ALGORITHMIC GOVERNANCE AND THE RIGHT TO EXPLANATION: BALANCING AI-DRIVEN DECISION-MAKING WITH ADMINISTRATIVE JUSTICE

Rajput Mounica Bai, Sri Prasunna College of Law

ABSTRACT

As the contemporary administrative state increasingly delegates welfare distribution, resource allocation, and risk profiling to Artificial Intelligence and Automated Decision-Making (AI-ADM) systems, public law faces an unprecedented legitimacy crisis.[1] Traditional administrative law is anchored in the twin pillars of natural justice and the duty to give reasons doctrines designed to ensure that state actions are transparent, rational, and contestable. However, the proprietary and complex "black box" nature of machine learning models creates a structural 'rationality gap,' rendering automated decisions opaque to both the citizen and the judiciary. This paper examines the constitutional friction generated when opaque algorithmic classifications undermine the guarantees of equality and due process. Moving beyond technical definitions of model interpretability, the paper argues for a robust administrative doctrine of 'algorithmic legibility,' bridging the gap between technological efficiency and constitutional accountability by embedding mandatory explainability and institutional contestability into India's digital governance framework.

Keywords: Algorithmic Governance, Right to Explanation, Administrative Justice, Artificial Intelligence and Automated Decision-Making (AI-ADM), Duty to Give Reasons, Constitutional Due Process.

[1] See Danielle Keats Citron, Technological Due Process, 85 Wash. U. L. Rev. 1249, 1251–55 (2008).

Introduction

The contemporary administrative state is undergoing a quiet, structural metamorphosis. Across jurisdictions, the traditional machinery of governance historically defined by human discretion, bureaucratic deliberation, and written ministerial files is increasingly being augmented, and in many instances supplanted, by Artificial Intelligence and Automated Decision-Making (AI-ADM) systems. From welfare distribution, credit scoring, and predictive policing to tax compliance and immigration screening, algorithms now actively shape, filter, and dictate the allocation of public resources and individual rights. This transition promises unprecedented institutional efficiency, speed, and cost-reduction. Yet, it simultaneously introduces a profound jurisprudential crisis: a fundamental collision between the relentless momentum of digital governance and the foundational tenets of public law.

At the heart of this friction lies the "Black Box" dilemma.[2] Modern machine learning models particularly deep neural networks and complex ensemble systems prioritize predictive accuracy over structural transparency. Their outputs are generated through non-linear mathematical correlations across multidimensional spaces, making it notoriously difficult, and at times technically impossible, even for their human designers to trace how a specific decision was reached. When an algorithm denies an individual a social welfare grant, flags a citizen as a high-risk security threat, or assigns an adverse tax assessment, the underlying rationale remains concealed within proprietary code and complex weights.

This opacity creates an acute structural conflict with traditional administrative jurisprudence, which is anchored in the twin pillars of Natural Justice (audi alteram partem) and the Duty to Give Reasons.[3] Public law has long maintained that arbitrary state action is anathema to the rule of law, requiring that any decision affecting a citizen's rights must be transparent, rational, and subject to judicial review. However, a citizen cannot effectively challenge an administrative decision they cannot comprehend, nor can a court review the rationality of a computational output that cannot be explained. When human accountability is outsourced to an inscrutable algorithm, the constitutional guarantee of administrative justice is effectively

[2] See generally Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information (2015).

[3] See Mohinder Singh Gill v. Chief Election Comm'r, (1978) 1 SCC 405 (India) (emphasizing that public orders must be judged by the reasons explicitly mentioned therein); see also Wade & Forsyth, Administrative Law 259–63 (11th ed. 2014).

reduced to a procedural echo.

This paper examines the friction generated when opaque algorithmic classifications undermine constitutional standards of equality and due process. Moving beyond purely technical definitions of model interpretability, the study argues that modern administrative law must evolve beyond human-centric assumptions to construct a robust doctrine of "Algorithmic Legibility." By exploring how traditional public law doctrines can be re-engineered, this paper contends that the codification of a judicially enforceable Right to Explanation is not merely a technical regulatory preference, but an indispensable constitutional prerequisite for the legitimacy of India's digital administrative state.

The Intersection: Public Administration Meets the "Black Box"

Deconstructing AI-ADM in State Action

To understand the constitutional friction of automated governance, one must first examine how AI-ADM systems are integrated into the modern administrative apparatus.[4] Unlike traditional computer systems that operate on explicit, rule-based programming (where a human administrator dictates if-then conditions), modern AI relies on machine learning (ML). These systems ingest vast historical datasets, identify statistical patterns, and autonomously generate rules to predict future outcomes or classify individuals.

In the realm of public administration, these tools manifest in diverse forms:

Predictive Risk Scoring: Used in criminal justice and social services to assess recidivism risk or child welfare vulnerabilities.

Automated Eligibility Engines: Deployed to process applications for unemployment benefits, housing assistance, or disability grants.

Automated Surveillance and Profiling: Utilizing facial recognition, automated traffic enforcement, and tax-evasion algorithms that trigger direct enforcement actions without prior human review.

[4] See Cary Coglianese & David Lehr, Regulating by Robot: Administrative Decision Making in the Machine-Learning Era, 105 Geo. L.J. 1147, 1153–60 (2017).

While these systems are framed as neutral arbiters designed to eliminate human bias and administrative backlog, they fundamentally alter the nature of administrative discretion. Discretion is no longer exercised by a public servant who weighs equities, listens to grievances, and applies statutory intent; instead, it is pre-packaged into statistical probabilities calculated by code.

The Rationality Gap and the Inherent Opacity of ML

The core legal challenge arises from what scholars term the "Rationality Gap" the fundamental mismatch between the demands of public law and the operational reality of machine learning models.

Epistemic Opacity: Deep learning models, gradient boosting algorithms, and neural networks utilize millions of parameters across hidden layers. Even when the source code and training data are open, the complex, non-linear interactions make it virtually impossible to isolate the precise weight or factor that tipped the scale against a specific individual.

Proprietary Secrecy: Governments frequently procure these AI systems from private technology vendors. Vendors routinely shield their algorithms behind intellectual property laws and trade secret protections, arguing that exposing the inner workings of their code compromises market competitiveness.

The Inductive Fallacy of Statistics: Administrative law evaluates decisions based on individual merit, statutory criteria, and direct causal nexus. Machine learning, by contrast, operates on correlation rather than causation. It groups individuals into risk categories based on population-level historical proxies (such as zip code, socio-economic background, or digital footprints), which often encode historical systemic biases.

When an administrative decision is grounded in a statistical correlation that cannot be unpacked, the traditional chain of administrative reasoning breaks down.[5] The administrator who signs off on the automated output cannot explain it, the aggrieved citizen cannot challenge its factual premises, and the reviewing court is left with nothing tangible to scrutinize. This transforms administrative discretion from a transparent, accountable power into an opaque

[5] See Jenna Burrell, How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, 3 Big Data & Soc'y 1, 2–4 (2016).

exercise of computational fiat.

Constitutional and Administrative Imperatives

The Duty to Give Reasons as a Pillar of Natural Justice

In administrative jurisprudence, the requirement to provide reasoned decisions is widely recognized as an indispensable safeguard against arbitrary state action. The principles of natural justice specifically audi alteram partem (the right to a fair hearing) are rendered hollow if an individual cannot ascertain why an adverse decision was rendered against them.

Indian administrative law has steadily evolved to recognize the duty to give reasons as an integral component of procedural fairness. Landmark jurisprudence has established that transparency is the hallmark of democratic governance, and administrative orders must reflect a clear application of mind to relevant considerations.

In foundational rulings such as A.K. Kraipak v. Union of India and subsequent expansions of administrative accountability, the Supreme Court emphasized that the dividing line between administrative and quasi-judicial powers is increasingly thin when fundamental rights are impacted.[6]

Similarly, the modern administrative state cannot shield arbitrary or unreasoned choices behind the cloak of institutional autonomy or bureaucratic convenience.

When applied to AI-ADM systems, the traditional duty to give reasons faces an existential crisis. If an administrative body adopts an automated output without possessing or providing the logical chain that generated it, the "reasons" provided to an aggrieved citizen are often generic boilerplate justifications (e.g., "system-generated risk score exceeded threshold"). This strips the right to challenge the decision of any meaningful substance, reducing procedural fairness to a digital formality.

Constitutional Standards and Fundamental Rights

Beyond administrative law principles, opaque algorithmic governance directly engages the text

[6] see also Maneka Gandhi v. Union of India, (1978) 1 SCC 248 (India) (establishing that procedure established by law under Article 21 must be right, just, and fair, and not arbitrary).

and spirit of Part III of the Constitution of India, primarily through the lenses of Article 14 and Article 21.

1. Article 14:[7] Equality, Non-Arbitrariness, and Algorithmic Classification[8]

Article 14 guarantees equality before the law and equal protection of the laws, striking down state action that is arbitrary, capricious, or lacking an intelligible differentia with a rational nexus to the object sought to be achieved.

AI-ADM systems classify citizens by sorting them into risk tiers or eligibility buckets based on historical data proxies.

Because machine learning models learn from past data, they frequently replicate, amplify, and institutionalize historical systemic biases (e.g., socio-economic or geographic disparities).

When an automated system denies benefits or imposes burdens based on opaque statistical correlations that cannot be inspected for bias, the classification becomes inherently arbitrary. It treats individuals not as unique legal subjects possessing independent agency, but as statistical remnants of historical prejudices, failing the rigorous test of non-arbitrariness established in Maneka Gandhi v. Union of India and K.S. Puttaswamy v. Union of India.

2. Article 21: Right to Life, Dignity, and Due Process

[9]The expansion of Article 21 to encompass substantive due process and human dignity provides a powerful anchor against algorithmic dehumanization.

State action that fundamentally affects a person's livelihood, personal liberty, or access to essential welfare must satisfy the threshold of fairness, justice, and reasonableness.

Delegating life-altering decisions such as access to healthcare, social security, or freedom from unwarranted surveillance to an inscrutable machine learning model reduces human dignity to a computational output.

---

[7] See K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1, 533 (India)

[8] see also Cathy O'Neil, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy 13–24 (2016).

[9] See Navtej Singh Johar v. Union of India, (2018) 1 SCC 791, 856 (India) (reaffirming that constitutional morality requires the state to respect individual autonomy and dignity).

A citizen subjected to an unexplained algorithmic rejection suffers a profound affront to their autonomy: they are judged by a process they cannot question, governed by logic they cannot inspect, and denied recourse to a human mind capable of exercising empathy, context, and discretion.

Comparative Frameworks & Their Shortcomings

Global Models and Regulatory Benchmarks

As jurisdictions worldwide grapple with the integration of AI-ADM in public administration, regulatory models have emerged attempting to balance technological innovation with citizen rights.

The European Union Model: The European Union has led legislative efforts through the General Data Protection Regulation (GDPR) and the subsequent EU Artificial Intelligence Act. Article 22 of the GDPR provides data subjects with the right not to be subject to decisions based solely on automated processing including profiling that produce legal effects or similarly significant effects.[10] Complementary recitals and provisions hint at a "right to explanation," requiring data controllers to provide meaningful information about the logic involved in automated processing.

The Limitations of Textual Convergence: While these international frameworks offer vital blueprints, comparative legal analysis reveals a structural disconnect. Many jurisdictions treat transparency as a consumer-protection or data-privacy metric rather than a core public law mandate. Consequently, rules designed primarily for commercial data processing often fail to capture the coercive power of the state when deploying automated systems for welfare distribution, policing, or resource allocations.

The "Explainability Trap"

A critical flaw in current regulatory discourse is the conflation of technical interpretability with legal contestability an oversight known as the "Explainability Trap."[11]

[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), art. 22, 2016 O.J. (L 119) 1.

[11] See Sandra Wachter, Brent Mittelstadt & Luciano Floridi, Why a Right to Explanation of Automated

The Fallacy of Technical Transparency: Policymakers frequently assume that forcing vendors or state agencies to disclose code, training datasets, or mathematical feature-importance scores (such as LIME or SHAP values) constitutes an adequate "explanation." However, raw algorithmic metrics are entirely incomprehensible to the average citizen and offer little utility to a court evaluating administrative fairness. An individual told that their welfare score failed due to a "negative coefficient in hidden layer three" is no closer to meaningful justice than one given no explanation at all.

Separating Comprehension from Contestability: As administrative law scholars emphasize, understanding how a machine learning model arrived at an output does not establish whether that output is legally or morally just. For instance, if an automated system accurately applies a flawed or discriminatory proxy (such as penalizing applicants based on historical postal-code default rates), technical transparency merely clarifies the mechanism of injustice without providing a remedy.

The Vendor Secrecy Barrier: Public authorities often procure proprietary software from private vendors who claim trade secret protection over their algorithms. When challenged in court, agencies hide behind intellectual property exemptions, rendering judicial review toothless.

True administrative justice requires shifting the focus from passive technical transparency to active institutional contestability. An effective legal framework must ensure that an explanation functions not merely as a descriptive technical manual, but as an actionable evidentiary key that allows a citizen to challenge, interrogate, and overturn an arbitrary state action.

Reimagining Administrative Law: Towards a Doctrine of Algorithmic Legibility

To bridge the chasm between technological efficiency and constitutional governance, administrative law must transcend the limits of traditional human-centric design. It requires the codification of a normative framework termed "Algorithmic Legibility." Rather than treating transparency as a mechanical software disclosure, this doctrine reimagines accountability across three institutional tiers:

Decision-Making Does Not Exist in the General Data Protection Regulation, 7 Int'l Data Privacy L. 76, 79–85 (2017) (analyzing the gap between technical transparency and legal contestability).

1. From Passive Transparency to Active Contestability

Courts and legislatures must decouple technical interpretability from legal contestability. A citizen does not need access to source code or raw tensor weights to contest a decision; they require an actionable rationale.

Counterfactual Explanations: Administrative agencies deploying AI-ADM should be legally mandated to provide counterfactual statements[12] (e.g., "Your application was denied under threshold X; had factor Y been different by Z amount, the outcome would have been positive"). This translates opaque machine learning logic into practical criteria that a citizen can verify, dispute, or correct.

2. Institutional Human-in-the-Loop Safeguards

Automated systems should never possess final, unmediated authority over fundamental rights or resource distributions. Administrative law must enforce a robust "Human-in-the-Loop" standard, requiring that:

Every adverse automated decision includes an automatic right to an independent review by a human administrator.

Reviewing officers must be trained to critically evaluate algorithmic outputs rather than rubber-stamping computer-generated scores out of institutional deference to technology.

3. Piercing the Veil of Vendor Secrecy

Private tech vendors contracting with the state cannot hide behind intellectual property and trade secret exemptions to evade public accountability.[13] Administrative procurement law must mandate that public-sector AI contracts include a waiver of proprietary opacity. If a private entity wishes to supply predictive governance tools to the government, open auditability and compliance with public law standards must be treated as non-negotiable statutory prerequisites.

[12] See Sandra Wachter, Brent Mittelstadt & Chris Russell, Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, 31 Harv. J.L. & Tech. 841, 845–50 (2018).
[13] See Rebecca Wexler, Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System, 70 Stan. L. Rev. 1343, 1355–70 (2018)

Conclusion

The digital transformation of the administrative state is irreversible. As algorithms increasingly dictate the distribution of welfare, liberty, and public resources, the legal system faces a defining test: whether constitutional governance will master technology, or technology will quietly erode the rule of law.

The traditional "black box" nature of machine learning creates an acute rationality gap that threatens to reduce the core tenets of natural justice the duty to give reasons, non-arbitrariness, and due process to empty digital formalities. However, technology should not be permitted to outpace constitutionalism. By moving beyond the superficial "explainability trap" and embedding a doctrine of Algorithmic Legibility into administrative frameworks, the state can preserve its technological gains without sacrificing public accountability. Ultimately, codifying a judicially enforceable Right to Explanation is not merely an administrative upgrade; it is an essential safeguard to ensure that in the digital age, human dignity remains superior to algorithmic fiat.