

Towards Just Algorithms: Risk, Rights and Regulation in AI-Driven Criminal Justice in India

Pooja Kumari^{1*}, Vishal Anand²

^{1*}Research Scholar, University Department of Law, Patna University, Bihar. Email: advpooja.law@gmail.com

²Research Scholar, University Department of Law, Patna University, Bihar. Email: vishalanand.law@gmail.com

Abstract: This paper intends to analyse the frequent use of artificial intelligence tools like prediction of crimes and facial recognition in criminal justice system of India. The researcher is of view that AI is being used rapidly without any ethical or legal protections. Since artificial intelligence is a brainchild of human only therefore, these systems can be biased as preventing different countries. In India this is a serious problem because of the existing social stigmas like caste, religion and difference in income. Like for example Delhi's crime prediction tools and the national facial recognition programme clearly shows how these systems are prone to treat some groups unfairly thereby harming their basic rights. The researcher also points that the manner in which AI is used now goes against the fundamentals of Indian Constitution like equal treatment and right to life and freedom. To solve this the researcher suggests making stronger laws for how AI is used in the justice system. The includes setting clear rules, showing how the AI makes decision, building a fair and transparent system, holding accountability and always letting humans have the final say. The objective is to ensure that the technology aids the justice system rather than creating any hindrance to it.

Keywords: Algorithmic Bias, Artificial Intelligence (AI), Criminal Justice, Human Rights, Regulation

1. Introduction: The Algorithmic Turn in Indian Criminal Justice

The criminal justice system of India has been grappling with major issues such as massive backlog of cases, limited resources and delay legal procedures. In an effort to address these issues, the government is taking keen interest in AI as a comprehensive solution. Key government bodies, including the policy think tank NITI Aayog, are actively supporting the use of AI as a part of broader national initiative towards modernization. They believe that AI can significantly improve the legal system by making it quicker, efficient and easily accessible for all. However, there is a hidden problem. While the government publicly highlights the positive uses of AI in areas like health and farming, it is also quietly using powerful AI tools for policing, investigation, and surveillance. This means that while the focus is on helpful applications, AI is also being increasingly used to control and monitor citizens, which receives much less attention.

Mapping the Landscape

The integration of AI into the Indian CJS is not a future prospect but a present reality. Several key systems are already operational, forming an increasingly interconnected technological infrastructure:

- i. **Policing and Investigation:** The foundational layer of this digital architecture is the Crime and Criminal Tracking Network & Systems (CCTNS). Conceived after the 2008 Mumbai attacks and approved in 2009, CCTNS is a centralized digital platform that has integrated all of India's 17,130 police stations. It digitizes police operations, including 100% of First Information Report (FIR) registrations, creating a



nationwide searchable database of crimes and criminals, which serves as the primary data source for more advanced AI applications.

- ii. **Predictive Policing:** Building on CCTNS data, several police forces have deployed predictive policing systems. Delhi Police's Crime Mapping Analytics and Predictive System (CMAPS), for instance, analyses historical crime data, satellite imagery, and calls to the 'Dial 100' helpline to identify crime 'hotspots' and inform resource deployment. Similarly, police in Telangana use tools like their own version of CMAPS and the Integrated People Information Hub (IPIH), a '360-degree' profiling system, to forecast criminal activity and identify potential offenders.
- iii. **Surveillance:** The National Automated Facial Recognition System (AFRS), managed by the National Crime Records Bureau (NCRB), represents a significant expansion of state surveillance capabilities. This system is built to assist in identifying suspects, tracing criminals, find missing people, and keep a watch over large crowds. It does this by matching images from CCTV cameras with a large, combined database of photos taken from different government records.
- iv. **Judicial Assistance:** The Apex Court of India has taken the lead in using AI to improve how courts work. Some important tools include the Supreme Court Portal for Assistance in Court's Efficiency (SUPACE), an AI-based tool that helps in judging case management, summarising facts and analysing legal information, and the Supreme Court Vidhik Anuvaad Software (SUVAS), a translation tool that used AI to translate court judgements into local Indian languages.

Table 1: AI-Driven Tools in the Indian Criminal Justice System

Tool/System Name	Deploying Agency/Body	Primary Function	Stated Objective	Operational Status/Key Facts
CCTNS	NCRB, Ministry of Home Affairs	Centralized Crime & Criminal Database	Modernize police force, enable data sharing	All 17,130 police stations integrated; 100% FIRs registered on platform.
AFRS	NCRB, State Police Forces	National Facial Recognition Surveillance	Identify criminals, missing persons, track suspects	Deployed nationwide; accuracy as low as 2% noted in Delhi High Court.
CMAPS	Delhi Police	Predictive Policing (Hotspot Analysis)	Geographic profiling of crime, predictive modeling	Operational; uses data from CCTNS and 'Dial 100' calls to plot crime hotspots.
IPIH	Hyderabad Police	Predictive Policing (360° Profiling)	Predict individuals to be arrested through profiling	Operational; part of a broader push for 24/7 surveillance and predictive arrests.
SUPACE	Supreme Court of India	Judicial Research Assistance	Enhance court efficiency, assist judges with data	Launched in 2021; officially an assistive tool that does not engage in decision-making.

SUVAS	Supreme Court of India	AI-Powered Legal Translation	Translate judgments into vernacular languages	Actively used; translated over 36,000 judgments into Hindi by 2024.
--------------	------------------------	------------------------------	---	---

This paper argues that the rapid, uncoordinated, and largely unregulated deployment of AI in the Indian CJS creates a critical and dangerous tension. While these technologies are championed for their potential to enhance efficiency, they are being integrated into a legal and ethical vacuum. This approach imports profound risks of algorithmic bias, opacity, and systemic rights violations, which are uniquely amplified by India’s deep-seated social stratifications of caste, religion, and class. The government appears to be pursuing a dual-track strategy: promoting a benign “AI for good” narrative to the public while simultaneously building a powerful, coercive techno-legal infrastructure for state control that is deliberately shielded from robust legal scrutiny through mechanisms like the broad law enforcement exemptions in data protection law. This creates a perilous governance gap where the highest-risk applications of AI are the least regulated. Without a comprehensive, rights-based regulatory framework, these technologies risk creating a new form of “digital caste” system. In this system, historical discrimination is not eliminated but is instead automated, amplified, and laundered through a veneer of technological objectivity, posing a direct and formidable challenge to India’s constitutional promise of justice for all.

2. Chapter I – The Global Precedent: Lessons from Algorithmic Risk and Bias

The promise of introducing algorithms into the criminal justice sphere is seductive: to replace the fallible, subjective, and often biased decisions of human actors with dispassionate, data-driven calculations. The reality, however, has proven far more complex. A growing body of research and real-world experience demonstrates that algorithms, particularly those trained on historically biased data, can replicate and even amplify existing societal prejudices, creating new vectors for discrimination. The international experience with these tools, particularly in the United States, offers critical cautionary lessons for India.

Case Study: The COMPAS Controversy

The most prominent and well-documented example of algorithmic bias in the CJS is the controversy surrounding the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool. COMPAS is a proprietary risk assessment algorithm used widely in U.S. courts to inform decisions on pretrial release, sentencing, and parole. The tool generates a risk score based on over 100 factors, including a defendant’s criminal history, age, employment status, and community ties.

In 2016, a groundbreaking investigative report by ProPublica analyzed over 10,000 COMPAS scores from Broward County, Florida, and uncovered stark racial disparities in its predictions. The key findings were:

- i. **High False Positive Rate for Black Defendants:** Black defendants who did not go on to re-offend were nearly twice as likely to be misclassified as “high risk” compared to their white counterparts (45% vs. 23%). This error exposes individuals to harsher consequences like unwarranted pretrial detention.
- ii. **High False Negative Rate for White Defendants:** Conversely, white defendants who did commit new crimes were much more likely to be mislabeled as “low risk” (48% vs. 28%). This error potentially compromises public safety.
- iii. **Overall Accuracy:** The algorithm’s overall accuracy in predicting recidivism was only 61%, slightly better than a coin flip, and its ability to predict violent recidivism was a mere 20%.

The COMPAS case sparked a crucial debate about what constitutes “fairness” in an algorithm. Northpointe (now Equivant), the creator of COMPAS, defended its tool by arguing that it was fair based on the metric of predictive parity or calibration. This means that a given risk score (e.g., a 7) predicts the same likelihood of recidivism for both Black and white defendants. ProPublica’s critique, however, centered on error rate balance, arguing that the *consequences* of the algorithm’s mistakes were not distributed equally; the harm of being falsely labeled high-risk fell

disproportionately on Black individuals. Subsequent research revealed a troubling mathematical reality: when base rates of an outcome (like recidivism) differ between populations, it is often impossible for an algorithm to satisfy both predictive parity and error rate balance simultaneously.

This impossibility reveals that the choice of a “fairness” metric is not a neutral, technical decision but a profound policy choice. It forces a value judgment about which type of unfairness a society is willing to accept: inconsistent meaning of risk scores or disparate distribution of harmful errors. Deploying such tools without a transparent, public deliberation on these trade-offs amounts to “policy laundering,” where developers make critical ethical decisions under the guise of mathematical formalism, a role that rightfully belongs to democratic institutions.

The “Black Box” Problem

Compounding the issue of bias is the problem of opacity, often referred to as the “black box” problem. This describes the inability of users, and sometimes even developers, to understand *how* an AI model arrives at its conclusions. This opacity can stem from the sheer complexity of the model (as in deep learning) or from intellectual property protections that shield the algorithm as a trade secret, as was the case with COMPAS.

The implications of this opacity for the CJS are severe:

- i. **Undermining Due Process:** The right to a fair trial includes the ability of a defendant to confront and challenge the evidence presented against them. If a decision affecting one’s liberty is based on an inscrutable risk score, this right is rendered meaningless. A defendant cannot meaningfully contest an outcome if its underlying logic is hidden. This was a central issue in the landmark case *State v. Loomis*, where the defendant challenged the use of a secret COMPAS score in his sentencing. While the Wisconsin Supreme Court ultimately rejected the due process claim, it highlighted the profound legal questions raised by algorithmic opacity.
- ii. **Eroding Trust and Legitimacy:** Justice must not only be done but must be seen to be done. When decisions are delegated to opaque, unaccountable systems, public trust in the legitimacy of criminal justice institutions is eroded. If citizens cannot understand why a decision was made, they are less likely to perceive it as fair or just, regardless of its accuracy.

3. Chapter II – The Indian Context: AI, Policing, and Systemic Injustice

The risks of bias and opacity identified in global contexts are not merely replicated in India; they are profoundly amplified by the country’s unique socio-political fabric, which is characterized by deep-rooted structural inequalities. The uncritical adoption of AI into the Indian CJS, without accounting for these contextual realities, threatens to create an automated system of injustice.

3.1. The Data Foundation: Garbage In, Gospel Out

The efficacy and fairness of any AI system are fundamentally dependent on the quality of the data it is trained on. In India, the data underpinning CJS algorithms is fraught with systemic biases, making the principle of “garbage in, garbage out” a critical concern.

- i. **Historically Biased Crime Data:** The primary data source for policing AI, the CCTNS database, is not a neutral record of criminal activity. It is a reflection of historical and ongoing policing practices that are notoriously biased against marginalized communities. Police records in India are shaped by a colonial legacy and are widely acknowledged to exhibit casteist and communal prejudices. Consequently, CCTNS data reflects *who is policed*, not necessarily who commits crime, with Dalits, Adivasis, and Muslims being disproportionately targeted, arrested, and incarcerated.
- ii. **Representational and Measurement Biases:** The datasets used to train AI models often fail to represent India’s diversity. For example, prominent facial recognition datasets like ImageNet contain a negligible percentage of Indian faces, leading to lower accuracy. Within India, available data is often skewed towards

urban, male, and dominant-caste Hindu populations, effectively rendering other groups invisible or mischaracterized. This digital divide means that models are over-fitted for digitally-rich profiles, further excluding and disadvantaging those already on the margins. Feeding this structurally flawed data into algorithms results in outputs that, while presented as objective and scientific, are inherently discriminatory.

Case Study 1: Predictive Policing and the Creation of a Digital Caste System

Predictive policing systems in India exemplify how biased data can create automated systems of oppression. The Crime Mapping Analytics and Predictive System (CMAPS) used by Delhi Police and the Integrated People Information Hub (IPIH) in Hyderabad are prime examples. These tools analyze historical crime data from sources like CCTNS to identify ‘hotspots’ or, in the case of IPIH, to create ‘360-degree’ profiles of residents to predict who should be arrested.

This process creates a pernicious discriminatory feedback loop:

- i. Police, influenced by historical biases, disproportionately patrol and make arrests in neighbourhoods predominantly inhabited by marginalized communities like Dalits and Muslims.
- ii. This targeted enforcement inflates the crime statistics for these areas in official databases like CCTNS.
- iii. The predictive algorithm, trained on this skewed data, flags these very neighborhoods as ‘crime hotspots’.
- iv. Police leadership then uses these “data-driven” predictions to justify and intensify patrols and surveillance in the targeted communities.
- v. The increased police presence naturally leads to more arrests and recorded incidents, which are then fed back into the algorithm, further “validating” the initial prediction and reinforcing the cycle.
- vi. This creates a technologically-obscured tautology that legitimizes and automates pre-existing police biases. It transforms policing from a system marred by individual prejudice into one governed by seemingly objective, systemic discrimination, effectively constructing what has been termed a “digital caste panopticon”.

Case Study 2: Facial Recognition and the Architecture of Mass Surveillance

The nationwide rollout of the Automated Facial Recognition System (AFRS) illustrates the grave threat AI-driven surveillance poses to fundamental rights in India. The system exhibits a clear pattern of “function creep.” Initially, its use by the Delhi Police was justified before the High Court for the noble purpose of tracing missing children. However, its application has rapidly and opaquely expanded into a tool for mass surveillance and social control.

Reports have documented the Delhi Police using AFRS to film and identify individuals at political protests, creating databases of “rabble-rousers and miscreants”. This practice directly infringes upon the fundamental rights to freedom of speech and peaceful assembly, creating a chilling effect on dissent. Civil society organizations like the Internet Freedom Foundation (IFF) have challenged this illegal expansion of the technology’s use, demanding a halt to the system’s deployment for profiling and surveillance.

Furthermore, the accuracy of these systems is highly suspect. In 2018, the Delhi Police’s FRT was reported to have an accuracy rate as low as 2%. While officials now claim much higher accuracy rates of 90% or more with upgraded systems, these figures are often generated in controlled environments and do not reflect real-world performance. Crucially, global studies have consistently shown that facial recognition algorithms exhibit significant bias, performing least accurately on women and individuals with darker skin tones—a critical flaw in a country as diverse as India. This inherent inaccuracy, coupled with existing societal biases, means the technology is likely to produce a high rate of false positives, leading to wrongful arrests and investigations that disproportionately harm marginalized individuals.

Case Study 3: AI in the Judiciary - The Opaque Assistant

Even within the judiciary, where AI is presented in a purely assistive role, significant risks loom. The Supreme Court's SUPACE portal is designed to enhance judicial efficiency by aiding judges with legal research, fact extraction, and case summarization. The judiciary has been explicit that SUPACE is not a decision-making tool.

However, the introduction of AI into the sanctum of judicial reasoning is not without peril. A key concern is automation bias: the cognitive tendency for humans to over-rely on the outputs of automated systems. Even if SUPACE only summarizes a case file, the way it selects facts, prioritizes precedents, and structures the summary can subtly frame the issues and influence a judge's cognitive process. The opacity of the underlying algorithms means that this influence is invisible and cannot be easily scrutinized.

In the long term, think tanks like the Vidhi Centre for Legal Policy have warned of the danger of "value lock-in". If AI tools are trained exclusively on historical legal precedents, their consistent application could cause the law to stagnate. This would prevent the organic evolution of legal principles in response to changing societal values, a process that is essential for a "living constitution" like India's. Furthermore, the use of opaque "black box" tools fundamentally conflicts with the judicial norm of providing reasoned orders, which is essential for accountability and appellate review.

The interconnectedness of these systems creates a dangerous pipeline. A bias in policing data can be amplified by a predictive algorithm, reified by a facial recognition match, and ultimately presented as objective fact in a case file summarized by a judicial AI. This "cascading effect of algorithmic bias" compounds the harm at each stage, creating an almost insurmountable burden for a defendant seeking to prove their innocence against a chain of flawed, automated processes.

4. Chapter III – Constitutional Scrutiny and the Legal Vacuum

The deployment of AI technologies in the Indian CJS, in its current form, stands in direct conflict with the foundational principles of the Indian Constitution. The lack of a specific legal framework, coupled with the inherent risks of bias and opacity, renders these systems vulnerable to constitutional challenge on the grounds of violating the fundamental rights to equality and life and personal liberty.

The Challenge to Article 14: The Right to Equality

Article 14 of the Constitution guarantees both 'equality before the law' and 'equal protection of the laws'. While the classic test for classification requires an intelligible differentia with a rational nexus to the objective, the Supreme Court has evolved a more potent doctrine. In cases like *E.P. Royappa v. State of Tamil Nadu* and *Maneka Gandhi v. Union of India*, the Court established that arbitrariness is the antithesis of equality. Any state action that is unfair, unreasonable, or capricious is violative of Article 14.

AI-driven decisions in the CJS fail this test of non-arbitrariness on multiple fronts:

- i. **Reliance on Flawed Data:** Decisions derived from data that is known to be a reflection of historical prejudice rather than objective reality are inherently unreasonable and arbitrary. Using data that over-represents marginalized communities to justify increased surveillance or harsher risk assessments is a form of institutionalized prejudice, not rational decision-making.
- ii. **Opacity and the "Black Box" Problem:** A decision-making process that is inscrutable and cannot be explained is the very definition of arbitrary action. It violates the principles of natural justice, which demand reasoned orders. An individual cannot be subject to a decision affecting their liberty without being able to understand its basis.
- iii. **Indirect Discrimination:** Even when algorithms do not explicitly use protected characteristics like caste or religion as inputs, they can engage in indirect discrimination. In the Indian context, factors like postal code, socio-economic status, or even name can serve as strong proxies for caste and religion. When an algorithm

uses these proxies and produces outcomes that have a disproportionately adverse impact on protected groups, it violates the guarantee of equal protection.

The Violation of Article 21: Right to Life and Personal Liberty

Article 21 guarantees that no person shall be deprived of their life or personal liberty except according to “procedure established by law.” This procedure must be fair, just, and reasonable. The Supreme Court’s landmark judgment in *Justice K.S. Puttaswamy (Retd.) v. Union of India* firmly established the Right to Privacy as a fundamental right protected under Article 21.

The *Puttaswamy* judgment laid down a three-pronged test for any state intrusion into privacy: it must be backed by legality, serve a legitimate state aim (necessity), and be proportionate to the objective pursued. AI-driven surveillance tools like AFRS and predictive policing systems fail this test:

- i. **Legality:** The deployment of these systems is largely based on executive decisions and lacks a clear, specific statutory framework passed by Parliament. This absence of a legal mandate is a direct violation of the first prong of the test and is already the subject of public interest litigations (PILs).
- ii. **Necessity:** While preventing crime is a legitimate state aim, it is not proven that indiscriminate mass surveillance through technologies like FRT is a *necessary* means to achieve it. The questionable accuracy and high potential for error and misuse undermine the necessity argument.
- iii. **Proportionality:** The use of AFRS to conduct blanket surveillance of public spaces and political protests is a grossly disproportionate infringement on the privacy of countless citizens in order to identify a few potential wrongdoers. This creates a pervasive chilling effect on the exercise of fundamental rights like free speech and assembly, making the intrusion far greater than any potential benefit.

Furthermore, a procedure that relies on opaque, biased, and unaccountable algorithms cannot be considered a “fair, just, and reasonable” procedure as required by Article 21. Depriving an individual of their liberty based on the output of a black box algorithm is a fundamental violation of due process.

The Inadequacy of the Digital Personal Data Protection Act (DPDP Act), 2023

While India has enacted the Digital Personal Data Protection Act, 2023, as its primary data protection legislation, its efficacy in regulating AI in the CJS is severely limited. The Act contains broad and vaguely worded exemptions for government agencies when processing personal data for the “prevention, detection, investigation or prosecution of any offence” and for reasons of national security and public order.

These exemptions effectively carve out most high-risk AI applications in the criminal justice sphere from the Act’s key safeguards, such as the requirements for consent, purpose limitation, and data minimization. This creates a paradoxical situation where the state, the most powerful data fiduciary with the ability to deploy the most intrusive technologies, is subject to the least amount of regulation. This renders the DPDP Act largely ineffective in preventing the constitutional harms discussed, highlighting the urgent need for a specific, overriding legal framework for AI in the CJS. The current legal landscape is forcing a necessary evolution in constitutional jurisprudence, compelling courts to apply established principles of arbitrariness and proportionality to these novel technological threats, potentially leading to the development of a new “techno-constitutional” jurisprudence in India.

5. Chapter IV – Forging a Path Forward: A Regulatory Framework for Just Algorithms

The uncritical adoption of AI in the Indian CJS poses a clear and present danger to constitutional rights. To harness the potential benefits of this technology while mitigating its profound risks, a paradigm shift is required—from a focus on techno-solutionism to a rights-first approach. This necessitates the creation of a comprehensive regulatory framework built on principles of legality, transparency, fairness, accountability, and human control. This framework should draw on global best practices, such as the EU’s AI Act, and the recommendations of Indian legal think tanks like the Vidhi Centre for Legal Policy.

Principle 1: A Comprehensive Statutory Mandate

The current ad-hoc deployment of AI tools by executive fiat is legally untenable. To meet the constitutional requirement of “legality” established in *Puttaswamy*, any use of high-risk AI in the CJS must be sanctioned by a specific, comprehensive law passed by Parliament.

Solution: Enact a dedicated “AI in Criminal Justice Act.” This legislation should explicitly define the permissible uses of AI, establish clear boundaries, and create binding obligations for all state agencies. It must supersede the general exemptions provided under the DPDP Act, ensuring that CJS applications are subject to stringent data protection norms. The law should also establish an independent, statutory oversight body with multi-stakeholder representation-including judicial experts, technologists, ethicists, and civil society members-tasked with approving, auditing, and continuously monitoring all AI tools used in the justice system.

Principle 2: Radical Transparency and Explainability (XAI)

The “black box” problem is fundamentally incompatible with the principles of due process and public trust. A just system cannot be an opaque one.

- **Solution:** The regulatory framework must mandate transparency by default.
- **Open-Source Algorithms:** All algorithms used for public functions within the CJS must be open-source, allowing for independent scrutiny by researchers, academics, and the public to identify flaws and biases.
- **Right to Explanation:** The law must codify a “right to an explanation” for any individual subjected to a significant AI-driven decision. This is a core principle of Explainable AI (XAI) and is essential for enabling meaningful appeals and challenges.
- **Public AI Incident Database:** Following recommendations from MeitY’s own subcommittee, a public, anonymized database of AI-related incidents, errors, and harms should be established to track real-world risks and inform regulatory action.

Principle 3: Fairness-Aware by Design

To counter the risk of automating discrimination, fairness must be embedded into the entire lifecycle of an AI system, from conception to deployment.

- **Solution:** Mandate a “fairness-by-design” approach.
- **Pre-Deployment Impact Assessments:** Require mandatory, independent Human Rights Impact Assessments and Algorithmic Bias Impact Assessments before any high-risk AI tool is procured or deployed. These assessments must include meaningful consultation with the communities most likely to be affected.
- **Fairness-Aware Machine Learning (FAML):** Developers must be legally required to employ state-of-the-art FAML techniques-including pre-processing of data to remove bias, in-processing constraints to guide the algorithm towards fair outcomes, and post-processing adjustments to correct for discriminatory predictions-to actively mitigate bias in their models.
- **Robust Data Governance:** Enforce strict protocols for data collection, annotation, and use. This includes moving beyond reliance on flawed and biased police data and investing in the creation of representative, high-quality datasets for training and validation.

Principle 4: Robust Accountability and Redressal

When an AI system causes harm, there must be clear and accessible pathways for accountability and redress. The current ambiguity regarding liability is unacceptable.

- **Solution:** Establish clear lines of accountability.
- **Legal Liability:** The law must explicitly state that the ultimate responsibility for any decision rests with the human authority using the AI tool. A police officer cannot blame a wrongful arrest on an algorithm, nor can a judge delegate their judicial reasoning.
- **Effective Redressal Mechanisms:** Create a simple, accessible, and effective mechanism for individuals to challenge automated decisions and seek redress for harms caused. This could take the form of a specialized tribunal or an ombudsman operating under the independent oversight body.

Principle 5: Human-in-Command

The most critical safeguard is ensuring that human judgment remains central to the justice process. AI must only ever assist, never supplant, human decision-makers in high-stakes contexts.

- **Solution:** Codify the non-negotiable principle of “human-in-command.”
- **Prohibition on Solely Automated Decisions:** The law must explicitly ban the use of AI to make final, consequential decisions about a person’s liberty—such as decisions to arrest, detain, or determine a sentence—without meaningful, independent human review and final determination.
- **A Risk-Based Approach:** India should adopt a risk-based regulatory model similar to the EU AI Act. This would involve completely banning unacceptable-risk applications (e.g., social scoring, predictive policing based solely on individual profiling). It would subject high-risk applications (e.g., FRT, risk assessments used in court) to the strictest regulatory requirements, including conformity assessments, mandatory audits, and human oversight.

The following table contextualizes India’s current regulatory vacuum against international approaches, underscoring the urgency for reform.

Table 2: Comparative Analysis of AI Regulatory Frameworks

Parameter	India (Current State)	European Union (AI Act)	United States
Legal Status	No specific, comprehensive AI law.	Comprehensive, binding regulation (AI Act).	No federal law; patchwork of state laws.
Regulatory Approach	Principles-based guidance (NITI/MeitY); no binding rules.	Risk-based (Unacceptable, High, Limited, Minimal risk tiers).	Fragmented; sectoral or technology-specific state laws.
Prohibited Practices	None defined in law.	Social scoring, real-time public biometric ID, manipulative AI banned.	Varies; some cities/states have moratoriums on facial recognition.
High-Risk AI Rules	None.	Strict obligations: conformity assessments, risk management, data governance, registration in EU database.	Varies; e.g., Colorado’s law on high-risk AI requires impact assessments.

Transparency Mandates	Voluntary commitments suggested.	Legally mandated explainability, human oversight, and transparency for high-risk systems.	Varies by state; includes disclosure requirements and impact assessments.
Enforcement Body	None.	EU AI Office and designated national competent authorities.	Fragmented; Federal Trade Commission (FTC), state Attorneys General.

6. Conclusion

The integration of Artificial Intelligence into the Indian criminal justice system represents a critical juncture. The analysis presented in this paper demonstrates that a path of uncritical, unregulated technological adoption-driven by a narrative of efficiency-is fraught with peril. The deployment of predictive policing, facial recognition, and other algorithmic tools in a legal and ethical vacuum does not merely import global risks of bias and opacity; it amplifies them through the unique prism of India's deep-seated social hierarchies. This trajectory is on a collision course with the fundamental rights to equality and liberty enshrined in the Constitution.

The current approach threatens to create a new, insidious regime of "algorithmic injustice." In this regime, the historical prejudices of the CJS are not eradicated but are instead codified, automated, and laundered through the seemingly neutral language of data and algorithms. This makes discrimination harder to detect, more difficult to challenge, and more systemic in its impact, fundamentally undermining the constitutional promise of a fair and just legal process for every citizen.

This moment demands a paradigm shift. India must move decisively from a posture of techno-solutionism to one of constitutional caution and rights-centric governance. The goal is not to reject technology but to subordinate it to the rule of law and human rights. The comprehensive regulatory framework proposed-founded on a statutory mandate, radical transparency, fairness-by-design, robust accountability, and inviolable human oversight-offers a pathway to achieve this balance. By implementing such a framework, India can ensure that its algorithms are instruments of justice, not its modern-day antithesis, thereby safeguarding the integrity of its democratic and constitutional foundations in the age of AI.

REFERENCES

1. Lawful Legal. (n.d.). *ARTIFICIAL INTELLIGENCE (AI) IN INDIAN CRIMINAL JUSTICE SYSTEM*. Retrieved from <https://lawfullegal.in/artificial-intelligence-ai-in-indian-criminal-justice-system/>.
2. A public policy and corporate affairs consultant. (2020, March 21). *How AI can be used in policing to reform criminal justice system*. ThePrint. Retrieved from <https://theprint.in/tech/how-ai-can-be-used-in-policing-to-reform-criminal-justice-system/384786/>.
3. JETIR Research Journal. (2023, August). *A legal study on role of AI in criminal justice system in India*. Retrieved from <https://www.jetir.org/papers/JETIR2308092.pdf>.
4. National Crime Records Bureau. (n.d.). *Crime and Criminal Tracking Network & Systems (CCTNS)*. Ministry of Home Affairs, Government of India. Retrieved from <https://ncrb.gov.in/en/cctns>.
5. Internet Freedom Foundation. (2022). *Predictive Policing in India: A new form of an old problem*. Retrieved from <https://internetfreedom.in/pdictive-policing-in-india-a-new-form-of-an-old-problem/>.
6. Srinivas, M. (2023, May 22). *'TS Cops' new app to profile citizens, track their real-time movements*. The New Indian Express. Retrieved from <https://www.newindianexpress.com/states/telangana/2023/May/22/ts-cops-new-app-to-profile-citizens-track-their-real-time-movements-2577627.html>.
7. Ministry of Home Affairs. (2020, January 21). *National Automated Facial Recognition System*. Press Information Bureau, Government of India. Retrieved from <https://pib.gov.in/PressReleasePage.aspx?PRID=1600122>.
8. Supreme Court of India. (2021, April 6). *Chief Justice of India launches Artificial Intelligence Portal 'SUPACE'*. Retrieved from https://main.sci.gov.in/supremecourt/2021/11199/11199_2021_0_1_27383_PressRelease.pdf.
9. Supreme Court of India. (n.d.). *SUVAS*. Retrieved from <https://main.sci.gov.in/suvas>.
10. Aneja, U., & A. (2021). *Fairness, Accountability and Transparency in AI: A study of the Indian ecosystem*. Digital Futures Lab. This report highlights the absence of a dedicated legal framework for AI in India and discusses the ethical challenges this poses for deploying AI in critical sectors.

11. R, P. (2023, September 11). *Racist, sexist, casteist: Is AI bad news for India?* The Economic Times. Retrieved from <https://m.economictimes.com/tech/technology/racist-sexist-casteist-is-ai-bad-news-for-india/articleshow/103567192.cms>.
12. *The Digital Personal Data Protection Act, 2023*, § 17(1)(c) & 17(2)(a) (India).
13. Mathiyazhagan, S. (2023, October 12). *Caste in the Machine*. The Wire. Retrieved from <https://thewire.in/caste/caste-in-the-machine>. The author explores how algorithmic decision-making can codify and perpetuate caste-based discrimination, creating what can be termed a “digital caste” system.
14. O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown. This seminal book provides numerous examples of how algorithms, despite being portrayed as objective, can reinforce and amplify societal inequalities across various domains, including the justice system.
15. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks*. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
16. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). *Inherent Trade-Offs in the Fair Determination of Risk Scores*. ArXiv:1609.05807 [Cs, Stat]. <http://arxiv.org/abs/1609.05807>.
17. Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin’s Press.
18. *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).
19. Narayan, S. (2020). *Past, Present, and Past as Present in India’s Predictive Policing*. XRDS: Crossroads, The ACM Magazine for Students, 27(2), 28–32. This paper discusses how the Crime and Criminal Tracking Network & Systems (CCTNS) digitizes historical police records that are rooted in colonial-era biases.
20. Sonavane, N., & Singh, A. (2021). *Settled Habits, New Tricks: Casteist Policing Meets Big Tech in India*. Transnational Institute. This report details the colonial origins of policing in India and the continued targeting of Denotified Tribes and other marginalized groups, whose identities are now being encoded into digital databases like CCTNS.
21. Amnesty International. (2016). *India: Impunity, Discrimination and Repression of Dissent*. This report notes the disproportionate representation of Dalits, Adivasis, and Muslims in India’s prison population, reflecting biased policing practices.
22. Marda, V., & Narayan, S. (2020). *Data in New Delhi’s Predictive Policing System*. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* ‘20). Association for Computing Machinery, New York, NY, USA, 237–246. The authors discuss how datasets used for AI are often not representative of diverse populations, leading to biased outcomes.
23. Project 39A. (2022, March 24). *Predictive Policing: Discrimination, Due Process and Democratic Accountability*. This article describes the functioning of CMAPS in Delhi and the IPIH in Hyderabad, highlighting how they use data to create “hotspots” and “360-degree” profiles.
24. Narayan, S. (2020). *Past, Present, and Past as Present in India’s Predictive Policing*. The author’s ethnographic research reveals police biases against lower-caste and Muslim communities, influencing where and how they patrol.
25. Marda, V., & Narayan, S. (2020). *Data in New Delhi’s Predictive Policing System*. The paper explains how targeted enforcement in certain areas leads to skewed crime statistics in databases like CCTNS.
26. Ibid. The predictive algorithms, trained on this biased data, then identify these same areas as high-crime “hotspots,” reinforcing the initial bias.
27. Sonavane, N., & Singh, A. (2021). *Settled Habits, New Tricks: Casteist Policing Meets Big Tech in India*. Transnational Institute. The authors use this term to describe the creation of a technologically-enforced system of surveillance and control based on caste.
28. Internet Freedom Foundation. (2019, December 28). *Notice to cease use of facial recognition technologies by the Delhi Police as it is an illegal act of mass surveillance*. The legal notice states that the Delhi High Court’s order for the use of AFRS was for the purpose of tracing missing children.
29. Mazoomdaar, J. (2019, December 28). *Delhi Police film protests, run its images through face recognition software to screen crowd*. The Indian Express. This report details the use of AFRS to identify protestors.
30. Internet Freedom Foundation. (2019, December 28). *Notice to cease use of facial recognition technologies by the Delhi Police....* This legal notice from IFF demanded a halt to the use of AFRS for surveillance of protests, arguing it was an illegal expansion of its purpose.
31. Press Trust of India. (2018, August 23). *Delhi police facial recognition software has only 2 per cent accuracy: HC told*. Business Standard.
32. Supreme Court of India. (2021, April 6). *Press Note on the launch of Supreme Court Portal for Assistance in Court’s Efficiency (SUPACE)*. The official announcement emphasized that SUPACE is an assistive tool and not for decision-making.
33. Vidhi Centre for Legal Policy. (2021). *Responsible Artificial Intelligence for the Indian Justice System*. While not using the exact term “automation bias,” the report discusses the risk of over-reliance on AI tools and the need for human oversight to ensure justice is not compromised by technological solutions.
34. Ibid. The report warns against the risk of legal stagnation if AI systems, trained on historical data, are not designed to accommodate the evolution of legal principles in line with societal changes.

35. Internet Freedom Foundation. (2019, December 28). *Notice to cease use of facial recognition technologies....* The notice argues that the use of AFRS for surveillance at protests violates fundamental rights under the Indian Constitution, including Articles 14, 19, and 21. See also, Marda, V., & Narayan, S. (2020). *Data in New Delhi's Predictive Policing System*.
36. *E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3.
37. *Maneka Gandhi v. Union of India*, (1978) 1 SCC 248.
38. Marda, V., & Narayan, S. (2020). *Data in New Delhi's Predictive Policing System*. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 237–246. The paper argues that using biased historical data makes resulting algorithmic decisions inherently arbitrary.
39. T-SID. (2021). *Use of Artificial Intelligence in the Indian Criminal Justice System: A study of the constitutional and legal implications*. This report discusses how the opacity of “black box” systems contravenes the principles of natural justice and reasoned decision-making. Retrieved from <https://www.t-sid.org/use-of-artificial-intelligence-in-the-indian-criminal-justice-system>.
40. Sonavane, N., & Singh, A. (2021). *Settled Habits, New Tricks: Casteist Policing Meets Big Tech in India*. Transnational Institute. This report details how seemingly neutral data points like location or economic status can act as proxies for caste and religion, leading to indirect discrimination.
41. *Maneka Gandhi v. Union of India*, (1978) 1 SCC 248. This case established that the “procedure established by law” under Article 21 must be fair, just, and reasonable, not merely any enacted procedure.
42. *Justice K.S. Puttaswamy (Retd.) & Anr. v. Union of India & Ors.*, (2017) 10 SCC 1.
43. *Ibid.* The judgment articulated a three-part test for any permissible limitation of the right to privacy: legality, necessity, and proportionality.
44. Internet Freedom Foundation. (2022). *#BanTheScan In India: Facial Recognition Technology*. IFF has filed and tracked PILs arguing that the use of AFRS lacks a specific legal mandate from Parliament. Retrieved from <https://internetfreedom.in/facial-recognition-technology-in-india-a-clear-and-present-danger-to-your-rights/>.
45. Internet Freedom Foundation. (2019, December 28). *Legal Notice to the Delhi Police against the use of facial recognition technology*. The notice questions the necessity of FRT due to its low accuracy rates and potential for misuse.
46. Amnesty International India. (2021). *Ban the Scan: The human rights impact of facial recognition in India*. The report details how FRT use at protests is disproportionate and creates a chilling effect on the rights to freedom of assembly and expression.
47. Jain, A., & Moses, B. (2021). *A New Jurisprudence for Artificial Intelligence*. The India Forum. The authors discuss how decisions based on opaque algorithms violate due process rights enshrined in Article 21.
48. *The Digital Personal Data Protection Act, 2023*, No. 22 of 2023, § 17(1)(b). The Act provides broad exemptions to the processing of personal data for the investigation or prosecution of an offence.
49. Internet Freedom Foundation. (2023, August 7). *The Digital Personal Data Protection Bill, 2023: An explainer*. IFF explains how the broad exemptions for the State limit the Act's protections against government surveillance and use of AI.
50. Vidhi Centre for Legal Policy. (2021). *Responsible Artificial Intelligence for the Indian Justice System*. This report outlines a rights-based framework for regulating AI in the Indian justice system.
51. *Ibid.* The Vidhi report recommends a specific, rights-based legislation for AI in the justice system, including the creation of an independent oversight body.
52. European Commission. (2019). *Ethics Guidelines for Trustworthy AI*. The guidelines identify explainability, which includes the concept of a ‘right to explanation’, as a key requirement for trustworthy AI.
53. Ministry of Electronics and Information Technology. (2021). *Report of the Ad-Hoc Committee of Experts on AI*. The committee recommended the establishment of a public repository for AI incidents to enhance transparency and safety.
54. European Union Agency for Fundamental Rights. (2022). *Getting the future right – Artificial intelligence and fundamental rights*. The report advocates for mandatory Fundamental Rights Impact Assessments before the deployment of AI systems.
55. Vidhi Centre for Legal Policy. (2021). *Responsible Artificial Intelligence for the Indian Justice System*. The report emphasizes that legal accountability must remain with the human decision-maker.
56. *Ibid.* The report strongly recommends that AI should only play an assistive role and that final decisions, especially those affecting liberty, must be made by a human.
57. European Parliament. (2023). *EU AI Act: first regulation on artificial intelligence*. The EU's risk-based approach, which bans or strictly regulates AI based on its potential for harm, is presented as a model for other jurisdictions.