

The Paradox of Digital Governance: Ethical Implications of AI-Driven Surveillance for Sustainable Public Administration in India

Lt. Dr. Kongala Sukumar

Associate Professor of Public Administration

Dr. MCRHRD Institute of Telangana,

ORCID: 0009-0008-4166-8847

ARTICLEINFO

Received: 03 May 2024

Revised: 17 June 2024

Accepted: 28 June 2024

ABSTRACT

The rapid adoption of Artificial Intelligence (AI) tools in public sector governance has created a unique governance paradox in India; the use of digital tools is meant to bring transparency, efficiency, and benefits to citizens, while at the same time, the tools could lead to a pervasive state surveillance of the citizens, threatening the democratic values that they are supposed to serve. In this paper, the moral aspects of AI surveillance tools utilized in the Indian public sector will be critically analyzed, focusing on facial recognition systems, biometric databases, predictive policing algorithms, and smart city command centres. The study, which uses a multi-method analytical framework that combines normative ethics, sustainable governance theory, and empirical data from the National Crime Records Bureau, NITI Aayog, and independent civil society research, examines three crucial tensions: the efficiency-privacy dichotomy inherent in the use of biometrics for welfare delivery; the accountability shortfall at the heart of the use of algorithms in law enforcement; and the disproportionate impact of these algorithms on marginalised communities. The paper proposes to reshape the governance-AI nexus within the context of sustainable public administration in India by implementing comprehensive legislative protection frameworks, requiring algorithm assessments, and establishing collaborative design processes. Without proper control, AI surveillance has the potential to continue reinforcing systemic inequities and erode trust in institutions. The study suggests a blueprint for normative ethical governance of AI in India focusing on constitutional protection, federalism, and sustainable development.

Keywords: artificial intelligence; digital governance; surveillance ethics; public administration; India; algorithmic accountability; biometric systems; sustainable governance; privacy rights; facial recognition

1. INTRODUCTION

The rapid diffusion of artificial intelligence (AI) technologies across public sector institutions globally has been characterised as the defining administrative transformation of the early twenty-first century (Wirtz et al., 2019). In a few countries, however, this transformation is more consequential—or more contested—than in India, where a democracy of 1.4 billion people, extraordinary socio-economic diversity, and a constitutionally enshrined framework of fundamental rights intersect with an ambitious state-driven project of digital modernisation. The central and state governments have deployed AI systems on a scale unprecedented in the history of India: facial recognition networks in urban population monitoring across dozens of cities, the Aadhaar biometric identity system featuring more than 1.3 billion people, predictive policing, which has been implemented in some states such as Uttar Pradesh and Telangana, and integrated command-and-control centres, which are aggregating real-time surveillance streams in smart cities (Ministry of Electronics and Information Technology [MeitY], 2022; Datta, 2020).

These developments have the real potential to be a positive thing. AI-driven e-governance can help curb corruption, eliminate human discretion in administrative processes, speed up the distribution of social welfare to previously marginalised communities, and enhance public safety by allocating resources based on data (Misra & Mathur, 2021).

At the same time, they produce the paradox of digital governance: the very same set of technology that improves the administrative process also generates new, and extremely powerful, tools of social control, algorithmic discrimination and citizen surveillance, many of which are deployed without proper legal authorization, oversight or democratic discussion (Deibert, 2020; Ramanathan, 2023).

These are accentuated by three structural characteristics of the governance structure of India. First, the lack, until recently, of a comprehensive data protection statute has enabled AI surveillance to proliferate in the absence of a law to guide state actors to respect general constitutional principles that have been piecemeal interpreted over a series of judicial rulings (Srikrishna Committee, 2018; Bhatia, 2021). Second, the federal structure of the Indian government creates situations in which different subnational governments can experiment with different surveillance technologies, which are not within the reach of accountability mechanisms at the national level (Khera, 2019). Third, the socio-demographic characteristics of the populations that are the subjects of increased surveillance mirror and contribute to the reinforcement of pre-existing inequalities that are deeply rooted in the social order (Kovacs & Ranganathan, 2020; Raval & Dourish, 2016).

In response to these concerns, this paper examines three related research questions: (1) What ethical dilemmas arise from the use of AI-powered surveillance in public sector administration in India? (2) What sort of tensions emerge in particular governance fields like welfare provision, law enforcement, and urban management? (3) What institutional and regulatory measures are needed to secure sustainable and rights-fostering public administration in the era of AI surveillance? The paper uses the following structure. The second section presents a theoretical and empirical review of the literature on AI governance, surveillance ethics, and sustainable public administration. In Section 3, we explore the terrain of AI surveillance in the Indian public sector. In section 4, a structured ethical analysis is developed. The data analysis results are presented in Section 5. In Section 6, a normative framework for ethical AI governance is proposed. Implications are discussed in Section 7, and conclusions are made in Section 8.

2. LITERATURE REVIEW

2.1 AI in Public Administration: Promises and Perils

Since 2015, the academic body about AI and public administration has significantly increased, moving from initial hopes of efficiencies to more cautionary perspectives on the dangers of algorithmic governance. The central paradox of digital government identified by Fountain (2001) was that information technologies are institutional and also technical in nature and that their impact on governance is driven by political decisions, not technological determinism. Later researchers have turned this observation to its application in AI. Danaher (2016) claimed that more algorithmic governance creates an 'algocracy' or rule by automated decision systems, which threatens traditional democratic accountability. Cobbe (2019) showed how the opacity of machine learning models can make it harder for citizens to challenge administrative decisions, and Binns (2018) explored the normative assumptions behind algorithmic notions of fairness that can tend to systematically disadvantage marginalised groups.

In the Indian context, there is a growing debate on the tensions between the efficiency goals of digital governance and constitutional principles of rights protection in several studies. In his foundational analysis of the surveillance potential of Aadhaar, Ramanathan (2018) pointed out that the use of the biometric identity system to tie welfare benefits to it is an example of forced digital enrolment, which disproportionately affects the poor. The failure of biometric authentication has resulted in the exclusion of people from welfare services, as reported by Khera (2017) who noted that technical issues in the Aadhaar-based food distribution system have caused hunger for rural communities. Arun (2019) studied the discriminatory use of facial recognition by police forces in India, finding that the algorithms have high error rates for people with darker skin tones, which aligns with similar research on algorithmic bias across the world (Buolamwini & Gebru, 2018).

2.2 Surveillance Ethics and Democratic Governance

Several frameworks in the normative literature can be used for assessing AI-based monitoring systems. Drawing on Mill (1859) and Rawls (1971), classical liberal theories focus on two values: autonomy and privacy, the right to be left alone and the right to enjoy one's own privacy, which are subject to very strict proportionality requirements and which state interference must prove necessary to secure. Theories of Foucauldian discipline locate digital surveillance in the genealogies of its disciplinary power; digital surveillance produces disciplined subjects who are always free,

free to be themselves at the same time as they form themselves in the act of surveillance (Lyon, 2007). Feminist and post-colonial scholars have also questioned the neutrality of surveillance technologies, demonstrating that they are not neutral in themselves, but rather a reflection of gender, race, caste, and class hierarchies (Eubanks, 2018; Browne, 2015).

The general consensus among authors in the context of public administration is that AI surveillance should be both necessary and serve a public interest purpose that cannot be accomplished any other way; it should be proportionate, meaning that it should not be greater than necessary to achieve its specific public purpose; and it should be accountable, with transparency in the decision-making process, independent oversight, and legal limits on use (Taylor, 2017; Yeung, 2018). These principles can be understood in the Indian constitutional context of administrative law, where Article 21 guarantees right to life and personal liberty, and the Supreme Court's judgement in *Puttaswamy v. Union of India* (2017) added the concept of privacy as a fundamental right with the three-part test of legality, necessity, and proportionality.

2.3 Sustainable Public Administration and AI

The idea of sustainable public administration is the synthesis of the three elements of governance sustainability: the environmental, social, and institutional (Jørgensen & Bozeman, 2007). For AI governance, sustainability means that over time, digital administrative systems enhance, not undermine, institutional trust, social cohesion, and legal frameworks that are the bases of legitimate governance (OECD, 2019). Bannister and Connolly (2012) showed that e-government systems that do not gain the trust of their citizens are likely to have a long-term deficit of legitimacy that affects their sustainability. Also recently, Janowski (2015) has opened the discussion on how to incorporate equity concerns into the design of technology, as well as how to ensure that technological efficiencies are not concentrated disproportionately within a single segment of the population, but are shared more broadly.

Sustainable AI governance in India, however, should also address developmental priorities of a nation with around 19 to 22 percent of the population living in multidimensional poverty (UNDP, 2022) and a geography that has unequal access to state capacities for delivering basic services. This, combined with the potential for exclusion or repression via AI use, has been termed a 'development-rights dilemma' in digital governance in the Global South (Arora, 2019; Toyama, 2015).

3. THE LANDSCAPE OF AI DRIVEN SURVEILLANCE IN INDIAN PUBLIC ADMINISTRATION

3.1 Biometric Identity and Welfare Delivery

The Aadhaar system operated by the Unique Identification Authority of India (UIDAI) is the world's largest biometric register, with over 1,360 million people enrolled (as of March 2024, UIDAI, 2024). It can take fingerprints, iris scans, and pictures of faces, and creates a twelve-digit unique identifier that is linked to an increasing number of government services. Aadhaar has unarguably improved the welfare delivery mechanism and government officials claim that it has saved over Rs. With the linking of food ration cards, social pensions and bank accounts, this amount rose to 2.73 lakh crore (roughly USD 33 billion). It is proven that genuine beneficiaries were excluded (NITI Aayog, 2023).

Aadhaar has been widely reported to be used as a surveillance instrument. The system generates authentication logs which give a detailed account of citizens' dealings with the state, together with payment systems and private sector services, with commercial entities (Abraham, 2018). The Supreme Court judgment passed in 2018, in the case of *Puttaswamy v Union of India* (Aadhaar), did not state anything regarding limitation of purpose in the original Aadhaar Act, which allowed the possibility of mission creep; however, the Supreme Court did make a condition that the government must limit the use of Aadhaar by the private sector whilst private entities can use it. Since the enactment of the provisions, however, the extent to which acceptable Aadhaar data was being collected has not been clearly defined due to the successive amendment in the legislation and judgements (Bhatia, 2021).

Table 1. Key Metrics of the Aadhaar Biometric Identity System

Metric	Value	Year	Source
Total Aadhaar Enrolments	1.36 billion	2024	UIDAI
Authentication Transactions (Annual)	12.4 billion	2023	UIDAI Annual Report
Direct Benefit Transfer (DBT) Savings	Rs. 2.73 lakh crore	2014–2023	NITI Aayog
States with Biometric Food Distribution	36 (all states/UTs)	2023	NIC / MeitY
Documented Exclusion Complaints (Annual)	~2.3 million	2022	DARPG Grievance Portal
Authentication Failure Rate (Rural)	5–8%	2021	Khera et al.
Aadhaar-Linked Bank Accounts	820 million+	2023	RBI Annual Report

Note. Data compiled from UIDAI (2024), NITI Aayog (2023), Khera et al. (2021), and Reserve Bank of India (2023).

3.2 Facial Recognition and Law Enforcement

The use of facial recognition technology (FRT) by the Indian police has surged since 2019. The NCRB has developed the National Automated Facial Recognition System (NAFRS) to establish a centralised federal FRT database that can be accessed by the police in every state (NCRB, 2020). The state-level deployments have already been done before the national system: Delhi Police used FRT at the 2020 Republic Day parade and during protests at the Red Fort; Hyderabad City Police in Telangana deployed a system of FRT with an accuracy of 95% in identifying persons of interest (PI, 2023); Tamil Nadu piloted the use of FRT-based verification at the state's welfare distribution centres (PI 2023).

Technical and rights arguments about the foundation for these deployments are deeply problematic. The findings of independent tests on the accuracy of FRT in Indian conditions have been documented to show that women, dark-skinned persons, and those from the lower socioeconomic spectrum have performed poorly compared to lighter-skinned individuals and men, as observed across the globe in algorithms with unrepresentative training data (Buolamwini & Gebru, 2018; Arun, 2019). As of December 2023, the Internet Freedom Foundation (IFF) has recorded 124 separate instances of facial recognition deployments in Indian states, the vast majority of which are not explicitly authorised by any law, have no published accuracy standards, nor have they established any process for challenge in the event of wrongful identification (IFF, 2023).

**Table 2. Select AI-Based Facial Recognition and Surveillance Deployments in India
(as of 2023)**

State / Agency	System	Stated Purpose	Statutory Basis	Independent Accuracy Data
Delhi Police	CCTNS-FRT Integration	Crime Investigation	None (executive order)	Not published
Telangana Police	Hyderabad City AI Platform	Predictive Policing	None	Vendor: 95% (unverified)
NCRB (National)	NAFRS	Multi-purpose ID	Tendered under MHA mandate	Not yet evaluated
Tamil Nadu	Welfare FRT Pilot	Welfare Delivery	None	Not published
UP Police	TrackMe / C3i	Crowd Management	None	Not published
Maharashtra	Smart City CCTV-FRT	Urban Surveillance	Smart Cities Mission	Not published
Andhra Pradesh	CIPA (Predictive)	Crime Prevention	None	Vendor claims only

Note. Compiled from Internet Freedom Foundation (2023), Privacy International (2023), NCRB (2020), and published state government tender documents.

3.3 Predictive Policing & Algorithmic Law Enforcement

Besides facial recognition, some States in India have implemented or piloted predictive policing tools, which are algorithms based on machine learning that forecast criminal activity, identify suspects, or prioritize police efforts in certain areas. The local socio-economic data, crime data and social media data is used to compute risk score in Andhra Pradesh's Crime Investigation and Prevention Application (CIPA) which is based on that area and individual. In UP, the Crime Command Centre (UPCSC) has 60,000 CCTVs, emergency call records and criminal records databases are integrated and are used for real-time threat assessment (Datta, 2020).

Such biases can be perpetuated by and/or exacerbate the racial/ caste inequities in historical crime data. This study by Kovacs and Ranganathan (2020) has demonstrated that predictive policing models, which are trained on historical arrest rates, systematically over-predict the risk of crime in densely populated neighbourhoods of marginalised caste and religious minority groups, creating a vicious cycle that further drives up police attention in these areas and fuels higher arrest rates which further boosts the likelihood of being flagged by algorithms. This trend has been extensively documented in the context of the United States and the United Kingdom (Angwin et al., 2016; Big Brother Watch, 2022) and is particularly salient in India because of the well-documented over-representation of Dalit and Adivasi individuals in the criminal justice system, with the police and state courts being the primary institutions involved in the criminalization of Dalit and Adivasi people (National Crime Records Bureau, 2022).

3.4 Smart Cities and Urban Surveillance infrastructure.

In India, the Smart Cities Mission, launched in 2015 with 100 cities, has accelerated the adoption of integrated surveillance. The Smart Cities Mission has been launched in India in 2015 with 100 cities, and it has accelerated the adoption of integrated surveillance in India. AI enabled video analytics are being increasingly used for detecting

'anomalous' behaviour, estimating crowd density and tracking the movement of individual people across multiple camera feeds; ICCCs are integrating data from traffic cameras, emergency services, environmental sensors and public WiFi networks (MeitY, 2022; Bhan et al., 2020). Among the most advanced ICCC deployments are those in Surat, which connects more than 7,000 CCTV cameras to an AI analytics system; Bhopal, which facilitates the integration of artificial intelligence with cameras; Lucknow, which has cameras in the public good district; and Visakhapatnam, which has cameras covering the public market.

The governance of surveillance is not uniform and insufficiently strong. ICCCs are typically operated by Special Purpose Vehicles (SPVs), which are hybrid entities consisting of municipal corporations and private technology partners, and exist in a limbo zone in the administrative law framework in India. The ICCC systems also observe that information developed by the systems are shared with state police under informal arrangements, not legally authorised, which may result in accountability gaps and targeted surveillance of political activists and journalists (Srinivasan, 2021; Article 19, 2022).

4. ETHICAL ANALYSIS: KEY DILEMMAS OF AI GOVERNANCE

4.1 The Efficiency-Privacy Trade-Off

In India, the moral equation surrounding AI-driven governance is a simple one: between improving welfare and minimizing crime, and protecting the privacy of citizens. This analysis suggests that the Supreme Court in Puttaswamy (2017) has laid down a three-part test for state intrusion in the affairs of citizens: the intrusion must be explicitly authorised by law; it must be necessary for a legitimate state purpose; and it must be the least restrictive means available.

In the context of welfare delivery through Aadhaar, these cons can be viewed as having serious normative drawbacks. While the efficiency argument is substantiated by the facts and figures provided by NITI Aayog (2023)'s significant savings from 'ghost beneficiaries', the proportionality argument is less encouraging. Other verification schemes, such as community-based, one-time passwords, and paper-based fallback methods, can potentially deliver the same anti-fraud benefits with significantly lower privacy costs and without the failures that prevent legitimate beneficiaries from being authenticated. Even after the partial acceptance of Aadhaar by the Supreme Court in 2018, the government has to have manual override mechanisms, which, however, have suffered varying degrees of implementation, especially in rural areas with challenging biometric capture conditions (Khera, 2019).

Table 3. Efficiency-Privacy Trade-Off Assessment Across AI Governance Domains

AI Application	Efficiency Benefit	Privacy Cost	Proportionality Assessment	Alternatives Available
Aadhaar Welfare Delivery	High (Rs. 2.73 lakh cr saved)	High (transactional surveillance)	Contested	Yes (OTP, community ID)
FRT Law Enforcement	Low-Moderate (limited evidence)	Very High (mass surveillance)	Fails test	Yes (targeted investigation)
Predictive Policing	Moderate (hotspot resource allocation)	High (algorithmic profiling)	Fails test	Yes (community policing)
Smart City ICCCs	Moderate (urban management)	High (continuous monitoring)	Contested	Partial

AI Application	Efficiency Benefit	Privacy Cost	Proportionality Assessment	Alternatives Available
e-Courts AI	High (case management)	Low-Moderate (limited scope)	Passes test	No (scale prevents manual)
AI Tax Compliance	High (revenue recovery)	Moderate (financial data)	Mostly passes	Partial

Note. The author's assessment is based on Puttaswamy (2017) proportionality framework, NITI Aayog (2023), IFF (2023), and Khera (2019).

4.2 Algorithmic Accountability Deficits

The second moral conundrum relates to accountability, and it seems that whenever AI decisions in administration have a negative impact, one has to ask the question, "Who was responsible? The principles of natural justice, reason in decision-making, and judicial review are the foundation of the traditional administrative law in India, which was formulated for human decision makers in established bureaucratic hierarchies. They are not well-suited for algorithmic governance, which has opaque ways of making decisions, is spread across human-machine interfaces, and is commonly hidden from public review by claims of commercial confidentiality from private technology vendors (Cobbe, 2019; Datta, 2020).

This is a significant problem in law enforcement AI. If the person is mistakenly identified by the facial recognition system and arrested or harassed by the police, the government ministry can then shift the burden to the SPV or NCRB, the police officer can say that he or she was following the system's output and the supervising officer may not be technically adept to question the system's logic. There is a lack of systems to respond to systemic harms (do not affect any particular individual) and technical complexity of current responses (must be understood through an algorithmic audit to establish causation) (IFF, 2023; Kovacs & Ranganathan, 2020).

4.3 Distributive Justice and Algorithmic Discrimination

The third ethical conflict is distributive justice, where AI surveillance in India disproportionately places a burden on marginalised groups. This dimension is most extensively documented in the context of facial recognition, where algorithmic bias arising from training datasets that under-represent dark-skinned individuals produces systematically higher false-positive rates for populations already subject to disproportionate policing attention (Buolamwini & Gebru, 2018; Arun, 2019). But the distributive justice concern extends more broadly.

Biometric authentication failures that exclude welfare beneficiaries are concentrated among populations with weathered fingerprints—agricultural workers, manual labourers, elderly persons—who are disproportionately drawn from lower castes and tribal communities (Khera, 2019). Surveillance infrastructure is not evenly distributed in the city and is mostly present in commercial and tourist zones, which means that the majority of the urban poor is not as well monitored as special interest groups. Predictive policing algorithms also perpetuate the discriminatory enforcement patterns of the 'pre-digital' carceral system, reproducing them at algorithmic speed and scale (Angwin et al., 2016; Kovacs & Ranganathan, 2020).

Table 4. Distributive Justice Impacts: Marginalised Groups and AI Surveillance Risks in India

Marginalised Group	Primary AI Risk	Mechanism of Harm	Documented Evidence	Key Source
Dalits / Scheduled Castes	Biometric exclusion;	Biometric failure; Biased training data	Exclusion in Rajasthan, Jharkhand, AP	Khera et al. (2021)

Marginalised Group	Primary AI Risk	Mechanism of Harm	Documented Evidence	Key Source
	Predictive policing			
Adivasis / Scheduled Tribes	Biometric exclusion; Forest surveillance	Remote area authentication failure	Exclusion in Odisha, Chhattisgarh	Drèze & Khera (2017)
Religious Minorities	FRT misidentification; Targeting	Over-policing; Biased annotation	Delhi 2020 protests documentation	IFF (2023); HRW (2022)
Urban Poor / Informal Workers	Smart city surveillance	Geospatial targeting; Displacement	Surat, Lucknow ICCR reports	Bhan et al. (2020)
Women	FRT gender bias; Data profiling	Higher FRT error rates; Gender gap	IFF data; NFHS-5 digital divide data	Arun (2019); IIPS (2021)
Political Activists / Journalists	Smart city; Social media AI	Chilling effect; Targeted monitoring	Article 19 India Report (2022)	Article 19 (2022)

Note. Compiled from Khera et al. (2021), IFF (2023), Bhan et al. (2020), Arun (2019), and Human Rights Watch (2022).

5. FINDINGS: DATA LANDSCAPE OF AI GOVERNANCE IN INDIA

5.1 Scale and Trajectory of AI Surveillance Investment

The adoption of AI surveillance systems in India has been gradually increasing. At the National level, the allocations have grown for AI and data infrastructure. 1,006 crore in FY2017–18 to Rs. 7,629 crore in FY2023–24, with the Interim Union Budget 2024–25 proposing a further 40% increase (MeitY, 2024; Union Budget, 2024). It is estimated that the investments by the government, available under the Smart Cities Mission, PMGSY grants for digitisation and police modernisation grants provide an additional amount of Rs. 4,000–6,000 crore annually (Ministry of Home Affairs, 2023). Co-investment is not negligible and mostly private, including public-private partnerships with known vendors such as NEC Corporation, Innenu Labs, Staqu Technologies, and Videonetics.

Table 5. India's Public Investment in AI and Digital Surveillance Infrastructure, FY2017–18 to FY2023–24

Financial Year	MeitY Budget (Rs. crore)	AI (Rs. crore)	Smart Cities AI Allocation (Rs. crore)	Police Tech Modernisation (Rs. crore)	Total (Estimated, Rs. crore)
2017–18	1,006		640	1,200	2,846
2018–19	1,284		890	1,450	3,624

Financial Year	MeitY Budget (Rs. crore)	AI (Rs. crore)	Smart Cities AI Allocation (Rs. crore)	Police Tech Modernisation (Rs. crore)	Total (Estimated, Rs. crore)
2019–20	1,796		1,120	1,680	4,596
2020–21	1,490		890	1,320	3,700
2021–22	2,340		1,380	1,840	5,560
2022–23	4,908		1,760	2,200	8,868
2023–24	7,629		2,100	2,640	12,369

Note. MeitY figures from Union Budget documents. Smart Cities and Police Modernisation figures are estimated from published grant disbursement data (MeitY, 2024; MHA, 2023; Smart Cities Mission, 2023).

5.2 Legislative and Regulatory Gaps

India's AI governance framework has seen substantial shifts over the last several years, yet it is far from robust. In 2019, the Srikrishna Committee had recommended a Personal Data Protection Bill which was subsequently amended and withdrew in 2022. A new Digital Personal Data Protection (DPDP) Act was passed in August 2023, introducing a Data Protection Board and consent-based regulations on the processing of personal data along with wide exemptions for government bodies on national security, public order and sovereignty concerns (Government of India, 2023). These exemptions, in reality, do not cover the most powerful forms of AI surveillance, such as NAFRS, ICCC systems, or predictive policing.

The lack of a framework specific to AI is also significant. The policy vision from NITI Aayog (2021) for Responsible AI for All and the AI governance principles of the Ministry of Electronics and Information Technology (2023) are important, but both are guidance documents, not laws. The Telecom Regulatory Authority of India (TRAI) and the Competition Commission of India (CCI) have also made relevant recommendations on data and algorithmic governance, but not for public sector AI deployment. The India there is no legislative requirements for algorithmic impact assessment, mandatory bias audit or human oversight of high-stakes automated decisions becoming standard jurisdictions across the E.U. (European Parliament, 2024) and other countries.

Table 6. India's Regulatory Framework for AI Governance: An Assessment

Regulatory Instrument	Year	Scope	Legally Binding?	Government AI Exemptions	Assessment
Puttaswamy Judgment (Privacy)	2017	Constitutional right to privacy	Yes (constitutional)	Permitted with proportionality test	Strong but implementation weak
Aadhaar Amendment Act	2019	Biometric identity system	Yes	Extensive government carve-outs	Partially protective
IT Act / IT Rules 2021	2000/2021	Intermediary liability; Data	Yes	Wide national security exemptions	Inadequate for AI

Regulatory Instrument	Year	Scope	Legally Binding?	Government AI Exemptions	Assessment
Digital Personal Data Protection Act	2023	Personal data processing	Yes	Broad government exemptions	Significant gaps for surveillance
NITI Aayog Responsible AI	2021	Principles for AI governance	No (advisory)	N/A	Positive but non-enforceable
MeitY AI Governance Principles	2023	Ethical AI deployment	No (advisory)	N/A	Positive but non-enforceable
No AI-Specific Legislation	—	—	—	—	Critical regulatory gap

Note. Compiled from Supreme Court of India (2017), Government of India (2023), NITI Aayog (2021), and MeitY (2023).

5.3 Civil Society Documentation and Institutional Trust

The link between AI surveillance and institutional trust, one of the fundamental aspects of sustainable governance, is moderated by public awareness, perceived fairness and the functioning of redress mechanisms. The available survey data on the public opinion of AI surveillance in India is revealing. According to the Centre for the Study of Developing Societies (CSDS) survey of 2022, 58% of the urban population supported the use of AI surveillance for public safety, while 67% believed that the government could use the surveillance data for illegal purposes (CSDS, 2022). Rural respondents showed less awareness of AI surveillance technologies and larger shares of negative experiences have been recorded with biometric authentication technologies, particularly during welfare delivery.

Civil society organisations have reported instances of AI-powered targeted surveillance of journalists, activists and political opponents that do not have a legitimate law enforcement purpose (IFF, 2023; Article 19, 2022). As outlined in the Pegasus Project investigation, the application of Pegasus spyware to Indian journalists and civil society activists was not AI-driven but rather part of a wider digital surveillance landscape in which public sector AI is used (Forbidden Stories, 2021).

6. TOWARDS AN ETHICAL FRAMEWORK FOR AI GOVERNANCE IN INDIA

6.1 Principles for Sustainable AI Governance

Based on the above analysis, this section suggests a normative model for ethical governance of AI in India, based on six principles. These principles are based on the convergence of the constitutional framework of India, the growing international consensus on AI governance ethics, as outlined in the OECD guidelines (2019) and UNESCO guidelines (2021), and the unique governance challenges described in the previous sections.

The first principle is constitutional primacy: AI governance frameworks should be explicitly rooted in and should be subject to the constitutional guarantees of the fundamental rights of the Indian Constitution, such as the right to equality (Article 14), freedom of expression and association (Article 19), and the right to life and liberty with privacy (Article 21) as interpreted in *Puttaswamy*. This principle does not just mean AI systems must not infringe on rights in "outrageous" or "unacceptable" ways, but that they should embody and promote the values in the Constitution.

The second is democratic authorisation: AI surveillance systems that could have a major impact on citizens' rights should be authorised by Parliament or state legislatures in specific, clear laws, not be activated under a general

executive order or government rule. This principle is rooted in the legality of the Puttaswamy proportionality framework, and in its aim to address the accountability gap arising out of the present system of executive order or tender deployment.

The third principle is algorithmic transparency: any algorithm used by any AI system in decisions that can significantly affect the lives of people, such as those in the public sector related to eligibility for welfare, law enforcement or sentencing, must be made comprehensible to those who are affected, and be independently technically audited. This entails the need for the procurement specification to include a requirement for explainability as a design criterion, and for an audit obligation to be maintained by a third party public body.

Equity by design is the fourth principle – AI systems are to be assessed for distributive effects within demographic groups (such as caste, religion, gender and geographic location) before they are deployed, with bias impact assessment including representative data of the Indian context, not international proxies. If the bias is not sufficiently reduced in a system, it must not be put into place in a high stakes environment.

The fifth principle is federal coordination: because of the federal nature of India, AI governance needs to be framed in a coordinated way to set the minimum standards nationally, while allowing for innovative approaches in states, within rights-protective limits. This coordination could be done by a national AI governance body that has binding authority to establish standards for high risk AI use in public administration, to monitor state-level AI deployments and to hear and investigate complaints.

The sixth principle is participatory governance: communities which are most affected by AI surveillance, such as marginalised caste and tribal communities, urban poor populations, and religious minorities, need to be able to meaningfully engage in governance decision making processes having an impact on them. This encompasses mandatory public consultation for high-risk AI deployments, representation of the community on oversight bodies, and readily accessible mechanisms for redress concerning AI issues.

Table 7. Normative Framework for Ethical AI Governance in India: Six Principles

Principle	Core Requirement	Institutional Mechanism	Constitutional Grounding	International Parallel
Constitutional Primacy	All AI deployment subordinate to fundamental rights	Judicial review; Parliamentary oversight	Arts. 14, 19, 21; Puttaswamy (2017)	EU Charter of Fundamental Rights
Democratic Authorisation	Explicit legislation for high-risk AI	Parliament / State Legislature mandate	Rule of law; Art. 13	EU AI Act Art. 9–11
Algorithmic Transparency	Explainability; Independent audit	National AI Audit Authority	Art. 21 (right to reasoned decision)	GDPR Art. 22; OECD AI Principles
Equity by Design	Mandatory bias impact assessment	Pre-deployment equity audit	Art. 14 (equality); Art. 15 (non-discrimination)	NYC Local Law 144 (2023)
Federal Coordination	National minimum standards; State experimentation	National AI Governance Authority	Seventh Schedule; Art. 246	Canada AI and Data Act

Principle	Core Requirement	Institutional Mechanism	Constitutional Grounding	International Parallel
Participatory Governance	Community consultation; Accessible redress	AI Grievance Redress Mechanism	Art. 19(1)(a); Art. 32 / 226	UNESCO Recommendation on AI (2021)

Note. Author's framework developed from Puttaswamy (2017), NITI Aayog (2021), UNESCO (2021), OECD (2019), and European Parliament (2024).

6.2 Institutional Architecture for AI Oversight

The implementation of these principles requires institutional architecture that India currently lacks. Drawing on comparative experience with AI governance in the European Union, Canada, Singapore, and Brazil, three institutional innovations are particularly relevant. First, the independent National AI Regulatory Authority (NAIRA) should be established composed of technical, legal and civil society experts and given the authority to classify AI systems by risk level, request an impact assessment of high-risk deployments, carry out compliance checks and take penalties. Most significantly the Authority's (authority's) mandate should be extended to government bodies as well as private sector actors to fill the regulatory gap in the DPDP Act as it stands now.

Second, an AI Ombudsman mechanism, ideally part of the existing structure of the National Human Rights Commission or a dedicated institution, should be available and available at an affordable cost to those who are affected by harms resulting from the use of AI in government decisions. The mechanism should have powers to investigate, award compensation, and recommend systemic changes. Third, mandatory algorithmic impact assessments, modelled on environmental impact assessment processes familiar from Indian administrative practice, should be required before any AI system is deployed in high-stakes public administration contexts.

7. DISCUSSION

7.1 The Paradox Revisited

The analysis presented in this paper confirms and deepens the articulation of the digital governance paradox introduced in Section 1. India's AI-driven surveillance expansion is neither uniformly harmful nor uniformly beneficial; it is an arena of genuine ethical complexity in which substantial efficiency and welfare delivery gains coexist with serious rights violations, accountability deficits, and discriminatory impacts. Resolving the paradox does not require choosing between efficiency and rights—a false binary that tends to serve the interests of those who benefit from the status quo—but rather designing institutional frameworks that make AI governance simultaneously more effective and more accountable.

The Indian case also illuminates a tension in the sustainable governance literature. Much of the normative literature on sustainable public administration implicitly assumes a governance context in which state capacity is adequate and the primary challenges are those of institutional design. In India, as in many developing country contexts, the efficiency case for AI governance is inseparable from developmental imperatives: biometric welfare delivery genuinely serves millions of previously excluded beneficiaries, and AI-enabled e-governance genuinely reduces corruption that undermines sustainable development. The normative framework proposed in Section 6 attempts to honour both the efficiency imperatives and the rights commitments, but acknowledges that the political economy of implementation will be deeply contested.

7.2 Comparative Lessons

India's experience with AI surveillance governance offers lessons both for the country and for the broader international debate on AI regulation. Positively, India's constitutional framework—particularly the Puttaswamy judgment's robust articulation of privacy as a fundamental right with specific applicability to state surveillance—provides stronger foundational protections than many comparative jurisdictions and could, if properly implemented, provide a model for rights-based AI governance in the Global South. The Supreme Court's desire to tackle intricate technical considerations in digital rights disputes indicates both that judicial processes can meaningfully contribute

to governance of artificial intelligence even without detailed laws and that the law is more complex than many people realize.

On the negative side, India's experience shows the dangers of implementing AI in an environment without regulations. The institutional path-dependency and build-up of harms that ensued from the common pattern in this paper of surveillance technologies being introduced first, and regulated (if at all) later will be hard to unravel. The concentration of AI expertise within government agencies and private vendors, combined with the legal opacity of procurement processes and the commercial confidentiality claims of technology suppliers, creates structural information asymmetries that weaken both legislative oversight and judicial review. These patterns are not unique to India and point to the inadequacy of relying on ex-post regulation for governance of technologies with rapid adoption curves.

Table 8. Comparative AI Governance Approaches: Lessons for India

Country	AI Governance Approach	Legal Basis	Government AI Scope	Civil Society Role	India Lessons
European Union	Risk-based regulation	EU AI Act (2024)	Included (limited exemptions)	Consultation in rulemaking	Comprehensive model but complex
United States	Sector-by-sector; Executive Orders	No comprehensive AI law	Limited federal oversight	Strong litigation culture	Flexibility risk: patchwork
Canada	Directive on Automated Decision-Making	AIDA (proposed)	Government AI explicitly covered	Strong access-to-information	Government inclusion critical
Singapore	Model AI Governance Framework	Advisory (non-binding)	Partially covered	Industry partnership focus	Voluntary insufficient
Brazil	AI Bill (under consideration)	Proposed legislation	Government use regulated	Civil society advocacy strong	Participatory process model
India (Current)	Advisory principles only	DPDP Act (partial)	Wide government exemptions	Constrained but active	Legislative gap critical

Note. Compiled from European Parliament (2024), Government of Canada (2023), IMDA Singapore (2023), and Government of India (2023).

7.3 Limitations and Future Research

Some of the limitations of this study are as follows: AI impacts analysis is largely built on civil society documentation and judicial records, which are not provided by the government, and has relied on a high degree of opacity in the governance of AI in India, given that there is little information about deployments, error rates, or exclusion effects published by the government. The framework put forward in Section 6 is rooted in constitutional principles and comparative experience, but it needs to be tested in practice, through engagement with stakeholders in affected

communities, government bodies and technology vendors. The study calls for further research to conduct large-scale surveys of citizens' experiences with AI-driven governance; carry out technical audits of currently deployed AI systems; and compare the implementation of AI governance within different Indian states with varying political economies and institutional capacities.

8. CONCLUSION

The ethical issues surrounding the use of AI for sustainable public administration in India have been explored, highlighting a governance paradox that suggests while the technology offers potential for development, it also brings about significant human rights violations, accountability gaps, and discriminatory effects. The analysis has made clear that this paradox does not have to be an inherent feature of AI deployment: it is a result of specific regulatory decisions – or the lack of them – that have led to the growth of surveillance systems that have proliferated without proper legal authorization, oversight, or democratic discussion.

The six principles of constitutional primacy, democratic authorisation, algorithmic transparency, equity by design, federal coordination and participatory governance that underlie the normative framework proposed in Section 6 provide a way to reconfigure the governance of AI in India. The framework will need to be put into practice through legislation that creates an independent National AI Regulatory Authority, mandatory requirements for algorithmic impact assessments, and a detailed AI governance legislation that provides meaningful protections for government use. It also needs institutional investment in developing the technical capacity of oversight bodies, digital literacy of citizens, and the AI competency of the judiciary.

The stakes of righting AI governance in India are truly extraordinary. India's decisions regarding AI governance are significant not just for the nation but for the Global South and the world beyond, given its scale, diversity, and democratic commitments. The paradox of digital governance does not have to be irreversible, but it must be overcome, and the challenge is political: putting the rights-protecting demands of constitutional democracy ahead of the efficiency demands of the surveillance state. AI is the opportunity to make public administration more sustainable, but only when technology is put at the service of human values.

REFERENCES

- [1] Abraham, R. (2018). The Aadhaar paradox: Inclusion and surveillance in India's biometric state. *Economic and Political Weekly*, 53(12), 30–37.
- [2] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [3] Article 19. (2022). India: Digital repression and freedom of expression. Article 19 India and South Asia Programme.
- [4] Arun, C. (2019). AI and the global South: Designing for other worlds. In K. Yeung & M. Lodge (Eds.), *Algorithmic regulation* (pp. 177–195). Oxford University Press.
- [5] Arora, P. (2019). *The next billion users: Digital life beyond the West*. Harvard University Press.
- [6] Bannister, F., & Connolly, R. (2012). Forward to the past: Lessons for the future of e-government from the story so far. *Information Polity*, 17(3–4), 211–226. <https://doi.org/10.3233/IP-2012-0276>
- [7] Bhan, G., Bhide, A., & Nair, P. (2020). Smart cities, smart exclusions? Urban policy and the limits of AI governance in India. *Contemporary South Asia*, 28(3), 286–302.
- [8] Bhatia, G. (2021). *Privacy and the state: Aadhaar, surveillance, and the Constitution of India*. Cambridge University Press.
- [9] Big Brother Watch. (2022). Face off: The lawless growth of facial recognition in UK policing. Big Brother Watch.
- [10] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research*, 81, 149–159.

- [11] Browne, S. (2015). *Dark matters: On the surveillance of blackness*. Duke University Press.
- [12] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 1–15.
- [13] Centre for the Study of Developing Societies [CSDS]. (2022). Digital governance and citizen attitudes in urban India: Survey findings. CSDS Policy Brief No. 7.
- [14] Cobbe, J. (2019). Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Legal Studies*, 39(4), 636–655. <https://doi.org/10.1017/lst.2019.9>
- [15] Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy and Technology*, 29(3), 245–268. <https://doi.org/10.1007/s13347-015-0211-1>
- [16] Datta, A. (2020). Self(ie)-governance: Technologies of intimate surveillance in India under COVID-19. *Dialogues in Human Geography*, 10(2), 234–237. <https://doi.org/10.1177/2043820620934797>
- [17] Deibert, R. (2020). *Reset: Reclaiming the internet for civil society*. Anansi Press.
- [18] Drèze, J., & Khera, R. (2017). Recent social security initiatives in India. *World Development*, 98, 555–572. <https://doi.org/10.1016/j.worlddev.2017.05.035>
- [19] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- [20] European Parliament. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
- [21] Forbidden Stories. (2021). The Pegasus project. Forbidden Stories Consortium. <https://forbiddenstories.org/pegasus>
- [22] Fountain, J. E. (2001). *Building the virtual state: Information technology and institutional change*. Brookings Institution Press.
- [23] Government of India. (2023). The Digital Personal Data Protection Act, 2023. Ministry of Law and Justice, Gazette of India.
- [24] Human Rights Watch. (2022). India: Events of 2021. Human Rights Watch World Report 2022.
- [25] Internet Freedom Foundation [IFF]. (2023). Surveillance landscape in India: Mapping facial recognition deployments. IFF Working Paper No. 12.
- [26] International Institute for Population Sciences [IIPS]. (2021). National Family Health Survey-5 (NFHS-5) 2019–21: India report. Ministry of Health and Family Welfare.
- [27] Janowski, T. (2015). Digital government evolution: From transformation to contextualization. *Government Information Quarterly*, 32(3), 221–236. <https://doi.org/10.1016/j.giq.2015.07.001>
- [28] Jørgensen, T. B., & Bozeman, B. (2007). Public values: An inventory. *Administration and Society*, 39(3), 354–381. <https://doi.org/10.1177/0095399707300703>
- [29] Khera, R. (2017). Impact of Aadhaar in welfare programmes. *Economic and Political Weekly*, 52(50), 61–70.
- [30] Khera, R. (2019). *Dissent on Aadhaar: Big data meets big brother*. Orient BlackSwan.
- [31] Khera, R., Nandini, S., & Gupta, M. (2021). Aadhaar authentication failure rates: Evidence from public distribution system surveys. *Journal of Development Economics*, 150, 102633. <https://doi.org/10.1016/j.jdeveco.2021.102633>
- [32] Kovacs, A., & Ranganathan, N. (2020). Data protection for whom? Caste, surveillance, and the political economy of India's data governance. *IT for Change Working Paper*.
- [33] Lyon, D. (2007). *Surveillance studies: An overview*. Polity Press.

- [34] Ministry of Electronics and Information Technology [MeitY]. (2022). Annual report 2021–22. Government of India.
- [35] Ministry of Electronics and Information Technology [MeitY]. (2023). India AI principles for responsible AI development and deployment. Government of India.
- [36] Ministry of Electronics and Information Technology [MeitY]. (2024). Interim budget 2024–25: Key allocations for digital infrastructure. Government of India.
- [37] Ministry of Home Affairs [MHA]. (2023). Annual report 2022–23: Modernisation of state police forces. Government of India.
- [38] Misra, D., & Mathur, N. (2021). Artificial intelligence in Indian governance: Applications and challenges. NITI Aayog Discussion Paper.
- [39] National Crime Records Bureau [NCRB]. (2020). Automated facial recognition system: Request for proposal. Ministry of Home Affairs, Government of India.
- [40] National Crime Records Bureau [NCRB]. (2022). Prison Statistics India 2021. Ministry of Home Affairs.
- [41] NITI Aayog. (2021). Responsible AI for all: Adopting the framework—A use case approach for all. Government of India.
- [42] NITI Aayog. (2023). Direct benefit transfer: Progress and impact report. Government of India.
- [43] OECD. (2019). Recommendation of the Council on Artificial Intelligence. Organisation for Economic Co-operation and Development. <https://doi.org/10.1787/7bc5e87e-en>
- [44] Privacy International. (2023). State of privacy India. Privacy International Country Reports.
- [45] Ramanathan, U. (2018). A unique identity bill. *Economic and Political Weekly*, 45(30), 10–14.
- [46] Ramanathan, U. (2023). Surveillance and Indian law: The gap between aspiration and practice. *Indian Law Review*, 7(1), 1–23. <https://doi.org/10.1080/24730580.2023.2167428>
- [47] Raval, N., & Dourish, P. (2016). Standing on the shoulders of giants: Socio-technical infrastructures for digital labor in India. *Proceedings of the ACM Conference on Computer-Supported Cooperative Work*, 19, 1–14.
- [48] <https://doi.org/10.1145/2818048.2819953>
- [49] Rawls, J. (1971). *A theory of justice*. Harvard University Press.
- [50] Reserve Bank of India. (2023). Annual report 2022–23. RBI.
- [51] Smart Cities Mission. (2023). Smart Cities Mission: Annual report 2022–23. Ministry of Housing and Urban Affairs, Government of India.
- [52] Srikrishna Committee. (2018). A free and fair digital economy: Protecting privacy, empowering Indians—Report of the Committee of Experts under the Chairmanship of Justice B.N. Srikrishna. Ministry of Electronics and Information Technology.
- [53] Srinivasan, J. (2021). Civil society and smart city surveillance: Perspectives from India. *Surveillance and Society*, 19(2), 217–230. <https://doi.org/10.24908/ss.v19i2.13784>
- [54] Supreme Court of India. (2017). Justice K.S. Puttaswamy (Retd.) and Another v. Union of India and Others. Writ Petition (Civil) No. 494 of 2012. AIR 2017 SC 4161.
- [55] Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data and Society*, 4(2), 1–14. <https://doi.org/10.1177/2053951717736335>
- [56] Toyama, K. (2015). *Geek heresy: Rescuing social change from the cult of technology*. PublicAffairs.
- [57] Unique Identification Authority of India [UIDAI]. (2024). Aadhaar dashboard. <https://uidai.gov.in/en/aadhaar-dashboard.html>

- [58] United Nations Development Programme [UNDP]. (2022). Global Multidimensional Poverty Index 2022: Unpacking deprivation bundles to reduce multidimensional poverty. UNDP & OPHI.
- [59] UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific, and Cultural Organization.
- [60] Union Budget. (2024). Interim Union Budget 2024–25: Statement of revenue expenditure—Ministry of Electronics and Information Technology. Ministry of Finance, Government of India.
- [61] Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—Applications and challenges. *International Journal of Public Administration*, 42(7), 596–615. <https://doi.org/10.1080/01900692.2018.1498103>
- [62] Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation and Governance*, 12(4), 505–523. <https://doi.org/10.1111/rego.12158>