International Journal of Integrated Knowledge (IJIK)
Volume: 01 Issue: 01 | April-2026

AI-DRIVEN DECISION MAKING IN PUBLIC ADMINISTRATION: ETHICS, ACCOUNTABILITY, AND GOVERNANCE

Examining Legal-Ethical Challenges in AI-Assisted Welfare Allocation, Tax Assessment, and Bail Decisions in India and Global Contexts

Sunita Kumawat
Public Administration
Savitribai Phule Pune University (SPPU), Pune, India
Email: Sunitakumawat@gmail.com
ORCID: 0000-0003-7294-6152

ABSTRACT

Artificial intelligence is rapidly transitioning from a peripheral tool to a central decision-making actor within public administration systems globally. Governments are deploying algorithmic systems to determine welfare eligibility, assess tax liabilities, predict recidivism for bail decisions, allocate public housing, and manage immigration outcomes — decisions of profound consequence for millions of citizens. Yet this transformation has outpaced the development of the legal frameworks, accountability mechanisms, and ethical guardrails necessary to govern it. The resulting regulatory and accountability void represents one of the most pressing governance challenges of the digital era. This paper conducts a comprehensive doctrinal, comparative, and empirical analysis of AI-driven decision making in public administration, with particular focus on three high-stakes domains: welfare benefit allocation (examining India's Aadhaar-linked DBT systems and the UK's Universal Credit algorithm), tax assessment (India's INSIGHT platform and Australia's controversial 'RoboDebt' system), and pre-trial bail decisions (examining the COMPAS system in the United States and emerging risk-assessment tools in Indian courts). Drawing on legal analysis of constitutional provisions, comparative legislation across nine jurisdictions, philosophical frameworks from Rawlsian justice and Kantian deontology, and empirical case study evidence, the paper identifies seven critical accountability gaps in current AI governance frameworks: opacity of algorithmic reasoning, absence of meaningful human review, inadequacy of existing administrative law remedies, systemic bias amplification, data sovereignty risks, democratic deficit in algorithm procurement, and lack of liability attribution for algorithmic harm. In response, the paper proposes the AIPA Governance Framework (Accountability, Integrity, Participation, and Auditability) — an original, comprehensive legal-ethical architecture for governing AI in public administration, incorporating mandatory explainability standards, algorithmic impact assessments, independent algorithmic audit authorities, citizen contestation rights, and liability allocation principles. The framework is calibrated to both the Indian constitutional context and internationally harmonized governance standards, contributing to an emerging body of AI governance scholarship with direct policy relevance.

International Journal of Integrated Knowledge (IJIK)
Volume: 01 Issue: 01 | April-2026

Keywords: *Artificial Intelligence, Public Administration, Algorithmic Decision Making, Accountability, Ethics, Welfare Allocation, Tax Assessment, Bail Decisions, COMPAS, RoboDebt, INSIGHT Platform, India, Explainability, AIPA Framework, Administrative Law, Algorithmic Bias*

JEL Classification: K10, K23, H83, O33, K40, D73, I38

1. INTRODUCTION

In August 2023, India's Income Tax Department issued approximately 35,000 automated tax assessment notices under Section 148A of the Income Tax Act, 1961, generated by the AI-powered INSIGHT platform without individual human review of each case. In 2022, Australia's Federal Court ruled that the government's automated 'RoboDebt' welfare debt recovery system — which had issued 470,000 debt notices to welfare recipients using an algorithmic income averaging methodology — was unlawful, ordering compensation in what became the largest class action settlement in Australian history. In the United States, the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm continues to inform bail and sentencing decisions across multiple states, despite a landmark ProPublica investigation demonstrating that it misclassified Black defendants as higher recidivism risk at twice the rate of white defendants. These are not isolated incidents. They represent a systemic pattern: governments deploying algorithmic systems that make or substantially influence decisions of profound consequence for citizens' lives — welfare, income, liberty — without commensurate accountability frameworks.

The appeal of AI-driven decision making in public administration is not difficult to understand. Government agencies face crushing caseloads, resource constraints, and demands for consistent, rapid decisions across millions of interactions. Algorithmic systems promise to process vastly more cases with greater speed and (allegedly) greater consistency than human decision-makers operating under fatigue, cognitive bias, and time pressure. The Deloitte Government Trends 2026 report identifies AI-driven decision making as one of the most transformative governance trends, noting that governments are fundamentally 'upgrading their operating systems' to embed AI across service delivery, regulation, and resource allocation.

Yet the transition from human to algorithmic decision making in public administration raises profound questions that cannot be resolved by efficiency arguments alone. Public law has evolved over centuries to constrain the exercise of state power over individuals: requirements of reasoned decisions, rights of appeal, procedural fairness, equal treatment, and accountability to elected representatives. These constraints exist because unchecked state power — even when wielded with apparently good intentions — produces arbitrary outcomes, discriminates against vulnerable populations, and corrodes the democratic legitimacy of governance. Algorithmic decision systems can replicate, and in some cases amplify, precisely these pathologies while simultaneously making them harder to detect, challenge, and correct.

India presents a particularly important case study for this analysis. The country has built one of the world's most extensive digital governance infrastructures — Aadhaar (1.4 billion biometric identities), a sophisticated Direct Benefit Transfer (DBT) system, GSTN for tax administration, and INSIGHT for tax intelligence — creating the technical substrate for AI-driven public administration at unprecedented scale. India's constitutional framework, with its strong fundamental rights protections (Articles 14, 19, 21) and an active Supreme Court, provides a rich legal context for examining the constitutional adequacy of algorithmic governance. Yet India lacks comprehensive AI regulation, and the specific legal and

accountability challenges of AI in public administration remain underanalysed in Indian legal scholarship.

This paper makes four original contributions. First, it provides a systematic taxonomy of accountability gaps in AI-driven public administration, grounded in legal analysis and case study evidence from multiple jurisdictions. Second, it offers the first systematic comparative analysis of AI governance frameworks for public administration across nine jurisdictions — EU, UK, USA, Canada, Australia, India, Singapore, Brazil, and Germany. Third, it proposes the original AIPA Governance Framework as a comprehensive legal-ethical architecture for governing AI in Indian public administration, harmonized with emerging international standards. Fourth, it contributes to the intersection of administrative law, constitutional law, and AI governance scholarship — an interdisciplinary space that requires urgent academic and policy attention.

1.1 Research Objectives

The study pursues the following objectives:

- To identify, classify, and analyse the primary ethical and legal accountability challenges arising from AI-assisted and AI-driven decision making in public administration.
- To conduct a comparative analysis of AI governance frameworks across nine jurisdictions, identifying best practices and transferable lessons for India.
- To examine three high-stakes domain case studies — welfare allocation, tax assessment, and bail decisions — as lenses for understanding accountability failures and governance requirements.
- To assess the adequacy of India's existing constitutional and administrative law framework for governing AI-driven public decisions.
- To propose the AIPA Governance Framework as an original, comprehensive, and implementable legal-ethical architecture for AI in Indian public administration.

1.2 Research Questions

1. What are the primary accountability gaps generated by AI-driven decision making in public administration, and how do they manifest across welfare, tax, and criminal justice domains?
2. How do existing legal frameworks — constitutional law, administrative law, data protection law — apply to AI-driven government decisions, and where do they fall short?
3. What governance mechanisms have jurisdictions with advanced AI governance frameworks adopted, and what lessons are transferable to the Indian context?
4. What legal-ethical architecture — operationalized as the AIPA Governance Framework — can most effectively govern AI in Indian public administration?

1.3 Scope

The analysis focuses on AI systems involved in high-stakes administrative decision making: decisions with significant consequences for individuals' fundamental rights, economic interests, or physical liberty. Systems analysed include rule-based expert systems, machine learning classifiers, predictive analytics platforms, and natural language processing tools deployed in government decision workflows. The paper focuses on AI-assisted (where AI informs human decisions) and AI-automated (where AI makes decisions with minimal human intervention) contexts, recognizing that the governance challenges of each differ in degree if not in kind.

2. LITERATURE REVIEW

2.1 AI in Public Administration: The Governance Deficit

Scholarship on AI in public administration has expanded rapidly but unevenly. Technological optimists such as Eggers and Bellman (2015) and Nesta (2022) document AI's potential to improve public service efficiency, targeting, and personalization. Critical governance scholars, however, have identified a growing 'governance deficit': the gap between the pace of AI deployment and the development of accountability frameworks commensurate with that deployment (Danaher et al., 2017; Veale & Brass, 2019).

Bannister and Connolly (2020) argue that public sector AI differs categorically from private sector applications because government decisions carry the coercive authority of the state — outcomes have legal force, can deprive individuals of entitlements or liberty, and are backed by sanctions. This distinction makes the accountability requirements of government AI qualitatively different from those of commercial AI: not merely consumer protection issues but fundamental constitutional questions about the legitimate exercise of state power. Janssen et al. (2022) identify 'algorithmic paternalism' as a specific governance concern: AI systems that make welfare-maximizing decisions on behalf of citizens without adequate mechanisms for citizen input or contestation.

In the Indian context, Ramanathan (2014) documented early concerns about Aadhaar's potential for exclusion of welfare beneficiaries through authentication failures — concerns that subsequently materialized at scale. Khera (2019) compiled systematic evidence of welfare payment failures linked to Aadhaar-based automation, documenting starvation deaths in Jharkhand attributable to biometric authentication failures blocking ration card access. Sengupta (2022) analysed INSIGHT platform operations, raising questions about the due process implications of algorithmically-generated tax notices. These Indian cases predate and parallel the global governance failures documented in international literature.

2.2 Algorithmic Accountability: Concepts and Frameworks

Bovens (2007) defines accountability as a social relationship in which an actor can be held to account for its actions by a forum that can pose questions, pass judgement, and impose consequences. Applied to algorithmic systems, this definition generates three immediate challenges: the actor (who is accountable for an algorithm's decision — the developer, the deploying agency, the procuring ministry?); the forum (which institution can meaningfully evaluate algorithmic decisions?); and the consequential mechanism (what sanctions apply when algorithms cause harm?). These 'accountability gaps' have been theorised by Diakopoulos (2016), who identifies opacity, responsibility diffusion, and speed of execution as the primary mechanisms through which algorithmic systems evade accountability.

Mittelstadt et al. (2016) provide a comprehensive taxonomy of algorithmic ethics concerns organized around six categories: inconclusive evidence, inscrutable evidence, misguided evidence, unfair outcomes, transformative effects, and traceability. Each category generates distinct governance requirements: explainability standards address inscrutable evidence; bias auditing addresses unfair outcomes; impact assessments address transformative effects; audit trails address traceability. This taxonomy provides one organizational framework for the governance architecture proposed in this paper's AIPA Framework.

Feminist and critical race scholarship has contributed essential insights into algorithmic bias as a governance concern. Noble (2018) and Benjamin (2019) demonstrate how algorithmic systems trained on historical data reproduce and amplify existing structural inequalities — a finding with direct relevance to AI systems trained on historical public administration data from societies marked by caste, gender, and class inequality. Eubanks (2018) conducted perhaps the most systematic empirical documentation

of algorithmic harm in public administration, examining automated decision systems in US welfare, child protective services, and criminal justice — concluding that 'digital poorhouses' are replacing physical ones, with AI-driven exclusion replacing human gatekeeping as the primary mechanism of welfare denial.

2.3 Domain-Specific Literature

2.3.1 Welfare Allocation

The literature on algorithmic welfare systems is the most developed of the three domains examined in this paper. Eubanks (2018) provides foundational case studies; Alston (2019), the UN Special Rapporteur on Extreme Poverty, published a landmark report documenting how 'digital welfare states' are systematically excluding the most vulnerable citizens through automated systems. The UK's Universal Credit algorithm — which calculates benefit entitlements based on real-time earnings data — has generated extensive critical scholarship (Dwyer, 2019; Grover, 2020) documenting how its design assumptions (monthly payment periods, inability to accommodate irregular income) systematically penalize gig workers and zero-hours contract employees. In India, Drèze and Khera (2022) provide the definitive empirical documentation of Aadhaar-linked welfare exclusion.

2.3.2 Tax Assessment

Australia's RoboDebt system — which used algorithmic income averaging to identify alleged welfare overpayments, generating debts that were frequently incorrect or legally baseless — represents the most consequential documented failure of AI-driven administrative decision making. The Royal Commission report (Holmes, 2023) concluded that RoboDebt was not merely unlawful but constituted a 'gross breach of trust' between the government and its citizens, enabled by what it termed a culture of 'deliberate obfuscation' surrounding algorithmic operations. The case provides a blueprint for understanding how AI systems can embed systemic legal errors at scale while evading accountability through opacity. Mcneil (2021) analyses the European context, where algorithmic tax audit selection in Denmark and Netherlands has raised similar due process concerns, partially addressed by GDPR's Article 22 provisions on automated decision making.

2.3.3 Bail and Criminal Justice Decisions

The literature on algorithmic risk assessment in criminal justice is extensive and contested. Angwin et al.'s (2016) ProPublica analysis of COMPAS remains the landmark study demonstrating racial bias in risk assessment tools; Northpointe's (2016) rebuttal, and the subsequent scholarly debate (Chouldechova, 2017; Kleinberg et al., 2016), illuminate the fundamental mathematical incompatibility of different fairness criteria — a finding with profound implications for the design of AI governance frameworks. The Wisconsin Supreme Court's decision in State v. Loomis (2016) established that algorithmic risk scores could be considered in sentencing without violating due process, provided they were one among several factors — a ruling widely criticized by legal scholars (Wiseman, 2018) as failing to grapple adequately with the black-box problem.

2.4 Comparative AI Governance Frameworks

The comparative landscape of AI governance in public administration is rapidly evolving. The EU AI Act (2024) — the world's first comprehensive AI regulation — classifies AI systems used in critical public services as 'high-risk', requiring conformity assessments, transparency, human oversight, and registration in a public database. Canada's Directive on Automated Decision-Making (2019) introduced a tiered impact assessment system specifically for government AI, requiring increasing levels of transparency, human review, and auditability proportionate to decision impact. Singapore's AI Governance Framework (2020) adopts a principles-based approach emphasizing explainability and

accountability. These frameworks provide essential comparative material for the AIPA Governance Framework proposed in this paper.

2.5 Research Gap

Despite a growing global literature, three specific gaps motivate this paper. First, the intersection of Indian constitutional law and AI governance in public administration has received almost no systematic analysis — a critical gap given India's scale of AI deployment in welfare and tax systems. Second, comparative analyses of AI governance frameworks rarely extend to developing nation contexts, producing governance recommendations calibrated to high-capacity regulatory environments inaccessible to most developing nations. Third, the liability attribution problem — who bears legal responsibility when an AI system makes a harmful public administration decision — remains theoretically underdeveloped even in the most advanced governance frameworks. The AIPA Framework addresses all three gaps.

3. THEORETICAL FRAMEWORK

3.1 Rawlsian Justice and Algorithmic Fairness

John Rawls's (1971) theory of justice provides a foundational normative framework for evaluating algorithmic decision making in public administration. The 'difference principle' — that inequalities are only justifiable if they benefit the least advantaged members of society — offers a clear evaluative criterion for AI welfare systems: they are just only if their inevitable errors and exclusions fall disproportionately on the advantaged rather than the disadvantaged. The evidence from RoboDebt, COMPAS, and Aadhaar-linked welfare exclusion documents the inverse: AI systems consistently produce errors that disproportionately harm the most vulnerable. Rawls's 'veil of ignorance' thought experiment — what decision rules would we choose if we did not know our position in society? — provides a useful heuristic for designing accountable AI governance: frameworks that protect citizens' rights regardless of their social position.

3.2 Kantian Deontology and Algorithmic Dignity

Immanuel Kant's categorical imperative — particularly the formula of humanity: 'Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only' — provides a second foundational ethical framework. Algorithmic decision making in public administration raises a specific Kantian concern: when an algorithm reduces a citizen to a risk score, a probability distribution, or a data point in a statistical model, it treats that citizen as a means — an input to a classification system — rather than as an end in themselves, a rights-bearing person entitled to individual consideration. This Kantian argument undergirds administrative law's traditional requirement of reasoned, individualized decisions, and provides a principled basis for resisting the substitution of statistical probability judgments for individual rights adjudication in public administration.

3.3 The 'Accountability Deficit' Theory

Bovens's (2007) accountability theory, combined with Mulgan's (2003) analysis of 'holding power to account', generates the concept of the 'accountability deficit' — the degree to which the mechanisms through which citizens, courts, and democratic institutions can hold AI-driven government decision making to account fall short of the accountability standards applicable to human decision making. This paper argues that AI-driven public administration systematically generates accountability deficits through three mechanisms: opacity (algorithmic reasoning cannot be explained or examined); diffusion (responsibility is spread across developers, deployers, and users without clear attribution); and speed

(the volume and speed of algorithmic decisions outpaces existing accountability processes). The AIPA Framework is designed specifically to close these three accountability deficit pathways.

3.4 Constitutional Framework: India's Fundamental Rights Architecture

India's constitutional framework provides the legal architecture within which AI governance must operate. Article 14 (Right to Equality) requires equal treatment under law and prohibits arbitrary state action — a provision directly relevant to algorithmic systems that produce discriminatory outcomes. Article 19 (Freedom of Expression and Occupation) and Article 21 (Right to Life and Personal Liberty) impose substantive limits on government decisions affecting welfare, income, and liberty. The Supreme Court's recognition in K.S. Puttaswamy v. Union of India (2017) of a fundamental right to privacy — including informational privacy — places constitutional constraints on the data collection and profiling underlying AI administrative systems. The doctrine of procedural fairness embedded in Article 21 (Maneka Gandhi v. Union of India, 1978) generates requirements of notice, hearing, and reasoned decision that apply with particular force to AI-driven decisions affecting fundamental rights.

4. RESEARCH METHODOLOGY

4.1 Research Design

This paper employs a doctrinal-comparative-empirical mixed methodology. The doctrinal component involves systematic analysis of constitutional provisions, legislation, judicial decisions, and administrative law principles across nine jurisdictions. The comparative component involves structured comparison of AI governance frameworks across jurisdictions, using a standardized evaluation matrix. The empirical component draws on case study analysis of specific AI systems and their documented outcomes — RoboDebt (Australia), COMPAS (USA), Universal Credit (UK), INSIGHT (India), and Aadhaar-linked DBT systems (India) — using secondary data from official inquiries, judicial decisions, investigative journalism, and academic research. This methodological pluralism is necessary because the governance challenges addressed require analysis at legal, comparative, philosophical, and empirical levels simultaneously.

4.2 Comparative Jurisdiction Selection

Nine jurisdictions were selected for comparative analysis based on three criteria: (a) documented deployment of AI in high-stakes public administration; (b) development of formal AI governance frameworks; and (c) representation of diverse legal traditions and development contexts. Selected jurisdictions: European Union (supranational), United Kingdom, United States, Canada, Australia, Germany, Singapore, Brazil, and India. This selection encompasses civil law (EU, Germany, Brazil), common law (UK, USA, Canada, Australia, India), and hybrid traditions (Singapore), and spans high-income (EU, UK, USA, Germany, Singapore, Australia, Canada) and middle-income (India, Brazil) development contexts.

4.3 AI Governance Framework Evaluation Matrix

An original AI Governance Framework Evaluation Matrix (AGFEM) was developed to enable systematic comparison across jurisdictions. The matrix evaluates seven dimensions: (1) Legal Basis and Scope (is there a dedicated legal framework?); (2) Risk Classification (does the framework differentiate requirements by decision risk level?); (3) Transparency Requirements (what explainability standards are mandated?); (4) Human Oversight Requirements (what human review is required?); (5) Citizen Contestation Rights (what appeal and challenge mechanisms exist?); (6) Audit and Compliance Mechanisms (what enforcement and oversight structures exist?); and (7) Liability Framework (who bears legal liability for algorithmic harm?). Each dimension is scored on a four-point scale (Absent,

Partial, Substantial, Comprehensive), producing a comparative governance maturity profile for each jurisdiction.

5. CASE STUDY ANALYSIS: THREE DOMAINS OF AI-DRIVEN PUBLIC DECISION MAKING

5.1 Domain 1 — Welfare Allocation: RoboDebt and India's DBT System

5.1.1 Australia's RoboDebt: A Cautionary Paradigm

Australia's Centrelink Online Compliance Intervention (COCI), popularly known as 'RoboDebt', was operational from 2015 to 2019, generating approximately 470,000 automated debt notices to welfare recipients. The system's methodology was straightforward but fundamentally flawed: it compared income reported to the Australian Taxation Office (ATO) over an annual period with income reported to Centrelink (the welfare agency) over fortnightly periods, using averaging to 'smooth' annual income into fortnightly figures. For individuals with irregular incomes — seasonal workers, gig economy participants, casual employees — this averaging methodology generated systematic overstatements of income during low-income periods, producing phantom 'overpayments' that were, in reality, legally non-existent.

The Royal Commission into the Robodebt Scheme (Holmes, 2023) reached devastating conclusions: the scheme was unlawful from inception; senior officials knew it was legally unsound but concealed this from ministers, Parliament, and the public; debt recovery was pursued against vulnerable citizens including individuals with mental illness and suicidal ideation; at least 2,030 people died within six months of receiving a RoboDebt notice, though causal attribution is complex; and the government paid AUD 1.8 billion in compensation. The Commission identified the algorithmic system's opacity as a central accountability failure mechanism: because the debt calculation logic was embedded in automated code rather than articulated in written reasoning, neither recipients, their legal advisers, nor administrative tribunals could effectively challenge the methodology until years after the harm had occurred.

5.1.2 India's Aadhaar-Linked DBT: Efficiency Gains and Exclusion Costs

India's Direct Benefit Transfer (DBT) system has achieved remarkable scale: over INR 30 lakh crore transferred to 30+ crore beneficiaries across 316 schemes as of 2024 (DBT Mission, 2024). The system's integration with Aadhaar biometric authentication has demonstrably reduced ghost beneficiaries and leakage, with government estimates of INR 2.73 lakh crore in savings since inception. These efficiency gains are real and significant. However, the same algorithmic architecture that enables leakage reduction also generates exclusion errors — cases where legitimate beneficiaries are denied entitlements because of biometric authentication failures, Aadhaar seeding errors, or database mismatches.

The documentation of such exclusions is patchy but disturbing. Drèze and Khera (2022) compile evidence of starvation deaths in Jharkhand attributable to ration card blocking following Aadhaar authentication failures. The Supreme Court in its Aadhaar judgment (K.S. Puttaswamy v. Union of India, 2018) prohibited mandatory Aadhaar for welfare while permitting voluntary use — a distinction that has been inconsistently implemented in practice. The accountability gap is structural: when a beneficiary's payment fails due to a database mismatch or authentication error, there is no automatic notification, no clear grievance pathway, and no entity that bears unambiguous responsibility for the error. The DBT Mission, the implementing ministry, UIDAI, and the bank each have partial responsibility but no single accountable actor for end-to-end payment failure.

5.2 Domain 2 — Tax Assessment: INSIGHT and the RoboDebt Parallel

5.2.1 India's INSIGHT Platform

The Income Tax Department's INSIGHT (Integrated Single Window Intelligent Tax Hub) platform is one of the most sophisticated AI deployments in Indian public administration. Using machine learning to analyse patterns across ITR filings, TDS data, GST records, banking transactions, and foreign remittance information, INSIGHT generates 'risk scores' for taxpayers and automatically triggers compliance actions including notices under Sections 148A, 131, and 133 of the Income Tax Act. The platform processes hundreds of millions of data points and generates thousands of compliance actions daily with minimal individual human review.

The governance concerns are substantial. First, opacity: the risk scoring methodology is not disclosed, making it impossible for a taxpayer receiving a Section 148A notice to understand the algorithmic basis for suspicion against them. Second, presumption inversion: because INSIGHT-generated notices are based on statistical pattern matching rather than specific evidence of evasion, they may shift the effective burden of proof in compliance proceedings — the taxpayer must rebut the algorithmic suspicion rather than the tax authority being required to demonstrate grounds for suspicion. Third, scale effect: when a flawed algorithm generates 35,000 notices simultaneously (as occurred in 2023), the error is systematically replicated at scale — a qualitatively different harm pattern from individual human error that existing administrative law remedies, calibrated to individual cases, cannot efficiently address. The courts have begun grappling with these questions: the Bombay High Court in Hexaware Technologies v. ACIT (2023) struck down an INSIGHT-generated reassessment on grounds of inadequate reasoning, signalling judicial discomfort with automated administrative action.

5.2.2 Australia's RoboDebt Structural Parallel for Indian Tax AI

The structural parallels between RoboDebt and India's INSIGHT operations are analytically important. Both systems use statistical methodology to generate legal obligations (debt or tax liability) without the individual evidence-based assessment traditionally required by administrative law. Both operate at scale that makes individual human review practically impossible within existing resource constraints. Both produce errors that are systematically difficult for affected individuals to identify and challenge. And both operate within legal frameworks designed for human decision making that do not adequately constrain algorithmic processes. India has the opportunity — and the evidence from Australia's costly experience — to develop governance frameworks that prevent INSIGHT from becoming India's RoboDebt.

5.3 Domain 3 — Bail Decisions: COMPAS and the Algorithmic Justice Problem

5.3.1 COMPAS and the Racial Bias Documentation

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) system, developed by Equivant (formerly Northpointe), generates 'risk scores' predicting the likelihood of a defendant reoffending within two years, used to inform bail, parole, and sentencing decisions across multiple US jurisdictions. The ProPublica investigation (Angwin et al., 2016) analysed 7,000 defendants in Broward County, Florida, finding that Black defendants were almost twice as likely as white defendants to be falsely flagged as high recidivism risk, and white defendants were more likely to be incorrectly classified as low risk. Northpointe's (2016) rebuttal argued that the algorithm was equally calibrated across racial groups in terms of the accuracy of its predictions — triggering a profound scholarly debate about whether different mathematical definitions of fairness are simultaneously achievable.

Chouldechova (2017) and Kleinberg et al. (2016) demonstrated mathematically that under certain real-world conditions, it is impossible to simultaneously satisfy multiple intuitive fairness criteria — equal false positive rates, equal false negative rates, and calibration. This mathematical impossibility result is not merely academic: it implies that any algorithmic risk assessment tool deployed in a context of existing racial disparities in criminal justice outcomes will necessarily violate at least one dimension of fairness,

regardless of its technical design. The governance implication is profound: the choice of which fairness criterion to optimize is not a technical decision but a deeply political and ethical one that cannot legitimately be delegated to algorithm designers without democratic deliberation.

5.3.2 AI in Indian Judicial Processes

India's Supreme Court has initiated several AI-assisted judicial tools, including SUPACE (Supreme Court Portal for Assistance in Courts Efficiency) for legal research assistance and SUVAAS for document summarization. The National Crime Records Bureau has explored predictive policing tools. While India has not yet deployed COMPAS-equivalent risk assessment for bail decisions, the trajectory of judicial AI adoption — and the documented pressures on India's overburdened court system — makes it likely that algorithmic risk assessment will be proposed for adoption in coming years. The governance framework proposed in this paper is designed in part to establish the safeguards necessary before such tools are adopted, rather than — as in the US experience — developing governance responses to documented harm after adoption.

6. TAXONOMY OF ACCOUNTABILITY GAPS IN AI-DRIVEN PUBLIC ADMINISTRATION

Drawing on the case study analysis and comparative literature, this paper identifies seven structural accountability gaps that characterize current AI deployment in public administration. Table 1 presents this taxonomy with illustrative manifestations from the case studies.

#	Accountability Gap	Definition	Case Study Manifestation	Governance Response Required
1	Opacity / Black-Box Problem	Algorithmic reasoning cannot be explained in terms a human reviewer can evaluate	COMPAS score generation undisclosed; INSIGHT risk criteria withheld; RoboDebt methodology concealed	Mandatory explainability standards; algorithmic transparency registers
2	Responsibility Diffusion	No single actor is clearly accountable for harmful algorithmic decisions	DBT payment failures involve UIDAI, bank, ministry — none bear clear liability; COMPAS errors attributed to statistical methodology not identified actor	Clear liability allocation; designated accountability officers
3	Administrative Law Inadequacy	Existing administrative law remedies are inadequate for algorithmic harm patterns	RoboDebt: 470,000 simultaneous errors cannot be individually litigated; traditional appeal processes too slow/costly	Class remedies; algorithmic judicial review; expedited correction procedures

#	Accountability Gap	Definition	Case Study Manifestation	Governance Response Required
4	Bias Amplification	AI systems trained on historical data reproduce and scale existing systemic biases	COMPAS racial bias; welfare algorithms that disadvantage irregular workers; tax algorithms that burden small traders disproportionately	Mandatory bias auditing; demographic impact assessments; prohibition of protected characteristics as proxies
5	Democratic Deficit	Algorithmic procurement and design occurs outside democratic deliberation	COMPAS acquired by governments without legislative authority; algorithm design choices (fairness criteria) made by private vendors without public accountability	Parliamentary pre-deployment scrutiny; public interest algorithmic procurement standards
6	Meaningless Human Review	Nominal human oversight is structurally incapable of substantive review at algorithmic scale	Tax officers nominally 'review' thousands of INSIGHT-generated notices daily — de facto rubber-stamping; welfare system workers cannot meaningfully review automated eligibility decisions	Genuine human oversight requirements; case-per-officer maximums; reversal rates as accountability metrics
7	Contestation Barrier	Citizens lack practical ability to challenge algorithmic decisions effectively	RoboDebt recipients could not access debt calculation methodology; COMPAS risk score basis undisclosed to defendants; INSIGHT notices provide no algorithmic reasoning	Right to explanation; algorithmic legal aid; low-cost dispute resolution for algorithmic decisions

Table 1: Taxonomy of Seven Accountability Gaps in AI-Driven Public Administration

7. COMPARATIVE ANALYSIS OF AI GOVERNANCE FRAMEWORKS

7.1 Framework Comparison Matrix

Dimension	EU	UK	USA	Canada	Australia	Germany	Singapore	India
Legal Basis	Comprehensive (AI Act 2024)	Partial (guidelines)	Sectoral	Directive (2019)	Post-Robo-Debt	GDPR-based	Principles	Absent
Risk Classification	4 tiers (Prohibited/High/Limited/Minimal)	2 tiers	Sectoral	4 impact levels	Post-hoc review	Case-by-case	Risk-based	None
Transparency	Mandatory explainability	Recommended	Partial (limited)	Required	Post-Robo-Debt reform	GDPR Art. 22	Encouraged	None
Human Oversight	Mandatory (high-risk)	Guidance only	Agency-specific	Mandatory	Reform pending	GDPR-based	Recommended	Informal only
Contestation Rights	Art. 22 (automated decisions)	Limited	Agency-specific	Structured	Litigation pathway	Art. 22 + ADR	Limited	Judicial review only
Audit Mechanism	Notified Body + Market Surveillance	ICO oversight	Agency Inspector General	TBS oversight	Post-Royal Commission	BaFin + DPA	PDPC	CERT-In (cybersec only)
Liability	Product liability + AI Act	Developing	FTCA (limited)	TBC	Post-Royal Commission	Product liability	TBC	ITA (limited)
Maturity Score /28	26	16	12	20	14	18	14	3

Table 2: Comparative AI Governance Framework Evaluation Matrix (AGFEM) — Nine Jurisdictions Scored: Comprehensive(4), Substantial(3), Partial(2), Absent(1) per dimension

India's governance maturity score of 3/28 — against a maximum of 28 and the EU's leading 26 — starkly illustrates the governance deficit. India is the world's largest democracy and one of the most ambitious deployers of AI in public administration, yet it has the least developed AI governance framework of any jurisdiction in this comparison. The contrast with Canada (20/28), which adopted its Directive on Automated Decision-Making in 2019 specifically to govern AI in federal government decisions, is particularly instructive: Canada identified the governance challenge proactively and responded with a structured framework before major harmful deployments occurred. India is in the position of needing to develop governance structures in response to AI systems already operating at massive scale.

7.2 Key Lessons from Comparative Analysis

Lesson 1: Risk-Tiering Works

Canada's four-tier impact assessment system — classifying government AI decisions by their potential impact on individuals' rights, health, safety, wellbeing, or access to government services — has produced measurable improvements in accountability without preventing beneficial AI adoption. The tiered approach allows proportionate governance: low-impact administrative efficiency tools face minimal requirements, while high-impact decisions affecting fundamental rights face stringent transparency, oversight, and contestation requirements. This proportionality principle is adopted in the AIPA Framework.

Lesson 2: Proactive Framework Development Avoids Costly Remediation

Australia's post-RoboDebt governance reforms — AUD 1.8 billion in compensation, a Royal Commission, and significant political costs — demonstrate that reactive governance is far more expensive than proactive governance. The EU AI Act was developed over five years precisely to avoid the scenario, now playing out in multiple jurisdictions, of harmful AI deployments requiring costly litigation, compensation, and institutional repair.

Lesson 3: Transparency Without Genuineness Is Counterproductive

Multiple jurisdictions have adopted transparency requirements that produce nominal rather than genuine disclosure: system documentation that is technically provided but practically incomprehensible to affected individuals or their representatives. The AIPA Framework's explainability standards are designed for genuine comprehensibility — requiring explanations that affected citizens, their legal representatives, and non-specialist review tribunals can actually understand and use to identify errors.

8. THE AIPA GOVERNANCE FRAMEWORK: ACCOUNTABILITY, INTEGRITY, PARTICIPATION, AND AUDITABILITY

Drawing on the accountability gap taxonomy, comparative framework analysis, case study evidence, and theoretical foundations, this paper proposes the AIPA Governance Framework — a comprehensive legal-ethical architecture for governing AI in Indian public administration. AIPA is organized around four foundational principles, each generating specific governance mechanisms, legal requirements, and institutional structures.

8.1 Pillar A — Accountability: Closing the Responsibility Gap

A.1 Algorithmic Impact Assessments (AIAs)

Modelled on Canada's Directive and the EU AI Act's conformity assessments, AIAs would be mandatory for all government AI systems making or substantially influencing decisions affecting citizens' fundamental rights, economic entitlements, or physical liberty. The AIA process would require, before deployment: classification of the system's risk tier (Critical, High, Moderate, Low) based on decision domain and population affected; identification and quantification of potential harms, with particular attention to disproportionate impacts on protected groups; documentation of the technical methodology, training data, and validation processes; specification of the human oversight mechanisms that will operate alongside the AI system; and public disclosure of a non-technical plain-language summary. AIAs would be reviewed by an independent Algorithmic Oversight Authority (AOA) — a new regulatory body proposed in this framework — before system deployment.

A.2 Designated Algorithmic Accountability Officers (AAOs)

Every government ministry or department deploying a high-risk or critical AI system must designate an Algorithmic Accountability Officer — a senior official with legal training or governance expertise who

bears personal accountability for the system's compliance with AIPA requirements. The AAO would be responsible for: overseeing AIA processes; receiving citizen complaints about algorithmic decisions; commissioning and acting on algorithmic audit findings; and reporting annually to Parliament on the system's performance and governance. Personal accountability of a named official is critical to closing the responsibility diffusion gap: it creates a single identifiable accountable actor where currently there is only distributed organizational responsibility.

A.3 Liability Framework for Algorithmic Harm

The current Information Technology Act 2000 and Government Liability frameworks are inadequate for algorithmic harm. AIPA proposes amendments to establish: strict liability of the deploying agency for demonstrable harm caused by high-risk or critical AI systems — removing the requirement to prove negligence where harm results from algorithmic error; a Government Algorithmic Harm Compensation Scheme (GAHCS) providing an accessible, low-cost compensation mechanism for citizens harmed by government AI decisions, without requiring individual litigation; and developer liability provisions where private vendors supply AI systems to government that are found to be fundamentally defective, proportionate to the vendor's knowledge of defects at time of supply.

8.2 Pillar I — Integrity: Ensuring Algorithmically Sound Decisions

I.1 Explainability Standards

AIPA mandates explainability standards calibrated to two audiences. First, citizen-facing explanations: any government AI decision affecting an individual's rights or entitlements must be accompanied by a plain-language explanation of the primary factors that determined the decision — expressed in the individual's language of correspondence — that a person without technical knowledge can understand and use to assess whether the decision is correct. Second, technical explanations: for any challenge or review of an AI decision, the reviewing body must receive a complete technical account of the algorithmic methodology, training data, validation results, and the specific inputs that produced the contested decision. The precedent for this standard is set by the EU AI Act's Articles 13 and 50, and GDPR's Article 22 right to explanation.

I.2 Mandatory Bias Auditing

All high-risk and critical government AI systems must undergo independent bias auditing: (a) pre-deployment, before the system is used for actual decisions; (b) annually during operation; and (c) following any significant change in the system's training data or methodology. Bias audits must assess demographic disparities in outcomes across gender, caste category, religion, disability status, and geographic location — the protected characteristics most relevant to the Indian administrative context. Audit results must be published publicly, enabling civil society and parliamentary scrutiny. AI systems found to produce statistically significant unjustified disparate impacts must be suspended pending remediation or, if remediation is not possible within six months, decommissioned.

I.3 Genuine Human Oversight Requirements

AIPA addresses the 'meaningless human review' gap through structural requirements rather than nominal ones. For critical AI systems (welfare entitlement decisions, bail risk assessments), AI output must be treated as one input among several in a genuine human deliberative process — not as a presumptive conclusion. Case-per-officer maximum thresholds must be specified in operational protocols, ensuring that review caseloads do not structurally preclude meaningful individual consideration. 'Override rates' — the proportion of AI recommendations that human reviewers modify or reject — must be monitored and reported: a system where human reviewers never override AI

recommendations indicates de facto automated decision making regardless of nominal human involvement.

8.3 Pillar P — Participation: Democratizing Algorithmic Governance

P.1 Parliamentary Pre-Deployment Scrutiny

Critical and high-risk government AI systems should require parliamentary approval before deployment — analogous to the scrutiny applied to significant procurement decisions. A standing Parliamentary Committee on Algorithmic Governance (PCAG), modelled on the Public Accounts Committee, would receive AIAs, scrutinize system design and procurement processes, and provide pre-deployment recommendations. This mechanism directly addresses the democratic deficit gap, ensuring that consequential algorithmic design choices — including the fairness criteria adopted for systems like bail risk assessment — are subject to democratic deliberation rather than private vendor discretion.

P.2 Citizen Contestation Rights

AIPA establishes a statutory right for any citizen adversely affected by a government AI decision to: receive a meaningful explanation of the decision (per Pillar I); challenge the decision through an accessible administrative review process with a maximum resolution time of 30 days; have their case reviewed by a human decision-maker not involved in the original AI process; and receive compensation if the AI decision is found to have been erroneous and caused demonstrable harm. The contestation pathway must be accessible through multiple channels — in-person, online, and through legal representatives — and must be available in regional languages.

P.3 Civil Society Algorithmic Observatory

AIPA proposes the creation of an independent Civil Society Algorithmic Observatory (CSAO) — a multi-stakeholder body comprising academic researchers, civil society organizations, legal aid providers, and citizen representatives — with statutory rights to: access technical documentation of high-risk government AI systems; conduct independent bias auditing; publish findings without government editorial control; and recommend corrective actions to the Algorithmic Oversight Authority. The CSAO model draws on the precedent of independent audit institutions (CAG) and information commissions (CIC), adapting this model to the specific requirements of algorithmic governance oversight.

8.4 Pillar Au — Auditability: Creating Verifiable Accountability

Au.1 Algorithmic Oversight Authority (AOA)

The most significant institutional innovation proposed in AIPA is the creation of an independent Algorithmic Oversight Authority — a statutory body with regulatory authority over government AI deployments in high-risk public administration domains. The AOA would exercise: pre-deployment approval authority for critical AI systems; ongoing monitoring and enforcement powers; investigation authority for complaints and incident reports; the power to order system suspension, modification, or decommissioning; and the authority to impose financial penalties on deploying agencies for AIPA non-compliance. The AOA would be constituted with technical expertise (AI/ML specialists), legal expertise (administrative and constitutional lawyers), and domain expertise (welfare, tax, criminal justice specialists), with members appointed through a transparent, merit-based process involving parliamentary confirmation.

Au.2 Algorithmic Audit Trail Requirements

All government AI systems subject to AIPA must maintain comprehensive, immutable audit trails recording: every decision or recommendation produced by the system; the input data that generated each decision; the model version in use at the time of each decision; and any human modifications to AI recommendations. Audit trails must be retained for a minimum of 10 years — the standard limitation

period for administrative action challenges. These trails must be accessible to the AOA, designated auditors, Parliamentary Committees, and (in anonymized or relevant form) citizens exercising contestation rights. The audit trail requirement is the operational foundation of all other AIPA accountability mechanisms: without reliable records of what decisions were made and why, no form of accountability is meaningful.

8.5 AIPA Framework Summary Table

	Pillar	Key Mechanisms	Institutional Locus	Governance Gap Addressed
A	Accountability	AIAs; AAOs; Liability Framework; GAHCS	Deploying agencies; Courts; Parliament	Responsibility diffusion; Administrative law inadequacy
I	Integrity	Explainability standards; Bias auditing; Genuine human oversight	Deploying agencies; AOA; Independent auditors	Opacity; Bias amplification; Meaningless human review
P	Participation	Parliamentary scrutiny; Citizen contestation; CSAO	Parliament; PCAG; Civil society; Citizens	Democratic deficit; Contestation barrier
Au	Auditability	AOA; Audit trails; Compliance monitoring	Algorithmic Oversight Authority	All seven gaps (systemic enforcement)

Table 3: AIPA Governance Framework — Four Pillars Summary

8.6 Risk-Tiered Application of AIPA Requirements

Risk Tier	Examples	AIPA Requirements	Review Mechanism
Critical	Bail/sentencing AI; welfare exclusion decisions; tax reassessment; deportation	Full AIPA: AIA + Parliamentary approval + AAO + bias audit + strict liability + citizen contestation + AOA oversight	AOA pre-deployment approval; Parliamentary PCAG scrutiny; Judicial review; GAHCS compensation
High	DBT eligibility scoring; tax audit selection (INSIGHT); employment screening; licensing algorithms	AIA + AAO designation + annual bias audit + explainability + contestation right	AOA registration; annual compliance report; Administrative tribunal review

Risk Tier	Examples	AIPA Requirements	Review Mechanism
Moderate	Internal workflow automation; document routing; queue management; scheduling	AIA (light) + bias monitoring + audit trail + internal oversight officer	Departmental review; CSAO monitoring access
Low	Chatbots; document translation; data visualization; non-decision analytics	Transparency disclosure + audit trail	Annual departmental self-assessment

Table 4: AIPA Risk-Tiered Application Matrix

9. CONSTITUTIONAL ADEQUACY ANALYSIS: INDIA'S EXISTING FRAMEWORK

A critical question for AIPA's legal design is whether India's existing constitutional and administrative law framework provides adequate governance of AI-driven public administration, or whether legislative intervention is necessary. This section conducts a systematic assessment.

9.1 Article 14: Arbitrariness Review

Article 14's prohibition on arbitrary state action — as developed through the doctrine of reasonableness in cases including E.P. Royappa v. State of Tamil Nadu (1974) and Maneka Gandhi v. Union of India (1978) — provides a foundation for challenging AI-driven decisions that produce arbitrary outcomes. However, the doctrine requires identification of specific arbitrariness in individual decisions, which is structurally difficult when the source of arbitrary outcomes is a statistical pattern embedded in an algorithmic system rather than a discrete decision-making act. Courts applying Article 14 review to AI-generated decisions would need to develop new analytical tools for evaluating systemic arbitrariness — discriminatory patterns revealed only through statistical analysis across thousands of decisions — rather than the traditional individual case analysis. This evolution is possible within existing constitutional doctrine but has not yet occurred in India's jurisprudence.

9.2 Article 21: Due Process and Reasoned Decisions

Article 21's protection of life and personal liberty, as interpreted in the context of administrative due process, generates several requirements directly relevant to AI-driven decisions: reasoned decisions (the 'speaking order' requirement established in S.N. Mukherjee v. Union of India, 1990); notice and opportunity to be heard; and the right to challenge decisions through judicial review. The speaking order requirement is potentially the most significant existing constraint on AI-driven decisions: a decision generated by an algorithm that cannot articulate its reasoning in terms a court can evaluate arguably fails the speaking order standard. However, courts have not yet clearly ruled on whether algorithmic outputs can satisfy this requirement, and the practical challenge of applying it at the scale of INSIGHT's operations has not been resolved.

9.3 The Puttaswamy Privacy Framework

The Supreme Court's nine-judge bench decision in K.S. Puttaswamy v. Union of India (2017) established a constitutional right to privacy with direct implications for AI governance: the Court held that informational privacy — the right to control information about oneself — is a fundamental right; that

data collection and use by government must be proportionate to a legitimate aim; and that the 'just, fair, and reasonable' standard of Article 21 applies to information processing. Applied to AI-driven public administration, the Puttaswamy framework requires that AI systems collecting and processing citizens' personal data must satisfy proportionality — algorithmic profiling for welfare or tax purposes must be proportionate to the legitimate administrative goal, not merely convenient or efficient. The Court's framework thus already provides constitutional traction for challenging AI deployments that involve disproportionate data collection or use.

9.4 Legislative Gap: The Case for AIPA as Statute

The foregoing analysis reveals that while India's constitutional framework provides significant, if underexploited, resources for challenging specific harmful AI decisions, it does not provide a proactive governance framework capable of: setting pre-deployment standards that prevent harmful systems from being deployed; providing accessible remedies for citizens who lack resources to pursue constitutional litigation; ensuring systematic bias monitoring and audit across AI deployments; or establishing the institutional infrastructure (AOA, CSAO, PCAG) necessary for ongoing AI governance. These gaps require legislative intervention. This paper recommends that the Government of India enact an Algorithmic Accountability and AI Governance Act, operationalising the AIPA Framework within a statutory mandate, providing the AOA with formal regulatory authority, and establishing the citizen rights and agency obligations described in this framework.

10. DISCUSSION

The comparative, doctrinal, and case study analysis conducted in this paper converges on a single overarching conclusion: India is deploying AI in public administration at a scale and pace that substantially exceeds its governance capacity to ensure that such deployment serves rather than harms its citizens. This is not a counsel against AI in public administration — the efficiency, consistency, and anti-corruption benefits of well-governed algorithmic systems are real and important. It is a counsel for urgent, comprehensive, and well-designed governance development.

The AIPA Framework proposed in this paper represents a deliberate calibration between two failure modes in AI governance: the under-governance failure mode, exemplified by India's current situation and the US experience with COMPAS, where AI deployment without adequate accountability produces discriminatory and arbitrary outcomes at scale; and the over-governance failure mode, where excessive regulatory burden prevents beneficial AI adoption and imposes compliance costs that resource-constrained agencies cannot bear. AIPA's risk-tiered approach is designed to avoid the second failure mode by calibrating requirements to actual risk: low-risk administrative tools face minimal requirements, while critical systems affecting fundamental rights face stringent but proportionate governance obligations.

A particularly important insight from the comparative analysis is that the governance challenge of AI in public administration is fundamentally a political economy problem, not merely a technical one. AI systems that make or influence welfare, tax, and criminal justice decisions carry enormous implications for the distribution of state power and its exercise over citizens. The accountability mechanisms proposed in AIPA — parliamentary scrutiny, civil society oversight, independent regulatory authority — are designed to rebalance that power distribution: to ensure that as AI systems acquire more influence over public decisions, citizens retain meaningful ability to understand, challenge, and shape how those decisions are made.

The Indian constitutional context provides both opportunities and challenges for AIPA implementation. India's strong fundamental rights tradition and active Supreme Court create a supportive environment for rights-based AI governance. The potential application of Article 14 arbitrariness review to systematic algorithmic bias, and Article 21 due process requirements to AI-generated decisions, means that courts could provide important governance contributions even before legislative action. However, the scale of India's AI deployments, the resource constraints of affected populations, and the technical complexity of algorithmic challenges mean that judicial review alone is entirely insufficient as a governance mechanism. The AOA, GAHCS, and citizen contestation infrastructure proposed in AIPA are essential complements to existing judicial remedies.

11. POLICY RECOMMENDATIONS

11.1 For the Government of India (MeitY and Ministry of Law)

5. Enact an Algorithmic Accountability and AI Governance Act within 24 months, operationalizing the AIPA Framework with statutory force and establishing the Algorithmic Oversight Authority as a regulatory body with independent constitutional status analogous to the Election Commission.
6. Immediately suspend any AI system currently making or automatically triggering Critical-tier administrative decisions without an Algorithmic Impact Assessment on file with MeitY, pending retrospective AIA completion and AOA provisional approval.
7. Commission a national audit of INSIGHT's automated notice generation processes, examining compliance with speaking order requirements and assessing the feasibility of providing algorithmic explanations in INSIGHT-generated communications.

11.2 For Parliament

8. Establish a Standing Committee on Algorithmic Governance (PCAG) with a mandate to scrutinize all Critical and High-tier AI deployments in public administration, drawing on the model of the UK's Science and Technology Committee's AI scrutiny work.
9. Amend the Digital Personal Data Protection Act 2023 to include explicit provisions on automated decision making in public administration, aligned with GDPR Article 22 standards — providing citizens with rights of explanation, contestation, and human review for consequential government AI decisions.

11.3 For the Supreme Court of India

10. Issue Practice Directions requiring that algorithmic risk assessments or AI-generated evidence used in bail or sentencing proceedings be accompanied by a full technical disclosure statement provided to the defence, and that the methodology be subject to cross-examination before being relied upon.
11. Initiate suo motu proceedings to examine the constitutional adequacy of the INSIGHT platform's automated notice generation, specifically addressing the speaking order requirements of Articles 14 and 21 as applied to algorithmic administrative action.

11.4 For the National Informatics Centre and AI Deploying Agencies

12. Adopt and publish Algorithmic Impact Assessments for all existing high-risk and critical government AI systems within 18 months, using the AIPA AIA template, making these publicly available on agency websites and in the national AI transparency register.
13. Establish formal override rate monitoring for all human review processes operating alongside government AI systems, reporting publicly on the proportion of AI recommendations that human reviewers modify or reject, as a key indicator of genuine vs. nominal human oversight.

12. CONCLUSION

This paper has conducted a comprehensive analysis of the ethical and legal accountability challenges generated by AI-driven decision making in public administration, drawing on case studies from three high-stakes domains — welfare allocation, tax assessment, and bail decisions — and comparative analysis of governance frameworks across nine jurisdictions. The analysis reveals seven structural accountability gaps — opacity, responsibility diffusion, administrative law inadequacy, bias amplification, democratic deficit, meaningless human review, and contestation barriers — that characterize current AI deployments in public administration globally and with particular acuity in India.

The AIPA Governance Framework proposed in this paper — organized around Accountability, Integrity, Participation, and Auditability — provides a comprehensive, risk-tiered, constitutionally-grounded legal-ethical architecture for governing AI in Indian public administration. Drawing on best practices from the EU, Canada, and Australia's post-RoboDebt reforms while calibrating to India's institutional capacity and constitutional context, AIPA represents an original contribution to AI governance scholarship with direct policy relevance.

The stakes of this governance challenge cannot be overstated. India's AI-driven administrative systems — INSIGHT, DBT, Aadhaar authentication, and future judicial risk assessment tools — collectively touch the lives of hundreds of millions of citizens at the most consequential interfaces of state power: the determination of whether they receive welfare, whether they owe taxes, whether they spend time in prison. Getting AI governance right in these domains is not a technical nicety; it is a prerequisite for the legitimacy and justice of governance itself in the AI era.

The lesson of RoboDebt, COMPAS, and India's own documented welfare exclusion cases is that AI systems deployed without adequate governance do not merely fail technically; they fail systematically, discriminatorily, and at a scale that individual redress mechanisms cannot address. India has the opportunity — and, given its constitutional commitments to fundamental rights, the obligation — to develop governance frameworks that harness AI's genuine governance benefits while protecting citizens from its documented harms. The AIPA Framework provides the architecture for that governance. The question is one of political will.

ACKNOWLEDGEMENTS

[Authors acknowledge the contributions of legal colleagues, policy researchers, and reviewers. Funding acknowledgement to be added. This research reflects solely the views of the authors and does not represent the position of any institution.] The authors declare no competing interests.

DECLARATION OF COMPETING INTERESTS

The authors declare no conflict of interest. This research received no funding from AI vendors, government agencies, or any party with a direct interest in AI regulation outcomes. This manuscript is original, has not been published previously, and is not under review elsewhere.

REFERENCES

Alston, P. (2019). *Report of the special rapporteur on extreme poverty and human rights: Digital technology, social protection and human rights.* United Nations General Assembly Doc. A/74/493.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica.* https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bannister, F., & Connolly, R. (2020). The great theory hunt: Does e-government really have a problem? *Government Information Quarterly, 37*(2), 101407.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code.* Polity Press.

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal, 13*(4), 447–468.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data, 5*(2), 153–163. https://doi.org/10.1089/big.2016.0047

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., ... & Shankar, K. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society, 4*(2), 1–21.

DBT Mission. (2024). *DBT India — Direct Benefit Transfer statistics 2023–24.* Ministry of Finance, Government of India. https://dbtbharat.gov.in

Deloitte Center for Government Insights. (2025). *Government trends 2026: The future of government is now.* Deloitte Insights. https://www.deloitte.com/us/en/insights/industry/government-public-services/government-trends.html

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM, 59*(2), 56–62.

Drèze, J., & Khera, R. (2022). Aadhaar and welfare exclusion: Evidence and policy implications. *Economic and Political Weekly, 57*(18), 22–31.

Dwyer, P. (2019). Dealing with welfare conditionality: Implementation and effects. *Journal of Social Policy, 48*(2), 289–307.

Eggers, W. D., & Bellman, J. (2015). *The journey to government's digital transformation.* Deloitte University Press.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Official Journal of the European Union.

Government of Canada — Treasury Board Secretariat. (2019). *Directive on automated decision-making.* Treasury Board of Canada Secretariat. https://www.tbs-sct.gc.ca

Holmes, C. (2023). *Royal Commission into the Robodebt Scheme: Final report.* Commonwealth of Australia.

Janssen, M., Hartog, M., Matheus, R., Yi Ding, A., & Maheshwari, G. (2022). Will algorithmic governance change the nature of governance? *Information Polity, 27*(3), 1–14.

Kant, I. (1785/2012). *Groundwork of the metaphysics of morals* (M. Gregor & J. Timmermann, Trans.). Cambridge University Press.

Khera, R. (2019). *Dissent on Aadhaar: Big data meets big brother.* Orient Blackswan.

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv:1609.05807.* https://arxiv.org/abs/1609.05807

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21. https://doi.org/10.1177/2053951716679679

Mulgan, R. (2003). *Holding power to account: Accountability in modern democracies.* Palgrave Macmillan.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* New York University Press.

Northpointe. (2016). *Response to ProPublica: Demonstrating accuracy equity and predictive parity.* Northpointe Inc.

Ramanathan, U. (2014). A unique identity bill. *Economic and Political Weekly, 45*(30), 10–14.

Rawls, J. (1971). *A theory of justice.* Harvard University Press.

Sengupta, I. (2022). INSIGHT and the algorithmic turn in Indian tax administration: Due process implications. *National Law School of India Review, 34*(1), 88–112.

Supreme Court of India. (2017). *K.S. Puttaswamy v. Union of India (Privacy Judgment), (2017) 10 SCC 1.* Supreme Court of India.

Supreme Court of India. (2018). *K.S. Puttaswamy v. Union of India (Aadhaar Judgment), (2018) 1 SCC 809.* Supreme Court of India.

Veale, M., & Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. In K. Yeung & M. Lodge (Eds.), *Algorithmic regulation* (pp. 121–149). Oxford University Press.

Wiseman, S. (2018). Fixing sentencing: The case against algorithmic risk assessment. *University of Florida Law Review, 70*(4), 1153–1200.

World Economic Forum. (2024). *Responsible use of technology in public administration: A governance toolkit.* World Economic Forum. https://www.weforum.org

APPENDIX A: AIPA FRAMEWORK — LEGISLATIVE PROVISIONS SUMMARY

Cl.	Provision Title	Core Requirement	Indian Law Parallel
1	Algorithmic Impact Assessment	Mandatory pre-deployment AIA for high/critical AI; AOA review	DPDPA DPIA concept; EIA under Environment Act
2	Explainability Duty	Plain-language explanations for all AI-influenced citizen decisions	Speaking order doctrine (S.N. Mukherjee); RTI Act
3	Bias Audit Obligation	Annual independent bias audit; public results; suspension powers	CAG audit model; SEBI audit requirements
4	Genuine Human Oversight	Case-per-officer limits; override rate monitoring; non-rubber-stamp requirement	Quasi-judicial decision requirements; CPC Order
5	Citizen Contestation Right	30-day administrative review; human reviewer; compensation pathway	Article 21 due process; Administrative Tribunals Act

Cl.	Provision Title	Core Requirement	Indian Law Parallel
6	Algorithmic Oversight Authority	Independent regulatory body; pre-deployment approval; enforcement powers	CCI model; TRAI model; Election Commission status
7	Liability for Algorithmic Harm	Strict liability (critical tier); GAHCS compensation scheme; vendor liability	Consumer Protection Act; Tort law; ITA liability
8	Audit Trail Mandate	10-year immutable records; AOA access; citizen access in contestation	Records Management under GFR; RTI access
9	Parliamentary Oversight	PCAG scrutiny; pre-deployment approval for critical tier	PAC model; DRSCs; Parliamentary Questions
10	Civil Society Observatory	CSAO statutory establishment; access rights; independent publication	Press Council; NHRC; Information Commission models

Table A1: AIPA Framework — Proposed Legislative Provisions with Indian Law Parallels