© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717917

Algorithmic Bias and the Regulatory Void: Rethinking India's Legal Framework for Artificial Intelligence Governance

ANAM SHAMS
Shibli National College, Azamgarh, U. P LL. M (Master of Laws)

Abstract- ***The integration of Artificial Intelligence into India's governance and legal infrastructure has accelerated dramatically, yet the constitutional and regulatory architecture required to contain its discriminatory potential remains embryonic. This article examines how algorithmic systems deployed in law enforcement, judicial administration, and public welfare reproduce and intensify pre-existing social inequalities, with particular consequences for marginalised communities including religious minorities, Scheduled Castes, and women. Drawing on constitutional doctrine principally Articles 14, 19, and 21 of the Constitution of India as interpreted in Justice K.S. Puttaswamy v. Union of India (2017) as well as comparative analysis of the EU AI Act, 2024, the article identifies critical doctrinal tools for contesting algorithmic bias and proposes a structured legislative and institutional framework for rights-protective AI governance in India***

I. INTRODUCTION

A machine, it is often said, cannot be prejudiced. Unlike a police officer shaped by social upbringing or a judge influenced by unconscious stereotypes, an algorithm operates on data and mathematical logic or so the argument goes. This comforting narrative of technological neutrality has driven the rapid integration of Artificial Intelligence into India's governance machinery, from the scheduling of court hearings to the identification of criminal suspects in public spaces. Yet the narrative is deeply misleading. Algorithms are not born in a vacuum. They are constructed by human beings, trained on data generated by human societies, and deployed within institutions shaped by centuries of hierarchy and discrimination. When these systems reflect, replicate, or amplify existing social inequalities, the claim of objectivity does not diminish the harm it conceals it.

India's investment in judicial and law enforcement AI has been substantial. Phase III of the eCourts Mission Mode Project allocated over seven thousand crore rupees toward digital court infrastructure, with explicit provision for AI-driven tools in High Courts covering intelligent scheduling, Natural Language Processing, and automated case management.

At the apex level, the Supreme Court of India has deployed SUVAS, the Supreme Court Vidhik Anuvaad Software, for regional language translation of judgments, and SUPACE, the Supreme Court Portal for Assistance in Court's Efficiency, to assist judges in research and case preparation.

State police forces have moved further still. Delhi Police's Crime Mapping, Analytics, and Predictive System (CMAPS) is positioned as a data-driven tool for anticipating criminal activity. Facial recognition technology has been deployed across multiple states for crowd surveillance, suspect identification, and post-incident investigation.

The pace of this adoption stands in stark contrast to the poverty of regulatory oversight. India's Digital Personal Data Protection Act, 2023 though a landmark step in data governance contains no provisions for mandatory algorithmic audits, no requirement that automated decisions be explained to those affected, and no independent mechanism to assess the discriminatory impact of AI systems before or after deployment.

The result is a governance landscape in which powerful, opaque, and potentially biased systems make or influence decisions that affect bail, sentencing, welfare, and employment with virtually

no constitutional or statutory check on their operation. This article proceeds in six parts.

Part II contextualises algorithmic discrimination by examining the international experience, with particular attention to the COMPAS recidivism-scoring controversy in the United States and the SyRI welfare surveillance litigation in the Netherlands.
Part III turns to India, analysing the discriminatory patterns documented in CMAPS and facial recognition deployments.

Part IV undertakes a constitutional analysis, examining how Articles 14, 19, and 21 as interpreted by the Supreme Court, provide doctrinal tools for contesting algorithmic bias. Part V considers the comparative regulatory model offered by the EU AI Act, 2024.

Part VI proposes a framework for rights-protective AI governance in India. A conclusion synthesises the argument and restates the central claim: that algorithmic systems deployed without constitutional accountability do not merely risk discriminating, they discriminate by design.

II. ALGORITHMS AND DISCRIMINATION

A. The Myth of Neutral Code

The concept of algorithmic discrimination challenges a foundational assumption of computational thinking: that mathematical processes, by virtue of their formal character, are immune to the social pathologies that afflict human judgment. This assumption collapses on examination.

Every algorithm is the product of choices, about what data to collect, which variables to weight, what outcomes to optimise for, and whose interests to prioritise. When those choices are made in societies marked by structural inequality, the resulting systems tend to encode and perpetuate that inequality. Computer scientist Cathy O'Neil has described such systems as 'Weapons of Math Destruction' models that are opaque, unquestioned, and destructive, operating at massive scale with disproportionate impact on the vulnerable. The opacity is not merely technical. It is frequently institutional: proprietary claims shield algorithms from public scrutiny, and the complexity of machine learning models makes even their designers unable to fully explain their outputs. The combination of scale, opacity, and institutional authority renders such systems uniquely dangerous.

B. COMPAS and the Racialisation of Risk

The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), developed by Equivant (formerly Northpointe Inc.), became the subject of international scrutiny following a 2016 investigation by ProPublica. Analysing data from Broward County, Florida, the investigation found that Black defendants were flagged as future high-risk offenders at almost twice the rate of white defendants who had comparable criminal histories. Simultaneously, white defendants who went on to reoffend were more frequently misclassified as low-risk.

The algorithm did not include race as an explicit variable yet it produced outcomes systematically correlated with race, because the variables it did use (neighbourhood, employment history, family background) functioned as proxies for racially stratified social conditions. The constitutional and judicial consequences crystallised in State v. Loomis, decided by the Wisconsin Supreme Court in 2016.

Eric Loomis challenged his sentence on the ground that the court had relied on a COMPAS score whose methodology was inaccessible due to its proprietary status. The court upheld the sentence, reasoning that the score was one factor among many, but acknowledged serious discomfort with the opacity of the tool. The United States Supreme Court declined to grant certiorari.8

The case thus established an uncomfortable precedent: that life-altering judicial decisions could rest, in part, on outputs that neither the defendant nor the judge could interrogate. What Loomis illuminates for India is not merely the danger of a specific algorithm but the structural vulnerability created when judicial or quasi-judicial processes incorporate opaque automated tools. The Indian Constitution, as this article argues below, provides far stronger grounds for challenging such opacity than the

procedural due process framework invoked in Loomis.

C. SyRI and the Right to an Explanation

A different dimension of algorithmic discrimination emerged in the Netherlands, where the government deployed the System Risk Indication (SyRI) to detect welfare fraud by combining data from multiple government databases and generating risk scores for individuals. In February 2020, the District Court of The Hague struck down SyRI as incompatible with Article 8 of the European Convention on Human Rights.

The court found that the system lacked the transparency necessary to enable individuals to understand, challenge, or contest the risk assessments made about them. Crucially, the court observed that SyRI was disproportionately deployed in lower-income, ethnically diverse neighbourhoods, a geographic targeting that produced discriminatory effects regardless of the system's formally neutral design. The SyRI judgment is instructive because it demonstrates that discrimination need not be intentional to be unlawful. A system deployed with the stated aim of administrative efficiency can produce outcomes that are constitutionally impermissible if those outcomes systematically disadvantage protected groups without adequate justification or transparency. This principle that effect, not intent, is the measure of discrimination has deep roots in Indian constitutional jurisprudence, as Part IV of this article demonstrates.

III. INDIA'S ALGORITHMIC INFRASTRUCTURE

(Predictive Policing, Facial Recognition, and the Architecture of Bias)

A. CMAPS: Encoding Historical Discrimination

Delhi Police's Crime Mapping, Analytics, and Predictive System (CMAPS) exemplifies the risks that arise when AI systems are built on data generated by discriminatory institutions. The system draws on decades of crime records, arrest data, and geographic information to generate risk predictions for specific areas and, in some configurations, individuals.

The foundational problem is that this historical data is not a neutral record of criminal behaviour, it is a record of policing behaviour, shaped by the biases, priorities, and prejudices of the officers and institutions that produced it.

Research by Vidushi Marda and Shivangi Narayan has documented that CMAPS disproportionately targets Muslim and Dalit communities, not because members of those communities are more likely to commit offences, but because historical overpolicing of their neighbourhoods has generated data that marks those areas as highrisk . The system then recommends intensified police presence in those areas, generating further arrests, producing more data confirming the original risk assessment, and tightening a feedback loop of algorithmic discrimination.

Scholars have termed this dynamic 'the surveillance trap': a self-reinforcing cycle in which predictive tools amplify the very patterns of selective enforcement they claim merely to observe.

The 2019 Status of Policing in India Report, produced jointly by Common Cause and the Centre for the Study of Developing Societies, documented widespread attitudinal bias among police personnel, with significant proportions of officers expressing prejudicial views toward minority communities.

When these attitudes have shaped the data on which CMAPS is trained, the algorithm does not transcend the bias of its human predecessors, it systematises, scales, and lends false scientific authority to that bias.

B. Facial Recognition Technology: Precision, Error, and Targeting

Facial recognition technology (FRT) has been deployed across Indian states for purposes ranging from the identification of missing persons to the surveillance of political protests. The technology operates by comparing a captured image against a database of stored images and generating a probability match.

The reliability of this matching process varies significantly across demographic groups, a technical limitation with serious legal consequences. The Vidhi

Centre for Legal Policy's 2021 empirical study of FRT deployment in Delhi found higher rates of false positives erroneous matches for women and individuals from minority religious communities compared to the overall population.

The study also documented that the technology was disproportionately deployed in areas with significant Muslim populations, creating a surveillance infrastructure whose geographic distribution mirrored the contours of religious minority residence. An RTI response obtained from Delhi Police revealed that an 80 per cent match was treated as a positive identification, a threshold that, when applied to a system with known demographic error disparities, produces a significant probability of wrongful identification in affected communities.

The pattern of deployment following communal violence in Delhi underlines the point. After incidents of intercommunal unrest, FRT was deployed to identify participants, with the vast majority of those subsequently charged belonging to the Muslim community.

Whether this outcome reflected accurate identification, algorithmic bias, or selective deployment or some combination of all three could not be determined because no independent audit of the technology's use was conducted. This opacity, as the following Part argues, is itself a constitutional problem.

"When a State deploys technology that produces higher error rates for some communities than others, then concentrates that technology's use in spaces where those communities live, the resulting pattern of misidentification is not an accident of engineering — it is an instrument of power."

IV. CONSTITUTIONAL ANALYSIS

(Fundamental Rights as a Framework for Algorithmic Accountability)

A. Privacy, Autonomy, and the Right to an Explanation: Article 21

The Supreme Court's unanimous nine-judge bench decision in Justice K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1 represents the most significant constitutional development for the governance of AI in India.

By recognising the right to privacy as an intrinsic component of life and personal liberty under Article 21, the Court established that the State cannot intrude upon the informational, bodily, or decisional autonomy of individuals without satisfying the requirements of legality, necessity, and proportionality.

Justice D.Y. Chandrachud's concurring opinion in Puttaswamy is of particular significance for algorithmic governance. His Lordship articulated privacy not merely as a shield against intrusion but as an affirmative entitlement to autonomy the capacity of individuals to make meaningful choices about their own lives and to receive explanations for state actions that affect those choices. Applied to algorithmic decision-making, this reasoning generates what might be termed a constitutional 'explainability imperative': where State action relies on automated systems to determine bail, custodial conditions, welfare entitlements, or access to justice, the individual affected has a constitutional right to understand the basis of that determination.

The earlier decision in Maneka Gandhi v. Union of India, (1978) 1 SCC 248 reinforced this interpretation by requiring that any procedure affecting personal liberty must be fair, just, and reasonable not merely formal compliance with enacted law.

An automated bail-scoring system that produces outputs no one can explain or contest cannot satisfy the standard of a fair and just procedure. The constitutional mandate of reasonableness thus operates as a direct constraint on the opacity of algorithmic systems used in the justice process.

B. Substantive Equality and Indirect Discrimination: Article 14

Article 14 of the Constitution guarantees equality before law and equal protection of the laws. The Supreme Court's evolution from formal to substantive equality most clearly articulated in E.P. Royappa v. State of Tamil Nadu, (1974) 4 SCC 3, where the Court declared that 'equality is antithetical to

arbitrariness' provides powerful constitutional tools for challenging algorithmic bias.

Formal equality requires only that the same rule be applied to all. Substantive equality requires that the application of rules not produce systematically unequal outcomes for protected groups without adequate justification.

Algorithmic systems that apply formally identical risk-scoring criteria to all individuals, but whose outputs are systematically correlated with religion, caste, or gender not because of explicit classification but because of proxy variables encoding social stratification violate the principle of substantive equality.

The doctrine of indirect discrimination, examined in Air India v. Nergesh Meerza, (1981) 4 SCC 335, captures precisely this form of differential impact. A facially neutral rule that disproportionately disadvantages a constitutionally protected group constitutes discrimination unless it is shown to be reasonably justified by a legitimate aim pursued by proportionate means.

CMAPS and FRT deployments, as documented above, produce differential impacts on Muslim and Dalit communities that cannot readily be justified by reference to any constitutionally legitimate objective. The arbitrariness doctrine independently supports constitutional challenge.

In Shayara Bano v. Union of India, (2017) 9 SCC 1, the Court applied the manifest arbitrariness standard to strike down a practice that lacked rational basis and was capricious in its operation.

An algorithmic system whose outputs cannot be explained, whose training data is unaudited, and whose discriminatory effects are undocumented satisfies the conditions for a finding of manifest arbitrariness.

C. Freedom of Expression and the Chilling Effect: Article 19

The deployment of AI surveillance tools and in particular the combination of facial recognition and predictive profiling poses a distinctive threat to the freedoms guaranteed by Article 19(1)(a) through (c): freedom of speech and expression, freedom of peaceful assembly, and freedom of association.

The mechanism of harm is not direct prohibition but the 'chilling effect' the rational anticipatory suppression of protected activity in response to the knowledge of pervasive surveillance.

The Supreme Court recognised the constitutional relevance of chilling effects in Shreya Singhal v. Union of India, (2015) 5 SCC 1, where the vagueness of Section 66A of the Information Technology Act was held to produce an unconstitutional chilling effect on free expression.

The logic applies with equal force to surveillance technology: when individuals know that their faces are being matched against databases of 'persons of interest', that their social media activity is being profiled for risk, or that their physical movements are being tracked and recorded, they rationally self-censor.

Political dissidents, religious minorities, journalists, and civil society activists precisely those for whom the freedoms of Article 19 are most vital are most acutely affected by this deterrent.

D. Natural Justice and the Right to Be Heard

The principles of natural justice, audi alteram partem (hear the other side) and nemo judex in causa sua (no person shall be judge in their own cause) constitute foundational components of procedural fairness under Indian constitutional law. The Supreme Court in Olga Tellis v. Bombay Municipal Corporation, (1985) 3 SCC 545 held that the right to be heard encompasses not merely the opportunity to speak but the entitlement to understand and meaningfully contest the basis of adverse decisions.23

Automated decision-making systems that produce outputs without human-interpretable reasoning violate this right at its core. An individual informed that they have been classified as a high bail-flight risk by an algorithm, but given no access to the variables, weights, or data that produced that classification, cannot meaningfully exercise the right to be heard. The hearing becomes a formality without substance

precisely the outcome that the natural justice requirement is designed to prevent.

V. THE EU AI ACT: A COMPARATIVE MODEL

(Risk-Based Regulation and the Architecture of Accountability)

The European Union's Regulation (EU) 2024/1689, the Artificial Intelligence Act represents the world's first comprehensive legislative framework for the governance of AI systems and offers important lessons for India.

The Act adopts a risk-based regulatory architecture, classifying AI systems into four categories unacceptable risk, high risk, limited risk, and minimal risk and imposing requirements of escalating stringency for higher-risk applications.

AI systems used in law enforcement including real-time biometric identification in public spaces, predictive policing tools, and risk assessment instruments used in criminal proceedings are classified as high-risk under Article 6 and Annex III of the Act.

Providers of such systems are required to implement risk management systems, ensure training data is sufficiently representative to minimise discriminatory outcomes, maintain technical documentation enabling regulatory audit, achieve appropriate levels of accuracy and robustness, and ensure meaningful human oversight before deployment. Affected individuals are entitled to explanations and to meaningful redress mechanisms.

Particularly significant is the Act's prohibition, subject to narrow exceptions, on real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes. This prohibition reflects a considered regulatory judgment that the risks of mass FRT deployment including chilling effects on public assembly, error-driven misidentification, and discriminatory geographic targeting outweigh the law enforcement benefits in the overwhelming majority of circumstances. India's current regulatory silence on this question is a conspicuous gap.

Three features of the EU model are of particular relevance to India's regulatory design challenge.

First, the mandatory ex ante conformity assessment the requirement that high-risk systems be evaluated for compliance before deployment rather than investigated only after harm has occurred shifts the burden of proof from affected communities to deployers.

Second, the requirement of data governance and bias mitigation as conditions of deployment rather than aspirational targets creates enforceable obligations.

Third, the establishment of a European AI Office as a dedicated oversight body with investigative and enforcement powers provides an institutional model for India, which currently lacks any body with the mandate, expertise, and independence to oversee AI deployments in high-stakes domains.

VI TOWARDS A CONSTITUTIONAL AI GOVERNANCE FRAMEWORK FOR INDIA

(Legislative Reform, Institutional Design, and Doctrinal Development)

A. Dedicated Algorithmic Accountability Legislation

India requires dedicated legislation, an Algorithmic Accountability and AI Governance Act, that moves beyond the patchwork of the DPDPA 2023 and the IT Act to address the constitutional challenges of automated decision-making comprehensively. Such legislation should be grounded explicitly in the constitutional values of equality, dignity, and procedural fairness, and should operate as a framework statute within which sector-specific rules can be developed.

The legislation should incorporate the following core obligations. First, mandatory algorithmic impact assessments before deployment of any AI system in a high-stakes domain, criminal justice, welfare, employment, immigration, or healthcare, examining discriminatory risks across protected categories, assessing the representativeness of training data, and documenting anticipated effects on marginalised communities. These assessments should be conducted by independent auditors and made

publicly available, subject only to genuinely narrowly defined exceptions for security-sensitive information. Second, a statutory right to explanation for all individuals subject to significant automated decisions by public authorities. The explanation must be meaningful articulating the principal factors that determined the output in terms an ordinary person can understand and contest not a formal recitation of system parameters.

This right should be accompanied by a right to human review: no automated decision affecting liberty, welfare, or fundamental rights should be final without the intervention of a human decision-maker with the authority and obligation to examine, question, and if necessary, override the algorithmic output.

B. Institutional Architecture: An Independent AI Oversight Authority

Legislative rights without institutional enforcement are aspirational rather than effective. India needs an independent AI Oversight Authority insulated from executive control, equipped with technical expertise, and endowed with investigative and enforcement powers.

The Authority should have the mandate to conduct or commission audits of deployed AI systems, receive and investigate complaints from affected individuals, issue binding orders requiring modification or suspension of non-compliant systems, and publish annual reports on the state of algorithmic governance in India. The Authority's composition must reflect the multidisciplinary character of the challenge. It should include legal experts with constitutional law specialisation, computer scientists and data engineers, social scientists with expertise in discrimination and inequality, and representatives of civil society organisations working with affected communities. Critically, its budget and staffing must be insulated from political pressure a lesson underlined by the experience of regulatory capture in analogous domains.

C. A Moratorium on Unaudited FRT in Public Spaces

Pending the establishment of the regulatory infrastructure proposed above, India should impose a moratorium on the deployment of facial recognition technology in public spaces for law enforcement purposes. The documented combination of demographic error disparities and discriminatory geographic targeting creates an unacceptable risk of wrongful identification and rights violation that cannot be managed under the current absence of regulatory oversight.

The moratorium should be lifted only upon the establishment of minimum accuracy standards disaggregated by demographic group, mandatory impact assessments, and robust oversight and redress mechanisms.

D. Doctrinal Development: Judicial Recognition of the Explainability Imperative

Courts have an important role to play, independently of legislative reform. The Supreme Court should, in an appropriate case, articulate the 'explainability imperative' as a component of the right to life and personal liberty under Article 21, establishing that no State authority may rely on an automated system to determine matters affecting fundamental rights without being able to explain, in terms that can be understood and contested, the basis of the system's output.

High Courts should exercise their jurisdiction under Article 226 to scrutinise the AI systems deployed by State police forces and administrative bodies, requiring disclosure of training data provenance, accuracy statistics disaggregated by demographic group, and impact assessment records.

CONCLUSION

The integration of artificial intelligence into India's legal and governance infrastructure is neither inherently good nor inherently bad. It is a political choice, a choice about whose interests are prioritised, whose risks are accepted, and whose voices are heard in the design and oversight of automated systems.

At present, that choice is being made by default rather than by deliberation, in the absence of the regulatory framework, institutional capacity, and constitutional accountability that a democracy committed to equality and dignity requires.

The Indian Constitution provides powerful tools for contesting algorithmic discrimination. The right to privacy and autonomy under Article 21, as interpreted in Puttaswamy, generates a constitutional explainability imperative that algorithmic opacity directly violates.

The substantive equality doctrine under Article 14 captures the indirect discrimination produced by systems trained on historically biased data. The natural justice principles embedded in procedural due process require that adverse automated decisions be meaningfully explained and genuinely contestable.

The chilling effect doctrine under Article 19 addresses the suppression of civil liberties produced by pervasive algorithmic surveillance. These doctrinal tools, however, require institutional vehicles to become effective. Without dedicated legislation, an independent oversight authority, and a judicially recognised explainability imperative, constitutional rights remain available only to those with the resources and knowledge to invoke them — typically not the marginalised communities most acutely affected by algorithmic discrimination.

The reform agenda proposed in this article is therefore not merely a technical regulatory exercise. It is a constitutional commitment: that India's technological future will be shaped by its founding values of equality, dignity, and justice, rather than by the unchecked reproduction of its hierarchical past.

BIBLIOGRAPHY

Books and Monographs

[1] Angwin, Julia et al., Machine Bias (ProPublica Investigations, 2016).

[2] Bostrom, Nick, Superintelligence: Paths, Dangers, Strategies (Oxford University Press, Oxford, 2014).

[3] Crawford, Kate, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence (Yale University Press, New Haven, 2021).

[4] O'Neil, Cathy, Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy (Crown Publishers, New York, 2016).

[5] Russell, Stuart, Human Compatible: Artificial Intelligence and the Problem of Control (Viking, New York, 2019).

[6] Zuboff, Shoshana, The Age of Surveillance Capitalism (PublicAffairs, New York, 2019)

Articles

[7] Barocas, Solon & Selbst, Andrew D., 'Big Data's Disparate Impact,' California Law Review, Vol. 104, No. 3 (2016), pp. 671–732.

[8] Marda, Vidushi & Narayan, Shivangi, 'Data in New Delhi's Predictive Policing

[9] System,' Proceedings of the ACM on Human-Computer Interaction, Vol. 5 (2021). Wachter, Sandra, Mittelstadt, Brent & Floridi, Luciano, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in GDPR,' International Data Privacy

[10] Law, Vol. 7, No. 2 (2017), pp. 76–99. Mittelstadt, Brent D. et al.,

[11] 'The Ethics of Algorithms: Mapping the Debate,' Big Data & Society, Vol. 3, No. 2 (2016), pp. 1–21

Cases

[12] Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1 Shreya Singhal v. Union of India, (2015) 5 SCC 1.

[13] Maneka Gandhi v. Union of India, (1978) 1 SCC 248.

[14] E.P. Royappa v. State of Tamil Nadu, (1974) 4 SCC 3. Shayara Bano v. Union of India, (2017) 9 SCC 1.

[15] Olga Tellis v. Bombay Municipal Corporation, (1985) 3 SCC 545.

[16] Air India v. Nergesh Meerza, (1981) 4 SCC 335.

[17] State v. Loomis, 2016 WI 68 (Wisconsin Supreme Court).

[18] District Court of The Hague, NJCM v. State of the Netherlands, C/09/550982/HA ZA 18-388 (5 February 2020).

Legislation and International Instruments

[19] Constitution of India, 1950.

[20] Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023).

[21] Information Technology Act, 2000 (Act No. 21 of 2000).

[22] Regulation (EU) 2024/1689 (EU Artificial Intelligence Act), OJ L 2024/1689