International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

The Legitimacy Paradox: How AI Surveillance Reshapes Citizen Trust in Indian Public Administration

Lt. Dr. Kongala Sukumar

Associate Professor of Public Administration
Dr. Marri Chenna Reddy, Human Resource Development (MCHRD) Institute of Telangana, Road no 25, Jubilee Hills, Hyderabad, Telangana 500033

Abstract:
As Indian public administration increasingly deploys artificial intelligence for surveillance, service delivery, and predictive governance, a paradox emerges: technologies introduced to enhance state legitimacy through efficiency and accountability simultaneously erode the very trust they are meant to build. This paper examines what is termed here the "legitimacy paradox" — the tension between the instrumental legitimacy that AI surveillance systems generate through improved service outcomes, and the normative legitimacy they undermine through opacity, function creep, and diminished citizen agency. Drawing on legitimacy theory, algorithmic governance scholarship, and India's evolving regulatory landscape under the Digital Personal Data Protection Act, 2023, the paper analyzes facial recognition deployment, predictive policing initiatives such as CCTNS 2.0, and biometric identity infrastructure including DigiYatra and Aadhaar-linked authentication. It argues that legitimacy in AI-mediated governance is not a fixed property but a continuously negotiated relationship shaped by transparency, procedural fairness, and redress mechanisms. The paper develops a three-dimensional framework — input, throughput, and output legitimacy — to analyze how AI surveillance systems in India succeed at output-oriented efficiency while failing at throughput-oriented accountability. It concludes with institutional recommendations for calibrating AI governance so that it sustains, rather than erodes, citizen trust in an expanding surveillance state.

Keywords: AI surveillance; algorithmic legitimacy; citizen trust; public administration; India; Digital Personal Data Protection Act; predictive policing.

1. INTRODUCTION

The Indian state has, over the past decade, become one of the world's most ambitious adopters of artificial intelligence in public administration. Facial recognition systems assist the Delhi and Hyderabad police in tracking suspects and missing persons; the proposed CCTNS 2.0 architecture aims to layer predictive analytics onto crime records; DigiYatra uses biometric facial matching to move travelers through airports; Aadhaar-linked authentication now underwrites everything from welfare disbursal to tax administration; and the judiciary's SUPACE platform assists judges with case-law retrieval. Each of these systems is justified in the language of legitimacy: efficiency, fairness, transparency, and improved service delivery

to citizens who have long experienced the Indian bureaucracy as slow, discretionary, and difficult to hold accountable.

Yet the same systems that promise to rebuild trust in the state through measurable performance frequently do so by expanding what the state can see, store, and infer about its citizens, often without a clear statutory basis, meaningful consent, or accessible mechanisms of redress. This paper calls this tension the "legitimacy paradox": AI surveillance systems are simultaneously legitimacy-generating, because they demonstrate state capacity and responsiveness, and legitimacy-eroding, because they concentrate discretionary power in opaque technical systems that citizens cannot see, question, or contest.

The paradox is not unique to India, but it takes a distinctive shape here. India's AI-enabled governance apparatus has expanded faster than its data protection architecture: The Digital Personal Data Protection Act, 2023 (DPDP Act) was passed in August 2023, yet its operative rules were not notified until November 2025, leaving a multi-year gap in which biometric and surveillance systems scaled without a functioning statutory oversight body. At the same time, India lacks a dedicated law governing law-enforcement use of predictive or facial-recognition technology, meaning that many of the systems discussed in this paper operate on administrative authority rather than legislative mandate.

This paper asks two connected questions. First, through what mechanisms does AI surveillance simultaneously build and erode legitimacy in Indian public administration? Second, what institutional design choices could reduce the paradox — that is, preserve the efficiency gains of AI-enabled governance while restoring the procedural and normative legitimacy that opaque, centralized surveillance tends to undermine?

The stakes of the paradox are not merely academic. Public administration scholarship has long treated citizen trust as the resource that allows the state to govern with the consent, rather than merely the compliance, of the governed; when that trust erodes, compliance costs rise, enforcement becomes more adversarial, and citizens who already distrust the state — often those with prior experience of discretionary policing or exclusionary welfare administration — are pushed further from institutions meant to serve them. AI surveillance sits at the center of this dynamic in India precisely because it is being deployed fastest in exactly the domains — policing, border security, welfare authentication — where the costs of eroded trust fall most heavily on citizens with the least capacity to seek redress.

To answer these questions, the paper proceeds in eight further sections. Section 2 reviews the literature on legitimacy theory and algorithmic governance, situating the Indian case within a broader comparative debate. Section 3 develops the paper's theoretical framework, adapting the input–throughput–output model of legitimacy to the specific dynamics of AI surveillance. Section 4 maps the architecture of AI surveillance currently operating across Indian public administration. Section 5 sets out the paper's methodology. Section 6 applies the framework analytically, using two comparative tables to assess where legitimacy is gained and where it is lost. Section 7 discusses the findings, Section 8 offers policy recommendations, and Section 9 concludes.

2. LITERATURE REVIEW

2.1 Legitimacy in Public Administration Theory

Legitimacy has long been understood in public administration and organizational theory as a socially constructed perception rather than an objective property of an institution. Suchman (1995) defines legitimacy as a generalized perception that the actions of an entity are desirable, proper, or appropriate within a socially constructed system of norms, values, and beliefs — a definition that has become

foundational across subsequent work on algorithmic governance. Classical accounts of political legitimacy, from Weberian rational-legal authority to Easton's systemic model of political support, treat legitimacy as something institutions accumulate through consistent, rule-bound behavior over time. More recent scholarship has translated these classical accounts into the operational logic of digital and algorithmic government, arguing that legitimacy must now be understood as something continuously produced — or eroded — through the design of computational systems themselves (Iwanowska, 2025).

A particularly useful analytical move, developed in the algorithmic governance literature, disaggregates legitimacy into three components: input legitimacy (whether a governing arrangement aligns with citizen interests and has a proper mandate), throughput legitimacy (whether the process by which decisions are made is open, accountable, and trustworthy), and output legitimacy (whether the arrangement is effective) (Grimmelikhuijsen & Meijer, 2022). This tripartite model is especially well suited to AI systems, because it allows a single technology to be simultaneously strong on one dimension of legitimacy and weak on another — precisely the paradox this paper investigates.

It is worth noting that legitimacy theory has historically treated institutions, not technologies, as the unit of analysis. Suchman's (1995) framework was developed to explain why organizations seek legitimacy and how they manage it strategically — through conformity to institutional norms, selective disclosure, and alignment with dominant cultural accounts of appropriate behavior. Translating this to an algorithmic system requires an additional step: an algorithm does not seek legitimacy for itself, but it is embedded within, and lends its outputs to, institutions that do. This is the basis for what Section 3 below calls "legitimacy borrowing," and it means that the legitimacy of an AI surveillance system cannot be assessed independently of the institution that deploys it, even though the system's own design — its transparency, contestability, and accuracy — has an independent causal effect on how that borrowed legitimacy is sustained or depleted over time.

2.2 Algorithmic Governance and Trust

A growing body of empirical work has examined how citizens and public administrators actually perceive the legitimacy of automated decision-making. Grimmelikhuijsen and Meijer (2014) found experimentally that transparency measurably increases the perceived trustworthiness of a government organization, a finding extended by Grimmelikhuijsen and Knies (2017), who developed and validated a multi-item scale for citizen trust in government organizations. De Fine Licht and De Fine Licht (2020) argue that explanations are the key mechanism through which artificial intelligence produces perceived legitimacy in public decision-making — without an explanation citizens can understand, even a technically accurate algorithmic decision is likely to be experienced as arbitrary. In one of the more recent experimental contributions, Hillo, Vento, and Erkkilä (2025) ran parallel survey experiments with citizens and public administrators and found that both algorithmic transparency and the presence of a "human in the loop" significantly shape legitimacy perceptions, though public administrators respond more strongly to the disclosure of algorithmic processes than citizens do.

Parallel work in AI ethics has focused on the conditions under which algorithmic systems can be considered accountable in the first place. Jobin, Ienca, and Vayena (2019) surveyed the global landscape of AI ethics guidelines and found transparency, fairness, and accountability to be near-universal principles, but noted the near-total absence of formal mechanisms to enforce or evaluate compliance with them. Floridi et al. (2018) proposed a broader ethical framework for AI in society built around explainability and institutional oversight. On the technical accountability side, Raji et al. (2020) introduced a framework for internal algorithmic auditing, arguing that trust in AI systems is contingent on institutionalized audit

processes capable of detecting bias before it causes public harm, while Selbst et al. (2019) demonstrated that algorithmic fairness cannot be reduced to technical metrics because it is embedded in sociotechnical systems shaped by institutional context and power. Eubanks's (2018) account of automated eligibility systems in the United States remains an influential warning that automation, far from neutralizing bureaucratic discretion, frequently relocates and hardens it against poor and marginalized citizens — a dynamic with clear relevance to predictive policing and welfare-linked biometric systems in India.

Taken together, this literature suggests two propositions that structure the rest of this paper. The first is that legitimacy perceptions of AI governance are multidimensional and can move independently of one another: a system can become more accurate (output) while becoming no more explainable (throughput), and the two trends will be experienced by citizens as contradictory rather than complementary. The second is that the mechanisms known to build throughput legitimacy — published audit results, accessible explanations, a visible human decision-maker of last resort — are largely institutional and procedural commitments rather than technical ones. This matters for the Indian case, because it means the throughput deficit identified in Sections 4 and 6 below is not primarily a problem of insufficiently advanced technology; it is a problem of insufficiently developed institutional practice around technology that is, in narrow technical terms, already quite capable.

2.3 The Indian Context

India's AI-enabled administrative apparatus has scaled without a settled legal architecture. The DPDP Act, 2023 was passed by Parliament in August 2023 and received presidential assent that same month, but its substantive provisions — including the constitution of the Data Protection Board — were only notified in stages beginning in November 2025, with further provisions scheduled through 2027 (Wikipedia, 2023/2026; DD News, 2026). During this extended interim period, facial recognition systems, predictive policing pilots, and biometric travel infrastructure such as DigiYatra continued to expand under executive and administrative authority rather than a dedicated statutory framework (SFLC.in / IFEX, 2026). Civil-society researchers have described this trajectory as a slide toward a "surveillance democracy," in which state, private, and even peer-to-peer monitoring combine without external oversight, disproportionately exposing already-marginalized citizens while surveillants themselves remain largely opaque (GIGA Focus, cited in Digital Surveillance and the Threat to Civil Liberties in India).

The clearest illustration is predictive policing. Reporting on the proposed CCTNS 2.0 system shows that India's paramilitary and police agencies — including the Border Security Force and Central Industrial Security Force — have proposed AI-driven surveillance, behavior analytics, and predictive hotspot identification built on legacy police databases (Medianama, 2026). Because these systems are trained on historically skewed enforcement data, researchers have repeatedly cautioned that they risk amplifying rather than correcting existing patterns of bias in Indian policing (Medianama, 2026). At the same time, government-facing commentary continues to frame AI adoption in overwhelmingly positive terms — as a tool for reducing case backlogs in the judiciary, improving tax fraud detection, and extending administrative reach into rural India through initiatives such as PMGDISHA (UPPCS Magazine, 2026). The coexistence of these two narratives — official optimism about efficiency and civil-society alarm about unchecked reach — is itself evidence of the legitimacy paradox this paper sets out to formalize.

3. THEORETICAL FRAMEWORK: THE LEGITIMACY PARADOX

This paper builds its analytical framework by combining Suchman's (1995) socially constructed definition of legitimacy with the input–throughput–output model developed in the algorithmic governance literature

(Grimmelikhuijsen & Meijer, 2022). The central claim is that AI surveillance systems in Indian public administration are rarely uniformly legitimate or illegitimate. Instead, they tend to score highly on output legitimacy — measurable improvements in detection rates, processing times, or service reach — while scoring poorly on throughput legitimacy, because the internal logic of the algorithm, the data it draws on, and the criteria by which it flags a citizen for attention are rarely disclosed, contestable, or subject to independent audit.

Input legitimacy asks whether a system has a proper mandate: was it authorized by a body accountable to citizens, and does it align with citizen interests as citizens themselves would define them? Several of the systems examined in this paper — facial recognition deployment by state police forces, for instance — were introduced through administrative circulars and procurement decisions rather than dedicated primary legislation, which weakens input legitimacy even where the stated purpose (identifying missing persons, deterring crime) commands broad public sympathy.

Throughput legitimacy asks whether the process of decision-making is open, contestable, and trustworthy. This is where AI surveillance systems are most consistently weak. Facial recognition match thresholds, predictive policing risk scores, and fraud-detection flags are, by design, not published in a form ordinary citizen or even most administrators can interrogate. Citizens who are misidentified, flagged, or denied a service on the basis of an algorithmic output typically have no accessible channel through which to learn why, let alone contest the underlying model.

Output legitimacy asks whether the arrangement works — whether it delivers the efficiency, safety, or convenience it promises. It is here that AI surveillance systems tend to perform best, and this is precisely what makes the paradox durable rather than self-correcting: each incremental gain in measurable output (faster identity verification at airports, quicker case disposal, more responsive tax administration) is used to justify further expansion of systems that remain weak on the throughput dimension. The paradox, in short, is that success on one axis of legitimacy is repeatedly used to paper over failure on another, and citizens experience this not as a coherent whole but as a state that is simultaneously more responsive and less accountable.

A second, related dynamic concerns what this paper terms "legitimacy borrowing." Because AI systems are frequently introduced as extensions of already-legitimate institutions — the police, the tax authority, the airport authority — they inherit a baseline of institutional trust that the technology itself has not independently earned. This borrowed legitimacy can dissolve quickly and asymmetrically: a single high-profile misidentification or data breach can damage trust in the technology far more than incremental improvements in accuracy can rebuild it, because the throughput failures (opacity, lack of redress) remain structurally unresolved even after any single incident is addressed.

A third dynamic worth naming explicitly is scope creep, which the surveillance-studies literature treats as a near-universal feature of administrative surveillance systems once they are established. A system authorized for a narrow purpose — locating missing persons, verifying identity for a specific welfare scheme, screening at a single airport — tends over time to be extended to adjacent purposes for which it was never separately authorized or publicly justified. Because input legitimacy is assessed at the moment of authorization, scope creep allows a system's operating legitimacy to drift away from its original mandate without ever triggering a fresh legitimacy assessment. This is one reason the throughput dimension — ongoing transparency and contestability, rather than a one-time authorization — is arguably the more consequential axis for long-run trust, even though input legitimacy dominates public debate at the point a system is first announced.

4. THE ARCHITECTURE OF AI SURVEILLANCE IN INDIAN PUBLIC ADMINISTRATION

AI-enabled surveillance and administrative decision-making in India is not a single system but a distributed architecture spanning law enforcement, immigration and travel, welfare and identity, and revenue administration. Table 1 summarizes the principal systems discussed in this paper, their administering authority, and their stated administrative purpose.

Table 1- *Selected AI-Enabled Surveillance and Administrative Systems in India*

System	Administering Authority	Stated Purpose	Legal Basis
Facial Recognition Technology (FRT)	Delhi Police, Hyderabad Police, and other state police forces	Identification of suspects and missing persons	Administrative/executive order; no dedicated statute
CCTNS 2.0 predictive analytics	National Crime Records Bureau (NCRB), state police	Crime-pattern prediction and risk scoring	Administrative; proposed expansion of CCTNS
Border and perimeter AI surveillance	Border Security Force (BSF), Central Industrial Security Force (CISF)	Detection of unauthorized crossings; CCTV analytics at critical infrastructure	Administrative/operational directives
DigiYatra	Airports Authority of India / Ministry of Civil Aviation	Biometric facial-match boarding at airports	Voluntary programme; administrative framework
Aadhaar-linked authentication	Unique Identification Authority of India (UIDAI)	Identity verification for welfare, banking, and tax services	Aadhaar Act, 2016, as amended
SUPACE	Supreme Court of India	AI-assisted legal research and case-material assistance for judges	Judicial administrative initiative
AI-based tax and financial fraud detection	Ministry of Finance / Income Tax Department	Detection of tax evasion and financial anomalies	Administrative; Income Tax Act provisions

Note. Compiled by the author from Medianama (2026), SFLC.in/IFEX (2026), and UPPCS Magazine (2026). This table describes publicly reported system characteristics rather than internal government specifications, which are not publicly disclosed.

What Table 1 makes visible is a pattern noted across the surveillance-studies literature on India: nearly every system listed operates on administrative or executive authority, procurement decisions, or a general-purpose statute (the Aadhaar Act) not written with AI-specific oversight in mind, rather than a dedicated legislative framework governing the specific technology in use (SFLC.in / IFEX, 2026; Medianama, 2026). This is the structural root of the throughput-legitimacy weakness identified in Section 3: without a statute that specifies audit rights, accuracy thresholds, or redress mechanisms, citizens have no fixed point of reference against which to hold any individual system accountable.

A second pattern visible in Table 1 is the diversity of administering authorities involved — state police forces, a national judicial body, a paramilitary force, an airport regulator, and a national identity authority — with no single coordinating institution responsible for AI governance across this architecture as a whole. Each authority sets its own procurement standards, accuracy thresholds, and (where they exist at all) grievance procedures, which means that a citizen's experience of AI-mediated governance can vary substantially depending on which agency's system they happen to encounter. This institutional fragmentation compounds the throughput deficit identified above: even a citizen motivated to seek an explanation or file a grievance faces a different, largely undocumented process depending on whether the system in question is police-operated, airport-operated, or identity-authority-operated, with no common standard of disclosure across them.

5. METHODOLOGY

This paper adopts a qualitative, document-based analytical approach rather than an original citizen survey. It combines (a) doctrinal analysis of India's principal data-governance statute, the DPDP Act, 2023, and related administrative frameworks; (b) a structured review of secondary and policy literature on AI surveillance systems operating in Indian public administration, drawn from government communications, civil-society research organizations, and investigative policy journalism; and (c) theory-driven analytical synthesis, in which the input–throughput–output legitimacy framework developed in Section 3 is applied systematically to the systems mapped in Section 4.

This design was chosen deliberately. Primary survey data on Indian citizens' trust in specific AI surveillance systems is not yet available at a scale that would support reliable generalization, and existing comparative survey evidence on algorithmic legitimacy comes overwhelmingly from European and East Asian contexts (Grimmelikhuijsen & Meijer, 2014; Hillo et al., 2025) whose institutional starting conditions differ substantially from India's. Rather than extrapolate from those contexts or present illustrative numbers as if they were empirical findings, this paper restricts its analysis to what can be documented from public sources and uses the resulting analytical tables (Tables 2 and 3) as structured comparative frameworks rather than as reports of primary data collection. This is a limitation the paper returns to in Section 7, and it defines a clear agenda for the primary survey research that would be needed to test the framework's propositions empirically.

6. ANALYSIS: MAPPING THE PARADOX

Applying the input–throughput–output framework to the systems described in Section 4 produces a consistent pattern: strong output performance, moderate-to-weak input legitimacy, and consistently weak throughput legitimacy. Table 2 presents this analysis for four representative systems.

Table 2- *Legitimacy Assessment of Selected AI Surveillance Systems*

System	Input Legitimacy	Throughput Legitimacy	Output Legitimacy	Net Pattern
Facial Recognition (policing)	Weak — no dedicated statute; introduced administratively	Weak — match criteria and error rates not published	Strong — reported utility in locating missing persons	Output-driven legitimacy; throughput deficit
CCTNS 2.0 predictive policing	Weak — proposed by security agencies with limited public consultation	Weak — training data and risk-scoring logic undisclosed	Uncertain/contested — efficacy claims not independently audited	Legitimacy claim rests on anticipated, unverified output
DigiYatra	Moderate — framed as voluntary, opt-in	Moderate — process is visible to the user at the point of use	Strong — measurable reduction in airport processing time	Comparatively balanced; strongest of the systems reviewed
Aadhaar-linked authentication	Moderate — statutory basis under the Aadhaar Act	Weak — limited recourse for authentication failures affecting welfare access	Strong — high enrolment and transaction volumes	Output-driven legitimacy; redress gap for authentication failures

Note. Author's analytical synthesis based on the sources reviewed in Sections 2 and 4

(Medianama, 2026; SFLC.in/IFEX, 2026; DD News, 2026). Ratings are qualitative assessments derived from the documented design and legal basis of each system, not survey measurements of citizen opinion. Table 2 shows why DigiYatra — a voluntary, opt-in, single-purpose system with a visible user-facing process — sits closer to balanced legitimacy than facial recognition policing or predictive policing, which are involuntary, opaque, and layered onto historically contested law-enforcement data. This comparison suggests that voluntariness and visibility of process, not accuracy alone, are strong predictors of where a system falls on the throughput-legitimacy axis.

Table 3 extends the comparison outward, situating India's regulatory posture on AI surveillance against three other major jurisdictions, to clarify what a stronger throughput-legitimacy regime would need to contain.

Table 3- *Comparative Regulatory Oversight of AI Surveillance Systems*

Jurisdiction	Core Instrument	Independent Oversight Body	Right to Explanation / Redress	Status (as of paper's analysis)
India	Digital Personal Data Protection Act, 2023	Data Protection Board of India	Limited; grievance redressal provisions exist but no AI-specific explanation right	Act passed August 2023; rules and Board operationalized in stages from November 2025
European Union	EU Artificial Intelligence Act	National market-surveillance authorities plus EU AI Office	Explicit risk-tiering; transparency and human-oversight obligations for high-risk systems including biometric surveillance	In force, phased implementation
United States	Sector-specific rules; no comprehensive federal AI statute	Sectoral regulators (e.g., FTC) and state-level initiatives	Uneven; varies by state and sector	Fragmented, evolving
United Kingdom	Data protection law (UK GDPR) plus sector guidance; no dedicated AI statute	Information Commissioner's Office	Data-protection-based rights; no AI-specific statute	Principles-based, non-statutory AI framework

Note. Compiled by the author from PRS Legislative Research (2023), the Digital Personal Data Protection Act, 2023 (Government of India), and DD News (2026), supplemented by publicly available comparative policy summaries.

The comparison in Table 3 underscores that India's DPDP Act, while a significant milestone, was designed as a general data-protection statute rather than an AI- or surveillance-specific instrument, in contrast to the EU's risk-tiered approach, which explicitly designates biometric surveillance as high-risk and attaches transparency and human-oversight obligations to it. This gap helps explain why systems such as facial recognition policing and predictive policing in India continue to operate largely outside dedicated statutory oversight even after the DPDP Act's passage.

7. DISCUSSION

Three findings emerge from the analysis above. First, the legitimacy paradox is not evenly distributed across the Indian AI-surveillance landscape: voluntary, single-purpose, user-visible systems such as DigiYatra generate comparatively balanced legitimacy, while involuntary, multi-purpose, opaque systems such as predictive policing generate the sharpest paradox, because their output claims are largely unverifiable by outside observers even as their throughput weaknesses are most consequential for the citizens they target. Second, the DPDP Act's long implementation runway — passed in 2023 but only substantially operationalized from late 2025 — meant that the systems with the weakest throughput legitimacy expanded during precisely the period when statutory oversight was least available, a sequencing problem as much as a design problem. Third, the borrowed-legitimacy dynamic described in Section 3 helps explain why public commentary on AI in Indian governance can be simultaneously celebratory (efficiency gains, reduced backlogs, expanded rural reach) and alarmed (surveillance democracy, unaudited bias, absent redress): both framings are accurate descriptions of different legitimacy dimensions of the same systems.

This analysis has clear limitations. Because it relies on secondary and policy-level sources rather than an original citizen survey, it cannot report the actual magnitude of trust gained or lost among specific citizen populations, nor can it disaggregate legitimacy perceptions by geography, class, caste, or prior experience of state surveillance — all of which the literature on surveillance and marginalization suggests would shape how the paradox is experienced (Eubanks, 2018). The analytical ratings in Tables 2 and 3 should therefore be read as a structured starting framework for empirical research, not as a substitute for it. A natural next step is a survey-experimental design, similar to that used by Hillo, Vento, and Erkkilä (2025) in the European context, adapted to Indian administrative systems and citizen populations, to test whether the input–throughput–output pattern identified here through document analysis holds when measured directly against citizen perception.

A further implication concerns sequencing. Comparative experience, particularly the European Union's decision to build risk-tiering directly into its AI statute rather than relying on a general data-protection instrument, suggests that throughput-legitimacy safeguards are far easier to design into a system before deployment than to retrofit afterward. India's DPDP Act was drafted primarily to address commercial and everyday data processing; extending its logic to cover the specific harms of biometric and predictive-policing surveillance — misidentification, chilling effects on lawful assembly and movement, and the absence of an appeals process for algorithmically flagged citizens — was not its central design problem, and the analysis above suggests this gap has practical consequences rather than being merely a matter of legislative drafting style. The systems that this paper found weakest on throughput legitimacy are also the ones for which a dedicated, purpose-built oversight regime is least developed, which is consistent with the view that legitimacy deficits in AI governance are, at root, institutional-design deficits rather than purely technical ones.

8. POLICY RECOMMENDATIONS

Four recommendations follow from the analysis. First, throughput legitimacy should be treated as a design requirement, not an afterthought: any AI surveillance system deployed in a public-administration context should be required, before deployment, to publish a plain-language description of what data it uses, what threshold triggers a flag or match, and what recourse a citizen has if flagged in error. Second, the sequencing gap between technology deployment and statutory oversight should be closed by extending the Data Protection Board's mandate, or creating a dedicated AI oversight function within it, to cover systems already in operation rather than only those deployed after the Board's full operationalization. Third, India's regulatory architecture would benefit from an explicit risk-tiering mechanism broadly comparable to the EU AI Act's approach, under which biometric and predictive-policing systems are treated as high-risk and subject to mandatory human-oversight and audit requirements, rather than being governed by the same general provisions that apply to routine commercial data processing. Fourth, voluntary, opt-in design — the feature that appears to explain DigiYatra's comparatively balanced legitimacy profile in this analysis — should be adopted as a default principle wherever a system's core administrative purpose does not require universal, involuntary participation.

A fifth, cross-cutting recommendation concerns measurement. If legitimacy is genuinely multidimensional, as this paper has argued, then government evaluation of AI surveillance systems should stop relying solely on output metrics — arrest rates, processing-time reductions, enrolment numbers — as evidence of a system's success. Periodic, independently conducted legitimacy audits, assessing input, throughput, and output dimensions separately and reporting on each publicly, would allow the throughput deficits identified in this paper to be tracked over time rather than surfacing only when an individual incident of misidentification or data misuse becomes public. Such audits need not be resource-intensive: publishing accuracy and false-positive rates disaggregated by demographic group, documenting the volume and outcome of citizen grievances filed against a given system, and disclosing the categories of data a system draws on would each, on their own, meaningfully close part of the throughput gap identified in Table 2.

9. CONCLUSION

AI surveillance in Indian public administration presents a genuine paradox rather than a simple trade-off: the same systems that most visibly demonstrate state capacity and responsiveness are frequently the ones that most significantly weaken the procedural conditions — transparency, contestability, redress — on which durable citizen trust depends. This paper has argued that resolving the paradox requires disaggregating legitimacy into its input, throughput, and output components, because a system can succeed dramatically on one dimension while failing on another, and current public debate in India tends to evaluate AI governance along only one of these dimensions at a time. As the Data Protection Board of India becomes fully operational and as predictive and biometric systems continue to scale, the central institutional challenge is not whether to use AI in public administration, but whether throughput legitimacy — the transparency, auditability, and redress mechanisms that make a system's internal logic answerable to the citizens it governs — can be built into these systems as deliberately as their output performance already has been.

REFERENCES:

1. De Fine Licht, K., & De Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & Society, 35*(4), 917–926. https://doi.org/10.1007/s00146-020-00960-w
2. Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
3. 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.
4. Grimmelikhuijsen, S., & Knies, E. (2017). Validating a scale for citizen trust in government organizations. *International Review of Administrative Sciences, 83*(3), 583–601.
5. 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
6. Grimmelikhuijsen, S., & Meijer, A. (2022). Legitimacy of algorithmic decision-making: Six threats and the need for a calibrated institutional response. *Perspectives on Public Management and Governance, 5*(3), 232–242.
7. Hillo, J., Vento, I., & Erkkilä, T. (2025). Algorithmic governance: Experimental evidence on citizens' and public administrators' legitimacy perceptions of automated decision-making. *Public Administration*. Advance online publication. https://doi.org/10.1111/padm.70028
8. Iwanowska, B. (2025). Algorithmic legitimacy and the digital state: Rethinking governance, trust, and accountability in the age of AI. *Journal of Law, Politics and Administration, 1*(2), 130–149. https://doi.org/10.5709/alp-01.02.2025-03
9. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.
10. Medianama. (2026, March/April). *India explores AI policing tools under CCTNS 2.0*. https://www.medianama.com/2026/03/223-government-developing-ai-system-predictive-policing-risk-scoring-entities/
11. Observer Research Foundation. (2023). *The Digital Personal Data Protection Act, 2023: Recommendations for inclusion in the Digital India Act*. https://www.orfonline.org/research/the-digital-personal-data-protection-act-2023-recommendations-for-inclusion-in-the-digital-india-act
12. PRS Legislative Research. (2023). *The Digital Personal Data Protection Bill, 2023*. https://prsindia.org/billtrack/digital-personal-data-protection-bill-2023
13. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery.
14. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery.
15. SFLC.in. (2026). *Artificial intelligence and surveillance in India: 2025 roundup*. https://sflc.in/artificial-intelligence-and-surveillance-in-india-2025-roundup/

16. Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, *20*(3), 571–610.
17. UPPCS Magazine. (2026). *Reimagining governance: The role of artificial intelligence in transforming public administration in India*. https://uppcsmagazine.com/reimagining-governance-the-role-of-artificial-intelligence-in-transforming-public-administration-in-india/