of the decision and the degree of human oversight retained – is likely to be both more proportionate and more politically sustainable than a blanket regulatory regime, a conclusion broadly consistent with the direction of comparative regulatory development discussed in Table 4 (Yeung, 2018).

7. Policy Recommendations

Drawing on the five-dimensional framework and the preceding analysis, this paper offers five policy recommendations directed at narrowing the algorithmic trust deficit in Indian public administration without foreclosing the legitimate efficiency and inclusion gains that motivate AI adoption.

First, mandatory disclosure standards should require that any AI-enabled administrative system affecting individual entitlements be publicly listed, with a plain-language description of its purpose, the categories of data it uses, and the general basis on which it reaches a decision, addressing the transparency dimension identified in Section 5.1 (Selbst & Barocas, 2018).

Second, a designated accountability mechanism – whether a cross-sectoral AI regulator or clearly assigned statutory responsibility within existing regulators – should be established, so that an identifiable actor can be required to explain and justify the design and operation of a given system, consistent with Bovens' (2007) accountability criteria and the recommendation of civil society commentary on NITI Aayog's strategy (Abraham et al., 2018).

Third, mandatory pre-deployment and periodic post-deployment fairness audits should be required for systems used in welfare, policing, and criminal justice, with disaggregated outcome monitoring by caste, gender, disability, and region to detect and correct disparate impact, addressing the fairness dimension identified in Section 5.3 (O'Neil, 2016; Eubanks, 2018).

Fourth, a comprehensive statutory data protection framework, with specific limits on cross-linkage and repurposing of administrative data collected for welfare authentication, should be enacted and independently enforced, addressing the data protection dimension identified in Section 5.4 (Zuboff, 2019; NITI Aayog, 2018).

Fifth, redressal infrastructure should be redesigned around accessibility rather than formal availability alone – including offline and low-literacy grievance channels, defined response timelines, and a presumption in favour of the citizen where a system's failure cannot be promptly explained – addressing the redressal dimension identified in Section 5.5 (Eubanks, 2018; CLPR, 2022). Alongside these five measures, the paper recommends that government agencies and research institutions commission systematic, India-specific survey research on public trust in AI-enabled administration, to close the evidentiary gap identified in Section 6 and allow future interventions to be targeted with greater precision.

Implementing these recommendations will require coordination across multiple institutional actors, and it is useful to be explicit about where responsibility for each might plausibly sit within India's existing administrative architecture. Disclosure standards and fairness-audit requirements could be operationalised through amendments to sectoral guidelines issued by the relevant nodal ministries, in coordination with the Ministry of Electronics and Information Technology, which already plays a central role in India's digital governance infrastructure. A dedicated accountability mechanism, whether a cross-sectoral regulator or a clearly designated function within an existing body, would likely require primary legislation, given the civil society critique that voluntary principles alone have proven insufficient to establish enforceable obligations (Abraham et al., 2018). Data protection reform is, at the time of writing, already under active legislative consideration, and the recommendations offered here are intended to reinforce, rather than duplicate, that broader effort by highlighting the specific risks associated with cross-linkage of welfare and surveillance-adjacent databases. Redressal infrastructure, finally, is likely to require investment at the level of individual implementing departments and state governments, given that the practical accessibility of grievance mechanisms depends heavily on local administrative capacity and last-

mile service delivery – domains in which state and district administrations, rather than central ministries, typically hold primary operational responsibility.

None of these recommendations is intended to be read as a call to halt or reverse India's investment in algorithmic governance. Rather, they are offered on the premise, argued throughout this paper, that the long-term success of that investment – measured not only in efficiency gains but in sustained public confidence – depends on treating trust as a design requirement rather than a downstream public-relations concern.

8. Conclusion

This paper has argued that the legitimacy of AI-enabled administrative decision-making in India depends on more than the technical accuracy of the systems involved; it depends on whether citizens can trust that such systems are transparent, accountable, fair, protective of their data, and subject to accessible redressal. Using a five-dimensional conceptual framework and illustrative evidence from Aadhaar-linked welfare authentication, CCTNS, FASTag, and SUPACE, the paper has shown that India's current AI governance architecture – anchored in non-binding strategic guidance – leaves each of these dimensions only partially addressed, with the accountability and redressal gaps particularly pronounced for the welfare and law-enforcement contexts in which marginalised citizens most frequently encounter algorithmic decision-making.

The analysis further suggests that trust deficits are not evenly distributed across the algorithmic state: they concentrate where the stakes of a decision are highest and where meaningful human oversight has been most substantially displaced, and they fall disproportionately on citizens with the least capacity to contest an erroneous outcome. This distributive pattern gives the trust-deficit problem a constitutional as well as an administrative dimension, implicating guarantees of equality and non-arbitrariness that lie at the heart of India's administrative law tradition (CLPR, 2022).

As India continues to expand the scope of algorithmic governance across welfare, policing, taxation, and judicial administration, closing this trust deficit will require binding accountability mechanisms, systematic fairness auditing, comprehensive data protection, accessible redressal infrastructure, and – as a precondition for evidence-based policymaking – sustained, India-specific research into how citizens actually experience and evaluate the algorithmic state. The conceptual framework developed in this paper is offered as one contribution toward that research agenda, and toward a mode of digital governance in which efficiency and public trust are pursued as complementary, rather than competing, objectives.

REFERENCES:

1. Abraham, S., Hickok, E., Sinha, A., Barooah, S., Mohandas, S., Bidare, P. M., Dasgupta, S., Ramachandran, V., & Kumar, S. (2018). NITI Aayog discussion paper: An aspirational step towards India's AI policy. Centre for Internet and Society.
2. Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. European Law Journal, 13(4), 447–468.
3. Centre for Law and Policy Research. (2022). Automated administration: Administrative law and algorithmic decision-making in India. In The philosophy and law of information regulation in India. CLPR.
4. Chatterjee, S. (2020). AI strategy of India: Policy framework, adoption challenges and actions for government. Transforming Government: People, Process and Policy, 14(5), 757–775.

5. Citron, D. K. (2008). Technological due process. Washington University Law Review, 85(6), 1249–1313.
6. Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89(1), 1–33.
7. Diakopoulos, N. (2016). Accountability in algorithmic decision making. Communications of the ACM, 59(2), 56–62.
8. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994.
9. Edelman. (2022). 2022 Edelman Trust Barometer. Edelman.
10. Engin, Z., & Treleaven, P. (2019). Algorithmic government: Automating public services and supporting civil servants in using data science technologies. The Computer Journal, 62(3), 448–460.
11. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
12. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.
13. Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705.
14. NITI Aayog. (2018). National strategy for artificial intelligence: #AIforAll. Government of India.
15. NITI Aayog. (2021). Responsible AI: Approach document for India, part 1 – principles for responsible AI. Government of India.
16. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown Publishing.
17. Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.
18. Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. Fordham Law Review, 87(3), 1085–1139.
19. Shah, H. (2018). Algorithmic accountability. Philosophical Transactions of the Royal Society A, 376(2128), 20170362.
20. Smith, A. (2017). Automation in everyday life. Pew Research Center.
21. Yeung, K. (2018). Algorithmic regulation: A critical interrogation. Regulation & Governance, 12(4), 505–523.
22. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.