

**INTERNATIONAL JOURNAL OF LAW
MANAGEMENT & HUMANITIES**
[ISSN 2581-5369]

Volume 9 | Issue 5

2026

© 2026 *International Journal of Law Management & Humanities*

Follow this and additional works at: <https://www.ijlmh.com/>

Under the aegis of VidhiAagaz – Inking Your Brain (<https://www.vidhiaagaz.com/>)

This article is brought to you for free and open access by the International Journal of Law Management & Humanities at VidhiAagaz. It has been accepted for inclusion in the International Journal of Law Management & Humanities after due review.

In case of **any suggestions or complaints**, kindly contact support@vidhiaagaz.com.

To submit your Manuscript for Publication in the **International Journal of Law Management & Humanities**, kindly email your Manuscript to submission@ijlmh.com.

Predictive Policing and Artificial Intelligence in India's Criminal Justice System: Constitutional Limits and a Framework for Accountable Use

HARINATH REDDY G*  AND MANNE. SAINATH** 

ABSTRACT

*Artificial intelligence is entering Indian criminal justice through expanding data infrastructure, analytics and machine-learning tools. Predictive policing promises better deployment of scarce police resources, earlier identification of crime patterns and faster investigation. Its constitutional difficulty, however, is not confined to technical error. Police data are produced by earlier enforcement choices; a model trained on such data may convert unequal visibility into a forecast of unequal risk. Opacity can then prevent an affected person from discovering or contesting the basis of surveillance, intervention, bail or sentencing. This paper uses a doctrinal and comparative method to examine predictive policing under Articles 14 and 21 of the Constitution of India, the emerging Digital Personal Data Protection framework, the United States decision in *State v. Loomis*, and the European Union's risk-based regulation of artificial intelligence. It argues for a constitutional permission line: place-based analytics may assist non-coercive resource allocation when supported by law, necessity, audited data and public oversight, but a person-based score must never create suspicion, justify coercive police action, or determine a judicial outcome. The paper proposes statutory authorization, algorithmic impact assessment, data-quality and equality audits, meaningful disclosure, contestability, human responsibility, procurement controls, independent supervision and sunset review. These safeguards treat efficiency as a legitimate public objective while preserving the rule that coercive state power must rest on lawful, individualized and reviewable reasons.*

Keywords: Artificial Intelligence, Predictive Policing, Algorithmic Bias, Right to Privacy, Due Process

* Author is a Final year student at Law, Aurora Deemed to be University, Bhongir, Telangana, India. ORCID: <https://orcid.org/0009-0000-8354-6735>

** Author is a Final year student at Law, Aurora Deemed to be University, Bhongir, Telangana, India. ORCID: <https://orcid.org/0009-0000-4545-838X>

I. INTRODUCTION

Predictive policing changes the location of judgment in criminal justice. In conventional policing, an officer decides how to respond to observed facts. In an algorithmic system, earlier decisions about data collection, categories, weights, target variables and acceptable error are embedded in a model before the officer sees its output. The resulting forecast may look numerical and impersonal, yet it remains a product of institutional choices. Predictive policing is therefore best understood not as a machine replacing the police, but as a system that reorganizes discretion across data collectors, vendors, analysts, supervisors and frontline officers.¹

India has the conditions in which this transformation can occur rapidly. The Crime and Criminal Tracking Network and Systems (CCTNS) links police stations, while the Inter-Operable Criminal Justice System (ICJS) is designed to connect police, courts, prisons, forensic laboratories and prosecution. The Ministry of Home Affairs now describes ICJS 2.0 as enabling data analytics and AI and machine-learning tools in investigations.² Public statements and government sources also record plans or uses of AI-assisted and predictive methods by police forces.³ These developments do not establish that every Indian police force uses the same model; they establish something more important for legal analysis: national data integration and institutional demand make predictive deployment a present governance question rather than a distant possibility.

The policy environment has also evolved. NITI Aayog's responsible-AI papers framed safety, equality, inclusivity, privacy, transparency and accountability as constitutional values for AI. The Ministry of Electronics and Information Technology released the India AI Governance Guidelines in 2025 and the Government constituted an AI Governance and Economic Group in 2026. These instruments favour risk management and human-centric innovation, but they are not a police-powers statute and do not themselves specify when an algorithm may contribute to a stop, search, arrest, bail decision or sentence.⁴ The central legal problem is consequently one

¹ ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 15–21 (2017); NAT'L INST. OF JUST., *Overview of Predictive Policing* (June 9, 2014), <https://nij.ojp.gov/topics/articles/overview-predictive-policing>.

² MINISTRY OF HOME AFFS., Gov't of India, *Inter-Operable Criminal Justice System (ICJS)* (updated Mar. 12, 2026), <https://www.mha.gov.in/en/commoncontent/icjsncrb-administration>.

³ INDIAAI, *Delhi Police Plans to Integrate AI for Crime Control* (Feb. 11, 2022), <https://indiaai.gov.in/news/delhi-police-plans-to-integrate-ai-for-crime-control>; INDIAAI, *The Focus Will Be on People-Centric and AI-Based Policing, Says UP DGP Prashant Kumar* (Feb. 8, 2024), <https://indiaai.gov.in/news/the-focus-will-be-on-people-centric-and-ai-based-policing-says-up-dgp-prashant-kumar>.

⁴ NITI AAYOG, *Responsible AI #AIForAll: Approach Document for India, Part 1 - Principles for Responsible AI* 37–41 (2021); MINISTRY OF ELECS. & INFO. TECH., Gov't of India, *India AI Governance Guidelines* (2025), <https://www.meity.gov.in/content/india-ai-governance-guidelines>; INDIAAI, *Government of India Constitutes AI*

of translation: how should broad principles become enforceable limits where predictive output can expose a person or community to state coercion?

This paper advances a constitutional permission line. A place-based forecast may be used to plan non-coercive deployment - for example, deciding where additional patrol visibility or community services may be useful - if the system has a clear statutory basis, demonstrable necessity, validated data and independent oversight. A person-based score is different. It should never generate legal suspicion, supply grounds for coercive action, substitute for evidence, or determine bail, sentencing or parole. Individual liberty must turn on proved or lawfully observed facts, not a statistical resemblance to a group. This distinction permits carefully bounded administrative assistance without allowing prediction to become an unreviewable source of police power.

Three questions structure the inquiry. First, what efficiency can predictive systems realistically add, and what types of error do their data create? Second, how do Articles 14 and 21, fair-process principles and India's data-protection framework constrain their use? Third, what institutional design can make any permitted use reviewable? The answers require attention to the complete sociotechnical process rather than to model accuracy alone.

II. RESEARCH DESIGN, SCOPE AND CONTRIBUTION

The study adopts a doctrinal and comparative method. It examines constitutional provisions, statutes, judicial decisions and official policy materials, and reads them alongside interdisciplinary scholarship on policing data, risk assessment and algorithmic accountability. The Indian analysis is primary. United States and European Union materials are used as functional comparisons because they expose two different regulatory problems: judicial management of a proprietary risk score and ex ante regulation of high-risk AI. Comparative rules are not treated as automatically transferable to India's constitutional and policing structure.

The paper uses 'predictive policing' to mean the application of statistical or machine-learning methods to data in order to forecast a place, event, person or group associated with future crime risk and to guide a police response. It distinguishes this from ordinary digitisation, database search and retrospective evidence analysis. An automated photo match, for example, may raise separate questions of identification and accuracy without necessarily being a forecast. The distinction matters because prediction adds a temporal and inferential step: past records are used to assign future risk.

Governance and Economic Group (AIGEG) to Steer National AI Policy (Apr. 16, 2026), <https://indiaai.gov.in/article/government-of-india-constitutes-ai-governance-and-economic-group-aigeg-to-steer-national-ai-policy>.

The analysis is qualitative and does not claim to measure the performance of any undisclosed Indian model. Public documentation does not permit a reliable nationwide inventory of systems, training data, error rates or operational responses. That limitation is itself legally significant. A system capable of influencing coercive power should not depend for legitimacy on information unavailable to the public, courts and affected persons. The paper therefore evaluates permissible uses and governance requirements, not the accuracy of a particular proprietary product.

Existing literature has persuasively identified biased inputs, feedback loops, opacity and procedural harm. This paper's contribution is to organize those concerns into a rule for Indian public law. It connects the degree of permissible algorithmic influence to the legal consequence of the decision. The closer an output moves toward individual restraint or adjudication, the stronger the requirements of legal authorization, disclosure and human reason-giving - until, for decisions that create suspicion or determine liberty, the score must cease to have operative force.

III. PREDICTIVE POLICING: PROMISE, MODELS AND FAILURE MODES

A. Place-based forecasting

Place-based systems estimate where an offence may occur during a defined period. Inputs may include reported crime, calls for service, time, location and environmental features. Their practical attraction is straightforward: a supervisor can allocate scarce patrol time to smaller areas instead of distributing officers uniformly. The National Institute of Justice places predictive methods within the broader family of place-based proactive policing, while emphasizing that evaluation must consider effects on individuals, neighbourhoods and police organizations, not crime counts alone.⁵

Location is not a rights-neutral proxy. A hotspot may describe recorded enforcement rather than underlying harm. Offences visible in public space, complaints from communities with greater access to police, and discretionary stops are more likely to enter official data than hidden, unreported or selectively enforced conduct. Sending more officers to the predicted area creates more observations and records, which may return to the model as apparent confirmation. Lum and Isaac demonstrated this feedback problem by showing how drug-policing data can repeatedly direct attention toward already policed communities even where underlying drug use is more widely distributed.⁶

⁵ NAT'L INST. OF JUST., *Research Will Shape the Future of Proactive Policing* (Oct. 24, 2019), <https://nij.ojp.gov/topics/articles/research-will-shape-future-proactive-policing>; NAT'L INST. OF JUST., *From Crime Mapping to Crime Forecasting* (July 10, 2019), <https://nij.ojp.gov/topics/articles/crime-mapping-crime-forecasting-evolution-place-based-policing>.

⁶ Kristian Lum & William Isaac, *To Predict and Serve?*, 13 SIGNIFICANCE 14, 14–17 (2016).

Place-based prediction is nevertheless distinguishable from individualized suspicion. A forecast that guides the location of a visible patrol need not authorize any stop, search or questioning. Its lawful use can be limited to general deployment and paired with non-enforcement responses such as lighting improvements, transport management, victim-support outreach or problem-oriented policing. The constitutional risk rises when presence is converted into saturation enforcement, because residents then bear a concentrated probability of observation and intervention.

B. Person-based prediction and risk assessment

Person-based systems estimate whether an identified individual may offend, be victimized, breach supervision or reoffend. Inputs can include prior contacts, arrests, convictions, age, address, associations or variables correlated with socio-economic position. Risk-assessment instruments used at bail, sentencing, prison classification or parole are not identical to street-policing forecasts, but they share a governing concern: a probability derived from group data can influence an individualized legal decision. Ferguson describes this shift as one component of big-data policing in which linked information expands the state's capacity to identify and rank people.⁷

The apparent precision of a score can conceal three normative decisions. The first is the outcome chosen for prediction: arrest, reconviction and actual harmful conduct are not interchangeable. The second is the cost assigned to false positives and false negatives. A police agency may prefer to over-identify risk, but the burden of that choice falls on individuals who receive additional surveillance. The third is the response attached to the score. A model with moderate statistical performance may still be unlawful if its output triggers coercion without individualized grounds.

Sandra Mayson explains a deeper limit: every prediction extrapolates from a past structured by inequality. Removing an explicit protected characteristic does not necessarily remove that structure, because other variables may carry the same social information.⁸ A person-based score therefore risks laundering a contested institutional history into a seemingly objective statement about an individual's future. That concern cannot be answered by claiming that human decision-makers are also biased. The proper comparison is between accountable and unaccountable exercises of power, not between an ideal machine and an imperfect officer.

⁷ Ferguson, *supra* note 1, at 64–72, 83–95.

⁸ Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224–31, 2251–58 (2019).

C. Data quality, proxies and feedback

Training data are generated within legal institutions. Arrest records depend on deployment and enforcement; case outcomes depend on investigation, representation and charging; missing values reflect whose conduct was never seen. Richardson, Schultz and Crawford describe data produced during unlawful or systematically biased policing as ‘dirty data’ and show why later analytics may inherit those practices.⁹ The defect is not simply an incorrect row in a database. It can arise from the process that decided which populations would become legible to the state.

Proxies create a related problem. Even if caste, religion, race or income is excluded, location, language, social network and contact history may reproduce similar classifications. Barocas and Selbst show how data-mining systems can create disparate effects without an instruction to discriminate.¹⁰ For Indian policing, the relevant audit must therefore test error and exposure across constitutionally and socially significant groups, and must examine whether a variable has a legitimate connection to the stated operational purpose.

Scale magnifies these effects. Big-data systems can integrate records that were once separated by practical barriers. Brayne’s empirical study of data-driven policing illustrates how surveillance capacity expands through institutional linkage, secondary use and alerts.¹¹ India’s ICJS architecture similarly seeks cross-pillar data integration. Integration may improve investigation, but it also makes purpose limitation, access control, retention and correction essential. A mistaken or context-poor record can travel farther and influence more decisions than it could in an isolated file.

D. Efficiency must be demonstrated, not presumed

Predictive policing is often defended by reference to efficiency, yet efficiency has several meanings: fewer crimes, faster response, more arrests, lower cost or more consistent deployment. These measures may conflict. More recorded offences in a predicted area could indicate either improved detection or concentrated enforcement. An evaluation that counts only arrests may reward the feedback loop it should detect. Official United States research has cautioned that available studies do not always establish an advantage over conventional hotspot methods.¹²

⁹ Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 192, 196–204, 211–19 (2019).

¹⁰ Solon Barocas & Andrew D. Selbst, *Big Data’s Disparate Impact*, 104 CALIF. L. REV. 671, 677–94 (2016).

¹¹ Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977, 979–84, 993–1001 (2017).

¹² NAT’L INST. OF JUST., *Evaluation of the Shreveport Predictive Policing Experiment* (Mar. 6, 2018), <https://nij.ojp.gov/topics/articles/evaluation-shreveport-predictive-policing-experiment>.

A lawful evaluation should predefine the public objective, the baseline for comparison and the distribution of harms. It should measure false alerts, displacement, community complaints and demographic exposure alongside operational outcomes. The NIST AI Risk Management Framework usefully treats validity, reliability, explainability, privacy and managed harmful bias as connected characteristics rather than interchangeable labels.¹³ A model cannot be declared successful merely because it predicts the label on which it was trained; the label, intervention and social effect must also be justified.

IV. CONSTITUTIONAL AND STATUTORY LIMITS IN INDIA

A. Article 14: equality and non-arbitrariness

Article 14 binds the state whether a decision is made by an officer, a rule or an algorithm. The Supreme Court's equality jurisprudence rejects arbitrariness as incompatible with equality before law.¹⁴ An agency cannot avoid that duty by outsourcing a model or accepting a vendor's trade-secret claim. Selection of training data, target variables and thresholds are attributable to the state when they form part of public decision-making. Procurement changes the maker of a technical component; it does not privatize constitutional responsibility.

Algorithmic inequality can appear in at least three forms. Direct inequality occurs when a protected trait is used without constitutionally sufficient justification. Proxy inequality occurs when other variables reproduce that trait. Institutional inequality occurs when apparently neutral data reflect prior patterns of unequal police attention. A system may therefore have equal code and unequal operation. Testing only whether the same formula is applied to everyone misses the constitutional question: whether state-created burdens are distributed on a relevant, reasoned and non-arbitrary basis.

Risk prediction also creates a classification problem. Persons grouped as 'high risk' may differ on facts that matter to responsibility, while persons given the same score may reach it through different variables. Mayson's analysis shows why predictive inequality is not cured merely by equalizing one statistical measure.¹⁵ For police action, the constitutionally safer rule is categorical: no person-based score should create reasonable suspicion or supply grounds for a coercive step. An officer must identify independent, legally relevant and contemporaneous facts.

¹³ NAT'L INST. OF STANDARDS & TECH., *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* 12–17, 20–24 (2023), <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.

¹⁴ INDIA CONST. art. 14; *E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3, 38.

¹⁵ Mayson, *supra* note 8, at 2224–31, 2284–91.

B. Article 21: privacy, legality and proportionality

Predictive policing implicates privacy before an adverse decision is made. Collection, linkage and inference can reveal associations, movement, habits and vulnerabilities. In *K.S. Puttaswamy*, the Supreme Court recognized privacy as a fundamental right protecting personal liberty, dignity and informational self-determination.¹⁶ The judgment also requires state interference to have a legal basis, pursue a legitimate aim and satisfy proportionality. A general executive desire for modernization cannot by itself specify who may be analyzed, what data may be used, how long it may be retained or what consequence may follow.

Proportionality requires more than a useful correlation. The state must show a rational connection to a legitimate aim, consider less restrictive alternatives and maintain a proper balance between public benefit and rights harm. *Anuradha Bhasin* reinforces the need for necessity, proportionality and review of restrictions on fundamental rights.¹⁷ Applied to predictive policing, the analysis is use-specific. Aggregated deployment may be less intrusive than naming individuals; temporary use may be less intrusive than persistent watch-listing; a tool aimed at serious violence may present a different justification from one used for minor public-order enforcement.

Purpose limitation is constitutionally important even when data were lawfully collected. A record created to investigate a completed offence does not automatically become a legitimate basis for predicting unrelated future conduct. *Selvi v. State of Karnataka* connected personal autonomy and mental privacy to substantive due process in the criminal process, and warned against disproportionate investigative techniques.¹⁸ The technology differs, but the principle is relevant: state convenience does not extinguish the boundary between investigation of facts and compelled or inferred claims about a person's mind and future.

C. Fair procedure, reasons and the presumption of innocence

Article 21 requires procedure affecting liberty to be fair, just and reasonable.¹⁹ An opaque score can undermine each element. Notice is ineffective if the person does not know that a score exists. A hearing is hollow if neither the person nor the decision-maker can identify the variables, limitations or error rate. Review is weak if a vendor controls the relevant information.

¹⁶ *K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1, 297–98, 505–06.

¹⁷ *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637, 691–93.

¹⁸ *Selvi v. State of Karnataka*, (2010) 7 SCC 263, 312–16, 335–36.

¹⁹ *Maneka Gandhi v. Union of India*, (1978) 1 SCC 248, 281–84.

Citron's concept of technological due process captures this danger: automation can combine rule-making and adjudication while avoiding the safeguards of both.²⁰

The presumption of innocence supplies a substantive limit. Indian law recognizes it as a basic principle of criminal jurisprudence and a component of a fair trial.²¹ Prediction addresses probability, not culpability. It cannot prove an act, intention or legally admissible fact in the case before the court. A risk score should therefore be inadmissible as proof of guilt and incapable of aggravating sentence. At bail or parole, where law already permits forward-looking judgments, any analytical assistance must remain subordinate to disclosed statutory factors and individualized reasons. The decision-maker must be able to reach and explain the result without the score.

Reason-giving is the bridge between human responsibility and judicial review. A conclusion such as 'the algorithm marks the person high risk' is not a reason; it is a reference to an undisclosed process. Indian administrative law requires reasoned decisions because reasons discipline discretion and permit appellate or judicial scrutiny.²² Where a permitted tool materially informs a decision, the record should state its purpose, data period, known limitations, output, weight and the independent facts supporting the result.

D. The data-protection and AI-governance gap

The Digital Personal Data Protection Act, 2023 is relevant but not sufficient. The Act's core obligations and data-principal rights were placed on a phased commencement schedule and, as of 1 September 2026, sections 3 to 17 were scheduled to commence eighteen months after the 13 November 2025 notification. More fundamentally, section 17(1)(c) exempts processing for prevention, detection, investigation or prosecution of offences from most fiduciary obligations and data-principal rights; section 17(2)(a) permits notified exemptions for specified state interests.²³ The future operation of those provisions does not create an explanation right tailored to an algorithmic police forecast.

The Act's access, correction and grievance architecture is valuable for ordinary processing, but the law-enforcement exception leaves the setting of greatest coercive consequence without

²⁰ Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1253–59, 1305–13 (2008).

²¹ *Babu v. State of Kerala*, (2010) 9 SCC 189, 197; *Mohd. Hussain v. State (Gov't of NCT of Delhi)*, (2012) 2 SCC 584, 600–01.

²² *Kranti Assocs. (P) Ltd. v. Masood Ahmed Khan*, (2010) 9 SCC 496, 510–12; *Siemens Eng'g & Mfg. Co. of India Ltd. v. Union of India*, (1976) 2 SCC 981, 986.

²³ Digital Personal Data Protection Act, No. 22 of 2023, INDIA CODE (2023), §§ 17(1)(c), 17(2)(a); Digital Personal Data Protection Act, 2023, Notification G.S.R. 843(E), Gazette of India, pt. II, sec. 3(i) (Nov. 13, 2025) (placing sections 3–17 in the tranche commencing eighteen months after publication).

equivalent statutory detail.²⁴ Nor do the India AI Governance Guidelines, though current and important, independently authorize police interference or create a cause of action. They are best treated as a policy foundation from which sector-specific binding rules must be built.

The Criminal Procedure (Identification) Act, 2022 and integrated criminal-justice databases expand the state's ability to collect, preserve and search measurements and records. They do not, without more, answer the separate normative question whether those data may be repurposed to forecast future offending or how a forecast may influence a decision. A lawful data repository is not a blank cheque for every later inference.²⁵ Predictive use requires its own purpose, necessity and consequence rules.

V. COMPARATIVE LESSONS

A. United States: *State v. Loomis* and the limits of cautionary use

State v. Loomis concerned a proprietary COMPAS risk assessment included in a presentence investigation. The Wisconsin Supreme Court permitted consideration of the assessment but imposed cautions: the score could not determine whether the offender was incarcerated, the severity of sentence, or whether supervision in the community was safe; the court had to rely on independent factors.²⁶ The case is valuable because it recognizes that an algorithmic score needs legal limits even when used as decision support.

Loomis is not an adequate model for India in every respect. A warning label cannot provide meaningful challenge if the method, validation data or relevant error distribution remain unavailable. Investigative reporting on COMPAS also demonstrated how different error patterns can produce competing claims about fairness.²⁷ The broader lesson is institutional: a court must not allow commercial secrecy to control the information necessary to test a state decision. India should go further by prohibiting person-based scores from forming grounds for coercion and requiring disclosure sufficient for an effective challenge wherever an output is lawfully considered.

²⁴ Digital Personal Data Protection Act §§ 11–13, 17(1)(c).

²⁵ Criminal Procedure (Identification) Act, No. 11 of 2022, INDIA CODE (2022), §§ 2–4, 7; MINISTRY OF HOME AFFS., *supra* note 2.

²⁶ *State v. Loomis*, 881 N.W.2d 749, 757–70, ¶¶ 34–105 (Wis. 2016).

²⁷ Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; Mayson, *supra* note 8, at 2248–58.

B. European Union: prohibition, high-risk classification and fundamental-rights assessment

The European Union Artificial Intelligence Act offers a more structured risk-based approach. It prohibits AI used to assess or predict an individual's risk of committing an offence when the assessment is based solely on profiling or personality traits, while preserving analysis that supports a human assessment already grounded in objective and verifiable facts linked to criminal activity. Other law-enforcement uses listed in Annex III are treated as high risk.²⁸ This distinction resembles the permission line proposed here, although Indian safeguards must be anchored in domestic constitutional doctrine and police law.

For high-risk systems, the EU framework combines risk management, data governance, technical documentation, logging, transparency to deployers, human oversight, accuracy and robustness. Public bodies deploying specified high-risk systems must conduct a fundamental-rights impact assessment.²⁹ This lifecycle design is more useful than a single demand for 'explainable AI'. Explanation must be supported by records: what the system was intended to do, what data were used, how performance was tested, what incident occurred and who was responsible.

European data law also illustrates the need for precision. The General Data Protection Regulation does not create an unqualified, universal 'right to explanation'. It supplies information rights concerning automated decision-making and restricts certain solely automated decisions, while law-enforcement processing is principally addressed by Directive 2016/680 and national implementation.³⁰ The comparative lesson is not to borrow a slogan, but to combine information, human intervention and sector-specific legal authority.

VI. A RIGHTS-PROTECTIVE GOVERNANCE MODEL FOR INDIA

A. Statutory authorization and a consequence-based use matrix

Parliament or the competent state legislature should define predictive-policing use cases rather than leave them to general executive procurement. The law should identify the responsible authority, permitted data, approved purpose, retention period, sharing limits, prohibited variables, review mechanism and consequence of an output. Executive rules may supply

²⁸ Regulation (EU) 2024/1689 of the European Parliament and of the Council, arts. 5(1)(d), 6(2), annex III, 2024 O.J. (L 1689) 1.

²⁹ Id. arts. 9–15, 26(10), 27.

³⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council, arts. 13–15, 22, 2016 O.J. (L 119) 1; Directive (EU) 2016/680 of the European Parliament and of the Council, art. 11, 2016 O.J. (L 119) 89.

technical detail, but the essential limits on liberty and privacy require legislative guidance. This is the first requirement of Puttaswamy's legality standard, not a bureaucratic formality.

The statute should classify use by consequence. First, low-influence analytics that aggregate sufficiently de-identified data for research or long-term planning may be permitted subject to security and publication controls. Second, place-based operational forecasts may inform general deployment after validation and impact assessment, but may not lower the legal threshold for action against any person found in the area. Third, person-based prediction that creates suspicion, places a person on a coercive watch list or affects arrest should be prohibited. Fourth, judicial and correctional risk tools should never determine guilt or sentence; any narrowly permitted use at bail or parole must be disclosed, independently supported and contestable.

B. Mandatory algorithmic and fundamental-rights impact assessment

Before procurement or pilot deployment, the agency should publish a non-sensitive impact assessment. It should define the problem, evidence that prediction improves on a simpler alternative, affected populations, data provenance, expected error, rights risks, mitigation, security and exit criteria. The assessment must consider less restrictive means and explain why the chosen use is necessary. This operationalizes proportionality before harm occurs rather than relying only on litigation after deployment.³¹

High-risk pilots should be time-limited and registered. Material changes in data, model, vendor, threshold or operational response should trigger reassessment. Emergency use should have a short statutory duration and retrospective independent review. No agency should convert a research pilot into routine policing merely because the infrastructure already exists.

C. Data governance and equality audits

A data-governance protocol should document provenance, lawful collection, field definitions, missingness, time coverage, known enforcement distortions and correction procedures. Arrest, complaint, acquittal and conviction must remain distinct labels. Records concerning children, victims, witnesses and acquitted persons require heightened exclusion or protection. Data obtained for one statutory purpose should not enter a predictive model without a fresh legal and proportionality assessment.

Independent audits should test validity over time and across relevant groups. They should report false-positive and false-negative rates, calibration, geographic concentration, drift and the effect of the police response. No single fairness metric is sufficient. The audit must also examine

³¹ See *K.S. Puttaswamy*, (2017) 10 SCC at 505–06; NITI AAYOG, Responsible AI #AIForAll: Approach Document for India, Part 2 - Operationalizing Principles for Responsible AI 11–25 (2021).

whether an apparently accurate system creates unequal surveillance or changes the data-generating process.³² Failed thresholds should automatically suspend deployment pending reasoned reapproval.

D. Transparency, notice and contestability

Transparency has public and individual dimensions. The public should know that a system exists, its purpose, vendor, cost, general data categories, validation results, policies and audit summaries. Operational details may be withheld only where a specific and reviewable security justification exists. A blanket claim of trade secrecy should never defeat disclosure to a court, regulator or security-cleared independent auditor. Accountable-algorithm scholarship shows that auditability depends on designing records and verification into the system rather than demanding a post hoc narrative from a complex model.³³

An affected person should receive notice whenever an output materially contributes to an adverse decision, unless a court temporarily authorizes delayed notice for a concrete investigative reason. Disclosure should include the output, its date, its role, material input categories, known limitations and a route to correct data or challenge use. The authority must provide an intelligible statement of reasons. Indian law's insistence on reasons is especially important where the technical system may otherwise acquire an undeserved presumption of accuracy.³⁴

E. Human responsibility and procurement controls

Human review must be operationally meaningful. An officer who ordinarily accepts the output is not exercising independent judgment. The system should display uncertainty and limitations, require the officer to record independent facts, and permit departure without penalty. Supervisors should review override and acceptance patterns for automation bias. For liberty-affecting decisions, the authority must be able to justify the same outcome without treating the score as evidence. Loomis's insistence on independent factors is a minimum safeguard, not a complete governance model.³⁵

Public contracts should guarantee access to model documentation, validation materials, version history, incident logs and audit interfaces. Vendors should be required to notify material changes and support correction, portability and secure decommissioning. The government must retain

³² Richardson, Schultz & Crawford, *supra* note 9, at 211–19; NAT'L INST. OF STANDARDS & TECH., *supra* note 13, at 20–24.

³³ Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 636–44, 658–81 (2017).

³⁴ *Kranti Assocs.*, (2010) 9 SCC at 510–12; Citron, *supra* note 20, at 1305–13.

³⁵ *Loomis*, 881 N.W.2d at 769–70, ¶¶ 98–105.

the right to disclose information necessary for judicial review and defense. NITI Aayog's responsible-AI work and the NIST framework both recognize that governance must extend through procurement and the system lifecycle.³⁶ A contract that prevents constitutional scrutiny is unsuitable for criminal justice.

F. Independent supervision, remedies and sunset review

A multidisciplinary supervisory mechanism should include legal, technical, policing, statistics and civil-rights expertise. It should maintain a registry, approve high-risk pilots, inspect systems, commission audits, receive complaints and order suspension. The India AI Governance Guidelines and earlier responsible-AI papers provide a policy basis for coordinated oversight, but criminal-justice review requires functional independence and enforceable powers.³⁷

Remedies should match the harm. They should include correction of source data, deletion where retention lacks authority, removal from a watch list, disclosure, reconsideration by an untainted decision-maker, exclusion of unlawfully generated material and compensation where established law permits. Constitutional remedies under Articles 32 and 226 must remain available.³⁸ Every high-risk authorization should contain a sunset clause. Renewal should depend on published evidence of effectiveness and rights compliance, not on sunk cost or administrative convenience.

VII. CONCLUSION

Artificial intelligence can help Indian criminal-justice institutions organise information, identify patterns and deploy resources. Those advantages do not make prediction neutral and do not alter the legal source of coercive power. A forecast is constructed from past records, institutional choices and selected error costs. When it is applied to communities or individuals, it can reproduce unequal visibility, expand surveillance and obscure the reasons for state action.

The Constitution supplies a disciplined answer. Article 14 requires non-arbitrary and equal state action. Article 21 requires legality, necessity, proportionality and fair procedure when privacy or liberty is affected. The presumption of innocence prevents probability from becoming proof. These principles support a permission line: audited place-based analytics may assist non-coercive planning under a clear law, but person-based prediction must not create suspicion, justify coercion or determine an adjudicative outcome.

³⁶ NITI AAYOG, *supra* note 31, at 11–25; NAT'L INST. OF STANDARDS & TECH., *supra* note 13, at 20–24.

³⁷ MINISTRY OF ELECS. & INFO. TECH., *supra* note 4; NITI AAYOG, *supra* note 31, at 11–25.

³⁸ INDIA CONST. arts. 32, 226.

India's integrated criminal-justice data infrastructure, the phased data-protection regime and the 2025 AI Governance Guidelines make sector-specific legislation urgent. A credible framework should require prior impact assessment, documented data provenance, equality and performance audits, public registration, meaningful notice, contestability, human reason-giving, procurement access, independent supervision and sunset review. These controls do not oppose innovation. They identify the conditions under which a technological tool can serve a constitutional democracy.

The decisive question is therefore not whether an algorithm is more accurate than an unaided officer in the abstract. It is whether a particular use is lawfully authorized, necessary for a defined public purpose, independently testable and incapable of displacing individualized legal judgment. Efficiency is legitimate only within those boundaries. Criminal justice remains just when the state can disclose, defend and review the reasons on which it acts.
