PREDICTIVE POLICING AND THE BIAS TRAP: ANALYSING SYSTEMIC CHALLENGES IN ALGORITHMIC CRIMINAL JUSTICE

Vijayaraghavan K, LLM, School of Criminal Law and Criminal Justice Administration, Tamil Nadu Dr. Ambedkar Law University, School of Excellence, Taramani, Chennai, Tamil Nadu, India

ABSTRACT

This abstract examines how place-based and person-based algorithmic tools embed systemic bias into public safety operations, converting flawed past data into automated, discriminatory futures. By analyzing the mechanics of machine learning models used in modern policing, this paper details the technical and institutional pipelines that codify racial and socioeconomic profiling into software infrastructure. Predictive policing technologies are data-driven innovations for optimized law enforcement. This method suffers lack in transparency and data reliability due to its increasing replicate and amplification as a result of historical human prejudices. At the core of algorithmic bias in law enforcement is the reliance on "dirty data." Machine learning models are trained on historical arrest, citation, and stop-and-frisk records rather than true, objective crime occurrences. Because these underlying datasets reflect decades of disproportionate policing in marginalized communities, racial minorities, and low-income neighborhoods, the algorithms inherit these skewed patterns. This reliance creates a destructive, self-fulfilling feedback loop. When software flags a specific zip code or demographic group as high-risk, police departments deploy additional patrol units to those areas. This heightened surveillance inevitably leads to more arrests and citations for minor infractions, which are then fed back into the system as new data points, validating and reinforcing the algorithm's initial, biased prediction.

Keywords: Artificial intelligence, predictive policing, crime predication, algorithmic bias, criminal justice system.

1. Introduction

Predictive policing is a technique used by the law enforcement agency that uses data and analytics to predict, avert and respond to criminal act. The predictive justice makes use of digital technologies in arriving solution in judicial process. Predictive policing as defined by Perry refers to the application of quantitative analytical techniques based on the statistical predictions by the police either to solve or prevent crimes. In recent years the use of algorithm has seen a increasing trend in adaptation by police forces and judicial authorities towards predictive profiling technologies in criminal justice and investigations.[1,2,3] Algorithmic bias in predictive policing refers to the systematic and unfair distortion of crime forecasts caused by feeding historically discriminatory data into automated systems. While marketed as objective, mathematical tools, predictive policing algorithms (PPAs) frequently automate, cement, and amplify human prejudices rather than removing them. This dynamic deepens racial profiling and creates hostile friction between law enforcement and marginalized communities. The predictive policing tools facilitates police swift action with respect to accuracy and efficiency. Thus AI predictive policing can improve crime prevention, it also brings within itself the risk associated with social injustices due to data bias and violation of privacy.[4 ,5]

Furthermore, the elimination of explicit demographic variables, such as race or gender, fails to neutralize algorithmic discrimination due to the persistence of "bias by proxy." The software utilizes highly correlated alternative data points such as geographic location, income brackets, educational background, and social networks to effectively substitute for protected characteristics. This discrimination is compounded by the "black box" nature of predictive software. Because these systems are typically developed by private vendors who protect their source code as proprietary intellectual property, they remain shielded from independent academic auditing, public scrutiny, and constitutional challenges regarding due process. The real-world consequences of these biased mechanisms are severe. In courtrooms and precinct stations globally, algorithms consistently output higher false-positive risk scores for Black and brown individuals compared to their white counterparts. This structural failure has prompted

[1] Perry WL. Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation; 2013.
[2] Perego B. The impact of predictive policing system on civil rights. Biolaw J 2019.
[3] Söderholm S. Fundamental rights control when implementing predictive policing–a European perspective. Peking Univ Law J 2023;11(1):91–104.
[4] Amit S., Mayuri S. Role of Artificial Intelligence in Predictive Policing: A Double-Edged Sword in Combating Crimes against Women, NeuroQuantology, 2022; 20(1), 1430-1452. DOI: 10.48047/NQ.2022.20.1.NQ22435
[5] https://www.privacyinternational.org/sites/default/files/2021-06/predictive_policing.pdf

significant civil rights pushback and shifting international legal landscapes. Robust frameworks like the European Union's AI Act[6] have placed strict prohibitions or high-risk classifications on autonomous risk-profiling tools in law enforcement. To counter these systemic harms, a combination of technical and governance interventions is required. This paper reviews essential mitigation strategies, including the application of mathematical fairness metrics during model training, conditional score recalibration to protect vulnerable subsets, and the integration of community-led oversight boards to democratize technology deployment. Ultimately, this analysis argues that algorithmic bias in predictive policing is not a minor technical glitch, but a fundamental structural flaw that demands rigorous regulation and systemic reform to safeguard human rights and civil liberties.

2. Mechanisms of Algorithmic Bias

The distortion inside predictive policing models stems from specific structural flaws:[7]

- **Dirty Data Ingestion**: Algorithms rely on historical arrest and citation files to make forecasts. If historical practices targeted specific racial minorities or low-income neighborhoods, the resulting data is inherently skewed.

- **The Feedback Loop**: When an algorithm flags a neighborhood as a "hot spot," law enforcement deploys more patrols there. Increased police presence naturally yields more arrests and citations, which are then fed back into the model. This training cycle treats heavily policed areas as dangerous purely because they are heavily policed.

- **Proxy Variables**: Even when developers explicitly omit protected categories like race, models substitute zip codes, socioeconomic indicators, and income levels as proxies. This results in indirect discrimination against the exact same demographics.

- **Proprietary Opacity**: Many systems are built by private vendors under trade-secret protections. The resulting "black box" prevents public audits, independent security evaluations, and algorithmic cross-examinations in court proceedings.

[6] The European Union's AI Act 2024/1689. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[7] Eylül B, Yueqi C, Sarah G, et al., Predictive Policing or Predictive Prejudice? A Study of the Legal, Historical and Ethical Implications of AI in Policing, 2025. https://www.oxjournal.org/predictive-policing-or-predictive-prejudice/

3. Core Modalities of Predictive Systems

Predictive policing generally operates through two core architectural styles, both vulnerable to severe data distortion (Table 1):

Table 1: Predictive systems

Predictive Modality	**Operational Goal**	**Data Sources Used**	**Inherent Bias Risk**
Place-Based Forecasting	Pinpoints geographic locations ("hot spots") where property or violent crime is expected to strike.	Historical crime maps, weather patterns, commercial zoning, transit layouts.	Replicates spatial patterns from stop-and-frisk tactics and systemic neighborhood segregation.
Person-Based Forecasting	Ranks individuals on their likelihood to commit a future offense or become a crime victim.	Arrest histories, social network connections, family ties, regional drug data.	Erases the presumption of innocence; profiles citizens based on associations rather than overt criminal actions.
Predictive Modality	Operational Goal	Data Sources Used	Inherent Bias Risk
Place-Based Forecasting	Pinpoints geographic locations ("hot spots") where property or violent crime is expected to strike.	Historical crime maps, weather patterns, commercial zoning, transit layouts.	Replicates spatial patterns from stop-and-frisk tactics and systemic neighborhood segregation.
Person-Based Forecasting	Ranks individuals on their likelihood to commit a future offense or become a crime victim.	Arrest histories, social network connections, family ties, regional drug data.	Erases the presumption of innocence; profiles citizens based on associations rather than overt criminal actions.

4. Prominent Case Studies & Legal Repercussions

Real-world deployments demonstrate the tangible civil rights violations caused by biased systems:[8,9]

- **The Chicago SSL Dataset**: Chicago's Strategic Subject List (SSL) intended to flag individuals at extreme risk of gun violence. Independent performance reviews from the city's Inspector General found the risk categorization framework entirely unreliable, leading to its cancellation after intense local pushback.

- **The COMPAS Recidivism Model**: Though used for sentencing and bail rather than street patrols, the COMPAS algorithm became a landmark study in algorithmic failure. It routinely assigned significantly higher recidivism risk scores to Black defendants than White defendants, even when their subsequent criminal paths proved otherwise.

- **The HART Constabulary System**: Durham Constabulary's Harm Assessment Risk Tool (HART) was deployed in the UK using historical data that disproportionately penalized poor socioeconomic classes. Civil rights organizations ultimately forced its retirement, arguing that predictive software directly violates the foundational right to a fair trial.

5. Mitigation and Evolving Frameworks

Addressing bias requires moving beyond simple software adjustments to install deep public oversight:[10]

1. **Regulatory Guardrails**: Comprehensive laws explicitly categorize algorithmic policing as a high-risk operation, enforcing rigid verification protocols.

2. **Mathematical Recalibration**: Data scientists utilize approaches like Conditional Score Recalibration (CSR) and demographic parity metrics to actively adjust risk scores

[8] Almasoud, A.S., Idowu, J.A. Algorithmic fairness in predictive policing. AI Ethics, 2025, **5**, 2323–2337. https://doi.org/10.1007/s43681-024-00541-3

[9] Chiara G. Predictive policing and predictive justice: Ethics, data protection, and the AI act. Computer Law & Security Review, 2026, 61,106282, 1-9

[10] Gültekin-Várkonyi, Dr. Gizem, Predictive Policing and Bias in a Nutshell: Technical and Practical Aspects of Personal Data Processing for Law Enforcement Purposes (June 5, 2022). Digital Criminal Justice: A Studybook Selected Topics for Learners and Researchers. Krisztina Karsai, Adem Sözüer, Liane Wörner (eds.). Istanbul, 2022, https://ssrn.com/abstract=4238774 or http://dx.doi.org/10.2139/ssrn.4238774

across age groups and ethnic populations.

3. **Mandatory Public Audits**: Civil advocacy groups require full transparency regarding training parameters, system vendor contracts, and routine independent accuracy evaluations.

4. **Community Co-Design**: Shifting power structures to let system-impacted communities directly oversee data boundaries ensures algorithms are deployed to address root structural needs rather than just optimizing containment and surveillance.

6. Prominent Software Tools Associated with Algorithmic Bias

Several specific AI software platforms and machine learning tools have faced heavy criticism, investigative audits, or legal bans due to documented algorithmic bias in predictive policing. In-spite of short coming the enforcement agencies relays on AI platforms as one the tools in unfolding criminal intent (Table 2).

Table 2: Software Tools Associated with Algorithmic Bias

Software Platform	Developer / Organization	Primary Predictive Modality	Documented Bias & Controversy
PredPol (later rebranded as Geolitica)	PredPol Inc. / SoundThinking	Place-based 500x500 foot geographic "hot spots".	Systematic Spatial Profiling: A landmark investigation by "The Markup"[11] revealed that the code disproportionately targeted low-income, Black, and Latino neighborhoods. Critics called it a "crime production algorithm" because increased patrols created an artificial feedback loop. It failed to accurately predict crimes in multiple cities and matching was less than 1% of actual reported crimes at Plainfield NJ which ceased operations at the end of 2023.
COMPAS	Northpointe (now equivant)	Person-based recidivism risk scoring for bail and	Racial Disparity in Error Rates: A foundational ProPublica study proved the software was biased

[11] https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes

		sentencing.	against Black defendants. It was twice as likely to incorrectly flag Black individuals as "high risk" for future crimes (false positives) compared to White individuals, while simultaneously making more false-negative errors for White defendants.
Gotham / Inside	Palantir Technologies	Person & Network-based data aggregation and link analysis.	Pervasive Mass Surveillance: Deployed covertly in cities like New Orleans, this software merged disparate databases-including social media, cell tower logs, and field interviews to map out criminal networks. Civil rights groups showed it essentially codified "guilt by association," over-indexing marginalized populations who interacted frequently with heavily policed family or social circles.
Patternizr	New York Police Department (NYPD)	Crime-pattern analysis linking separate active cases.	Codification of Stop and Frisk: Built internally by the NYPD to find hidden connections across theft and grand larceny records. Because it relied on legacy suspect descriptions recorded during the height of the department's racially discriminatory "Stop-and-Frisk" era, the AI consistently replicated those skewed demographic patterns in its automated search results.
HART (Harm Assessment Risk Tool)	Durham Constabulary / University of Cambridge	Person-based custody and re-offending risk profiles.	Socioeconomic Punishment: Used by UK police, this software relied heavily on marketing data and regional postal codes to predict recidivism. It directly penalized individuals living in impoverished zip codes, embedding structural class bias into custody recommendations before it was eventually restricted.

7. Why These Specific Systems Fail Safely

The algorithmic bias embedded across these software systems stems from a shared technical

reality: AI cannot separate criminal behavior from policing behavior. If a tool is trained on an arrest database, it does not learn where crime happens; it learns where arrests happened in the past. Because marginalized or low-income neighborhoods have historically faced disproportionately heavy police deployment, the AI mistakes this systemic human footprint for objective, mathematical truth.

8. Practice of AI-driven Predictive Policing in India

In India, the adoption of AI-driven predictive policing and algorithmic surveillance has expanded rapidly across various state police departments. Because India lacks an anchoring statutory framework or data protection laws specific to law enforcement AI, these systems face severe criticism from digital rights groups such as Software Freedom Law Centre and the Vidhi Centre for Legal Policy for codifying caste, religious, and socioeconomic biases.[12,13,14,15]

Table 3: Prominent AI & Predictive Software Used by Indian Law Enforcement

Platform / Software Name	Deploying Agency / State	Operational Function
CMAPS (Crime Mapping, Analytics and Predictive System)	Delhi Police	Place-based hot-spot mapping. Uses real-time helpline data and FIRs to identify criminal clusters.
Prophecy Alethia / Prophecy Guardian	Developed by Innefu Labs (used by multiple national agencies)	Predictive policing & Intelligence Fusion. Integrates open-source intelligence (OSINT), predictive algorithms, and case linking.
MahaCrimeOS AI	Maharashtra Police (Co-developed with Microsoft)	Generative AI & Predictive Analytics. Leverages Azure OpenAI and Retrieval-Augmented Generation (RAG) to link cases and

[12] Tracing The Rise of Predictive Policing in India, Febuary 26, 2026. https:// sflc.in/tracing-the-rise-of-predictive-policing-in-india/

[13] Akriti B. India's Tryst With Predictive Policing: The need for safeguards for greater transparency. vidhilegalpolicy.in/blog/indias-tryst-with-predictive-policing/

[14] Purushraj P. Algorithmic Policing in India: The Case for a Governing Framework. Centre for Digital Societies at Observer Research Foundation, May 29, 2026, https://www.orfonline.org/expert-speak/algorithmic-policing-in-india-the-case-for-a-governing-framework

[15] Hyderabad Police launch AI-powered social media surveillance platform 'SOCEYE'. May 18, 2026, The Hindu Bureau Hyderabad. https://www.thehindu.com/news/national/telangana/hyderabad-police-launch-ai-powered-social-media-surveillance-platform-soceye/article70994663.ece

		analyze regional crime trends.
TSCOP & SOCEYE	Telangana / Hyderabad Police	Real-time facial recognition & Social Media Surveillance. TSCOP enables field biometric matching; SOCEYE uses AI to scan social media for public sentiment and "threat narratives".
PAIS (Punjab) & Trinetr (Uttar Pradesh)	Punjab Police & UP Police	Person-based biometric matching. PAIS processes over 390,000 criminal databases and voice samples; Trinetr holds over 900,000 digital records.

9. Unique Structural Drivers of Algorithmic Bias in India

The amplification of algorithmic bias in India differs significantly from Western contexts due to unique legal and social structures:[16]

- **The Trap of "Dirty Data"**: Indian predictive software trains on First Information Reports (FIRs) and arrest logs. In India, structural prejudices mean Dalits, Adivasis, Muslims, and Denotified Tribes are heavily over-represented in prison populations relative to their actual demographics. The AI treats this systemic over-policing as an objective mathematical reality.

- **Weaponization of Discretionary Powers**: Algorithms that output "high-risk hot spots" are frequently used by local police to justify aggressive preventive actions, such as detentions under Section 151 of the CrPC (Section 170 BNSS 2023). This allows the police to act purely on anticipated algorithmic risk rather than a committed offense.

- **The Legacy of the Criminal Tribes Act**: Though repealed, the historical stigma against "Denotified Tribes" persists. When police manually feed legacy colonial-era surveillance books and local history-sheets into modern databases like PAIS or Trinetr, the AI automates generation-generational discrimination.

[16] Keerthi L., Suresh C. Data-driven justice: Survey on approaches to identify and mitigate biases to build fair artificial intelligent models for Indian justice system. Intelligent Decision Technologies, Online 2025, 19(2) https://doi.org/10.1177/18724981241301281

- **Total Legislative Opacity**: Unlike the EU AI Act, India lacks a statutory framework explicitly regulating law enforcement's use of AI. Furthermore, police departments regularly deny Right to Information (RTI) requests regarding their software's algorithms, citing national security exemptions.

10. AI role in Tamil Nadu Police

The Tamil Nadu Police, particularly the Greater Chennai Police, primarily rely on FaceTagr for facial-recognition-driven tracking, alongside an internal GIS-based Hotspot Analysis System integrated into the freshly deployed CCTNS 2.0 (Crime and Criminal Tracking Network and Systems 2.0). Rather than purchasing an all-in-one Western predictive policing software suite like PredPol, the state department layers localized AI, facial recognition portals, and geographic information systems (GIS) on top of centralized police databases to execute predictive actions. [17]

10.1 Core Software Systems Used by Tamil Nadu Police

A. FaceTagr (Facial Recognition System)

- **What it is**: An AI-powered facial recognition application developed by a Chennai-based startup (*Backbone Technologies*) used extensively by the Tamil Nadu Police.[18]

- **How it works**: Field officers use smartphones to capture photos of individuals on the street and compare them in real time against a centralized suspect database.

- **Algorithmic Bias Concerns**: Real-world tests demonstrate that FaceTagr and similar facial recognition algorithms suffer from severe racial, ethnic, and skin-tone biases, sometimes registering 10 to 100 times higher false-positive rates for darker-skinned individuals compared to lighter-skinned subjects. In practice, this results in dark-skinned individuals from marginalized backgrounds being disproportionately flagged as "suspect matches" during routine street stops.

[17] Soumyarendra B. Tamil Nadu Police nab criminal using face recognition app, questions of safeguards and accountability remain. November 24, 2020. https://www.medianama.com/2020/11/223-tn-police-catch-criminal-using-facial-recognition-issues/

[18] Stalin to launch CCTNS 2.0, digital fingerprint collection system on February 26. The Hindu Bureau Chennai, February 25, 2026. https://www.thehindu.com/profile/author/the-hindu-bureau-14355/

B. CCTNS 2.0 & Digital Fingerprint Collection System (DFCS)

- **What it is**: Launched by the Tamil Nadu government in early **2026** with a ₹124.37 crore budget, CCTNS 2.0 updates legacy police networks with integrated machine learning models.

- **How it works**: The platform consolidates active First Information Reports (FIRs), criminal profiles, and a massive network of biometric signatures across all 17,000 national police stations into an analytical pool.

- **Algorithmic Bias Concerns**: By aggregating historical arrest databases, the software essentially automates historical policing prejudices. The AI uses this data to generate "suspect scores" or predict likely future repeat offenders. This locks generations of heavily-policed communities e.g. Denotified Tribes and low-income migrants, into an automated feedback loop of perpetual state suspicion.

C. GIS Hotspot Analysis and Predictive Mapping

- **What it is**: An evidentiary predictive tool implemented by the Greater Chennai Police.

- **How it works**: The software ingests five years of historical FIR records, Community Service Register (CSR) logs, distress calls from emergency helplines (100, 112, 181), and GPS data from the Kavalan SOS app[19]. It runs algorithms to generate localized geographic crime maps, predicting where incidents like chain-snatching or house burglaries are likely to repeat.

- **Algorithmic Bias Concerns**: Because call logs and police reports heavily skew toward wealthier enclaves reporting on lower-income neighborhoods, or reflect areas where police already maintain heavy station density, the algorithm treats these structural biases as natural law enforcement priorities. It directs patrol vehicles into marginalized neighborhoods, creating an artificial surge in localized arrests that feeds back into the algorithm as "proof" of its predictive accuracy.

10.2 Key Systemic Fallout & Incidents in Tamil Nadu

- **Massive Data Security Leaks**: In **2024**, a massive data security leak exposed the Tamil Nadu Police's Facial Recognition Portal database, leaking the personal data, biometrics,

[19] https://kavalan-sos.en.aptoide.com/app

and photos of thousands of unconvicted suspects. Digital liberties advocates noted that exposing unregulated blacklists creates a breeding ground for systemic profiling.

- **Political & Protest Profiling**: During regional demonstrations (such as the CAA protests), Chennai field officers were documented scanning the faces of peaceful protestors waiting at transit terminals using AI tools. Privacy watchdogs argue that using predictive tracking appends an algorithmic "threat label" to political dissidents, chilling freedom of assembly.

- **Complete Legislative Blind Spot**: The state's digital policing frameworks operate without a governing statutory law passed by the assembly, preventing independent audit access to FaceTagr's code or CCTNS 2.0's mathematical parameters.

11. Indian Judiciary Ban on Algorithmic Decisions in the Judicial System

The Indian judiciary has acted directly to prevent predictive software from replacing human judgment in criminal justice and litigation.[20]

11.1 Draft Regulations for Use of AI in Courts, 2026 (Supreme Court of India)

Following an incident in 2025 where a trial court relied on fabricated, AI-generated legal citations, the Supreme Court issued strict, explicit restrictions on software autonomy. The regulations institute a blanket prohibition on algorithmic judicial decision-making and predicting bail outcomes. While AI software can handle legal translations (via tools like SUVAS) or case-load transcription, it is legally barred from determining a citizen's guilt, assessing witness credibility, or deciding incarceration.[21]

11.2 Preventing Discriminatory Bias (Articles 14 & 15)

Predictive policing software can inadvertently reinforce structural discrimination if its training datasets focus primarily on historically over-policed neighborhoods or marginalized demographics.

[20] Ashish B., Krishna D.C. The bench has spoken on AI. Now the law school must respond. The Hindu, June 29, 2026. https://www.thehindu.com/education/the-bench-has-spoken-on-ai-now-the-law-school-must-respond/article71154102.ece

[21] Supreme Court of India, Draft Regulation for use of Artificial Intelligence in Courts, 2026 https://cdnbbsr.s3waas.gov.in/s3ec0490f1f4972d133619a60c30f3559e/uploads/2026/06/2026060342.pdf

11.3 The Core Clash

The AI Act is a product liability and market harmonization regulation. The Law Enforcement Directive (LED) focuses strictly on fundamental rights and data processing during criminal investigations. When law enforcement agencies use AI, these two distinct frameworks collide.

Major Gaps and Inconsistencies

- **Different Standards of Risk**: The AI Act classifies certain policing tools (like recidivism prediction) as "high-risk," requiring strict technical compliance. However, the LED allows Member States wider discretion based on "strict necessity," leading to uneven protection across borders.

- **Enforcement Blind Spots**: Market surveillance authorities enforce the AI Act. Data Protection Authorities (DPAs) enforce the LED. It remains unclear who takes precedence when a police AI tool violates both product safety and data privacy laws.

- **National Security Loophole**: The AI Act explicitly excludes national security from its scope. However, the line between ordinary law enforcement (covered by the LED) and national security is often blurred, creating a massive regulatory vacuum.

- **Biometric Surveillance Confusion**: The AI Act bans real-time remote biometric identification (like facial recognition) in public spaces, but includes broad exceptions for law enforcement. This conflicts with the LED's stricter requirement for authorization only under "essential public interest."

The Resulting Impact

- **Legal Uncertainty**: Police forces lack clear guidelines on which rules take priority during active investigations.

- **Accountability Deficit**: Victims of biased AI algorithms face a fragmented path to seek legal remedies.

- **Fragmented Implementation**: Different EU countries interpret the overlap differently, weakening the goal of a unified digital market.

12. Supreme Court of India's Draft Regulations

Digital Personal Data Protection (DPDP) Act, 2023 regulate data processing, they lack explicit AI governance clauses. To plug this gap, the Supreme Court Draft AI Rules (2026) have recently been introduced, which explicitly prohibit AI from introducing bias or dictating judicial/bail outcomes.

The Supreme Court of India's Draft Regulations for Use of Artificial Intelligence in Courts, 2026, establish a comprehensive legal framework that mandates AI must remain strictly assistive and subservient to human judgment. Released by the Supreme Court's AI Committee (chaired by Justice P.S. Narasimha), the regulations aim to improve efficiency while protecting the constitutional core of justice delivery.[22,23]

12.1 Scope of Application

The rules uniformly apply nationwide to the following judicial bodies:

- The Supreme Court of India
- All High Courts
- All Subordinate/District Courts
- Tribunals and Statutory Commissions exercising adjudicatory functions

Courts, lawyers, and litigants are encouraged to adopt AI for administrative and support tasks, including:

- **Legal Research & Verification**: Cross-referencing laws and verifying citations.
- **Drafting & Document Assistance**: Assisting with legal summaries and administrative templates.
- **Language Tools**: Real-time transcription of proceedings and translation of judgments.

[22] https:// www.scconline.com/blog/post/2026/06/05/sc-issues-draft-ai-regulations-for-courts/
[23] https://indianexpress.com/article/legal-news/supreme-court-ai-regulations-2026-draft-rules-courts-ban-ai-decisions-10723503/

- **Court Administration**: Automating case management, creating cause lists, and scheduling hearings.

12.2 Absolute Prohibitions (Regulation 20)

To prevent algorithmic errors or "hallucinations" from corrupting justice, the draft draws non-negotiable red lines. AI is **completely barred** from:

- **Core Adjudication**: Reaching judicial outcomes, deciding verdicts, or writing final judgments.
- **Liberty Determinations**: Evaluating bail eligibility, sentencing, or deciding on personal freedoms.
- **Risk Scoring**: Using predictive software to calculate recidivism or flight risks.
- **Profiling**: Inferring the credibility or future conduct of witnesses, litigating parties, or lawyers.
- **Surveillance**: Tracking or continuously monitoring judges, advocates, or litigants.
- **Opaque "Black-Box" Systems**: Using non-explainable AI in any context that affects a citizen's lawful rights.

12.3 Key Accountability Mandates

1. Mandatory Disclosure Regime

Any party or lawyer utilizing AI to prepare pleadings, documents, submissions, or evidence **must explicitly disclose its use** to the court via a declaration or certificate at the time of filing.

2. Human Primacy & Personal Liability

The principle of "human primacy" dictates that humans remain entirely legally responsible.

- **For Lawyers**: If an AI hallucination generates fake or inaccurate case citations or facts, the lawyer faces full liability and cannot blame the software.

- **For Judges**: Judges cannot rely on technical errors or system malfunctions to excuse miscarriages of justice or incorrect choices.

3. Data Protection and Fairness

AI integrations must strictly adhere to the Digital Personal Data Protection Act 2023. Furthermore, the algorithms must undergo strict technical and ethical audits to ensure they do not perpetuate bias related to caste, gender, religion, or economic status.

12.4 Proposed Governance Framework

The draft framework structures a multi-tiered administrative layer to manage safety:

- **Permanent Apex Body**: Established at the Supreme Court level to guide policy, Innovation, and standard-setting. It includes Supreme Court and High Court judges, technical specialists from the Ministry of Electronics and Information Technology (MeitY), and AI academic experts.

- **Monitoring Systems**: Mandates controlled testing environments, an "AI Register" documenting deployed legal tools, and a centralized "AI Incident Database" to record failures.

13. Case Law Analysis: Based on USA Court Verdict

State v. Loomis[24]

In the State v. Loomis (2016) the defendant, Eric Loomis, was arrested in connection with a drive-by shooting and subsequently pleaded guilty to attempting to flee an officer and operating a vehicle without consent. During the sentencing phase, the trial court reviewed a pre-sentence investigation report that incorporated a recidivism risk score calculated by COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), a proprietary software tool built by Northpointe, Inc. (now Equivant). The algorithm assigned Loomis a high risk of recidivism across three distinct categories, and the trial judge cited this score explicitly when handing down a sentence of six years of imprisonment.

[24] State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

Loomis appealed the ruling, asserting that the judicial reliance on a proprietary, closed-source algorithm violated his constitutional right to Due Process under the Fourteenth Amendment. His challenge rested on two arguments:

1. The Infringement of the Right to Accuracy: Because Northpointe kept its source code and underlying weighting metrics protected as proprietary trade secrets, Loomis could not review, test, or challenge the inner logic of the software that contributed to his incarceration.

2. Unconstitutional Use of Gender: The algorithm explicitly factored gender into its predictive equations, violating the right to individualized sentencing.

The Wisconsin Supreme Court denied Loomis's post-conviction motion, ruling that the use of COMPAS at sentencing does not violate due process. The court reasoned that because the algorithm provided an assessment based on aggregates rather than determining the final sentence outright, the judge retained ultimate discretionary authority. However, recognizing the inherent structural dangers of automated systems, the court issued mandatory cautionary warnings that must accompany any future judicial use of COMPAS. Judges must be explicitly informed that:

- The proprietary design bars code verification.

- It calculates scores based on group averages rather than individual traits.

- Studies (such as the landmark *ProPublica* analysis) have shown the tool to produce systemic, racially disparate false-positive rates against Black defendants.

State v. Loomis stands as a pivotal legal anchor point, showing the deep institutional friction between private corporate trade secrets and an individual's fundamental constitutional right to transparent, open justice. It can be concluded that algorithmic bias in predictive policing cannot be resolved by simple code refactoring or minor technical adjustments. Because these models are built on historical data that reflects systemic disparities, they naturally codify and perpetuate those inequalities under a veneer of mathematical objectivity. True mitigation demands structural intervention: establishing independent algorithmic audits, implementing rigorous socio-legal oversight, and building clear legislative boundaries like those outlined in the EU AI Act. Without these structural boundaries, predictive software will continue to

convert historical discrimination into automated, self-reinforcing mandates for targeted surveillance.

United States v. Lopez[25]

The case related to United States v. Lopez (2020) deals with the issue where the defendant's vehicle was stopped by law enforcement officers based primarily on data generated by a place-based predictive policing algorithm, which had designated the specific city block as a high-density "hot spot" for violent crime. During the stop, officers discovered a firearm, leading to federal charges. Lopez moved to suppress the evidence, arguing that a software program's generalized statistical prediction does not constitute the individualised reasonable suspicion required for a physical stop under the Fourth Amendment.

The Federal District Court ruled that while an algorithmic hot-spot designation can be considered part of the "totality of the circumstances" (akin to an officer's knowledge of a "high-crime area"), an algorithmic score or prediction alone can never substitute for individualized, observable suspicion. The court emphasized that allowing automated probability models to justify physical detentions risks exposing entire communities to unconstitutional, circular over-policing.

Legal Aid Society v. New York City Police Department[26]

The Legal Aid Society filed a Freedom of Information Law (FOIL) lawsuit against the NYPD to compel the disclosure of records detailing the department's acquisition and deployment of predictive policing software, including facial recognition systems and historic crime-mapping tools. The NYPD had resisted these requests, citing operational secrecy and vendor-dictated proprietary protections. The court ruled against the NYPD's broad opacity, forcing the disclosure of procurement records and technical vendor contracts. This litigation directly exposed the "black box" issue where private tech firms dictate public surveillance terms. The backlash from this case ultimately led to the passage of New York City's POST Act (Public Oversight of Surveillance Technology), which legally mandates transparency reports and civil rights impact assessments before any predictive policing tool can be deployed.

[25] United States v. Lopez, 443 F. Supp. 3d 991 (D. Minn. 2020).
[26] Legal Aid Society v. New York City Police Department, Index No. 154321/2020 (N.Y. Sup. Ct. 2020).

14. Case Law Analysis: Based on Argentina Court Verdict

CELS v. Ministry of Justice and Security of Buenos Aires[27]

The CELS v. Ministry of Justice and Security of Buenos Aires (2022 and 2023) case deals with the suit filed by the Civil rights organizations which sued the city of Buenos Aires over its implementation of the Fugitive Facial Recognition System (CONARC), a live automated profiling tool designed to scan public spaces and cross-reference biometric data with a government database of wanted individuals. Discovery revealed that the underlying system had executed millions of unauthorized biometric requests, generating sweeping false-positive matches that disproportionately misidentified innocent citizens.

The court ordered the immediate suspension of the algorithmic system, ruling that operating an un-audited, opaque tracking model in public spaces violates constitutional rights to privacy and freedom of movement. The court barred its return until the government established independent oversight and rigorous algorithmic impact assessments—establishing an international precedent that automated police surveillance cannot bypass fundamental civil liberties.

15. Case Law Analysis: Based on Indian Court Verdict

Lt. Col. Nitisha v. Union of India (Applied to Algorithmic Contexts)[28]

Lt. Col. Nitisha v. Union of India (2021) case dealt with employment discrimination and its framework has become the primary constitutional benchmark used by Indian courts and digital rights groups to challenge opaque policing algorithms (such as Delhi's Project Panoptic and automated facial recognition deployments). The Law entrenched the doctrine of indirect discrimination under Indian equality law. It ruled that institutional frameworks or neutral-looking metrics are unconstitutional if they result in a disproportionate, adverse impact on a specific protected class. If a state agency deploys software were its algorithmic relevance disproportionately targets specific pincodes, low-income groups, or marginalized communities based on biased historical arrest data, the deployment can be challenged as a violation of Article 14 and 15 of the Constitution of India. The Supreme Court adopted a two-step test: a court

[27] Centro de Estudios Legales y Sociales (CELS) v. Ministerio de Justicia y Seguridad, Expte. 43812/2021-0 (Contencioso Administrativo y Tributario de la CABA, 2022; upheld on appeal, 2023)
[28] Lt. Col. Nitisha v. Union of India, (2021) 15 SCC 125.

must evaluate whether a facially neutral measure disproportionately burdens a protected group, and whether it reinforces historical structural disadvantages. In the context of algorithmic policing, this framework means that even if a law enforcement algorithm is designed without explicit bias, if its outputs yield a demographically skewed pattern of false positives, it is unconstitutional under Indian law. Litigants can rely on "judicial common sense" to challenge these outputs, shifting the burden of proof to the state to prove the model's accuracy.

Sadhan Haldar v. The State NCT

In the Sadhan Haldar (2017) case[29] concern the tracking of a missing 13-year-old child, the Delhi High Court authorized law enforcement to deploy an Automated Facial Recognition System (AFRS) to trace and reunite missing or trafficked children. However, this case is highly significant in tech-policy jurisprudence as a premier example of "function creep". While the court sanctioned the software strictly for child rescue, subsequent Right to Information (RTI) disclosures revealed that the Delhi Police systematically widened the scope of this judicial permission to justify deploying facial recognition for routine criminal investigations, mass surveillance, and crowd control.

Karthik v. The State of Tamil Nadu

The Karthik case[30] addresses procedural fairness and the abuse of preventive detention based on arbitrary law enforcement mechanics. The Madras High Court set aside a detention order, establishing that algorithmic recommendations, administrative automated profiling, and mechanical police reports cannot substitute individualized "probable cause" or reasonable suspicion. The court emphasized that using opaque, generic risk-assessment profiles without providing the accused with the exact material grounds for defence severely violates procedural due process established by law.

Digital Rights Foundation v. Union of India & Ors.

The Delhi High Court[31] evaluated the state's construction of an aggressive, bulk internet and telecom dragnet surveillance system. The court held that algorithmic profiling and the

[29] Sadhan Haldar v. The State NCT of Delhi & Ors. W.P. (Crl.) 1560/2017
[30] Karthik v. The State of Tamil Nadu MC No. 45/2025/B7 / Criminal Revision Case No. 634 of 2026 (Madras High Court).
[31] Digital Rights Foundation v. Union of India & Ors. (NCT of Delhi) WP (C) No. 4812/2024

harvesting of digital footprints without explicit statutory authorization or consent constitutes a prima facie violation of personal liberty and the Right to Privacy under Article 21. It laid down that machine-driven, bulk intelligence tools cannot process citizen datasets without setting up independent, judicial oversight bodies to prevent systemic data bias and profiling abuse. The Delhi High Court held that algorithmic profiling without explicit consent is a prima facie violation of Article 21's right to digital footprints and personal liberty.

Pooja Ramesh Singh v. Jammu and Kashmir Bank case[32]

Pooja Ramesh Singh case the Supreme Court of India completely quashed the orders of the National Company Law Tribunal (NCLT) and the National Company Law Appellate Tribunal (NCLAT) after finding that the initial tribunal judgment relied heavily on fake, non-existent, and hallucinated legal precedents generated by Artificial Intelligence (AI). The landmark ruling was delivered by a Division Bench comprising Justices P.S. Narasimha and Alok Aradhe in this case, which originated from an insolvency dispute involving Essel Infraprojects.

The "Methyl Isocyanate" of the Legal System

In a strongly worded judgment, the bench likened the entry of unverified, fake AI hallucinations into court rulings to a silent toxic leak, stating it is "like the release of methyl isocyanate in the province of law and justice: invisible, insidious, and catastrophic by the time anyone notices". The Court established strict directives regarding the boundaries of technology in the judiciary:

- "No Decision" Status: The apex court ruled that any judicial determination relying on fake or hallucinated AI material is entirely void and "no decision at all in the eyes of the law," regardless of whether it significantly impacted the outcome.
- Professional Misconduct: Citing unverified AI-generated judgments is officially labeled as a professional misconduct for advocates. It is deemed an equally severe lapse for judges to adopt them without independent verification.
- Zero-Tolerance Directive: Courts and tribunals across the country must exercise a zero-tolerance approach toward unverified AI submissions. The underlying case has been

[32] Pooja Ramesh Singh v. Jammu and Kashmir Bank Ltd. and Anr. 2026 INSC 668

restored to the NCLT to be decided afresh strictly on verified facts.

16. Conclusion

The transition toward algorithmic governance in the criminal justice system was championed as a technological remedy for human bias, promising an era of objective, data-driven public safety. Predictive policing software is often marketed as an objective, tech-driven solution to crime reduction, yet it fundamentally operates as an automated mirror of historical inequality. By treating flawed police data as neutral truth, these algorithms institutionalize systemic bias under the guise of mathematical objectivity. The resulting hyper-surveillance of minority neighborhoods erodes public trust and compromises constitutional protections against discriminatory policing. Moving forward, technology cannot serve as a substitute for systemic reform. Minimizing algorithmic harm requires strict regulatory oversight, independent third-party audits, and a fundamental shift away from punitive predictions toward community centered public safety models.

References

1. Perry WL. Predictive policing: The role of crime forecasting in law enforcement operations. Rand Corporation; 2013.

2. Perego B. The impact of predictive policing system on civil rights. Biolaw J 2019.

3. Söderholm S. Fundamental rights control when implementing predictive policing–a European perspective. Peking Univ Law J 2023;11(1):91–104.

4. Amit S., Mayuri S. Role of Artificial Intelligence in Predictive Policing: A Double-Edged Sword in Combating Crimes against Women, NeuroQuantology, 2022; 20(1), 1430-1452. DOI: 10.48047/NQ.2022.20.1.NQ22435

5. https://www.privacyinternational.org/sites/default/files/2021-06/predictive_policing.pdf

6. The European Union's AI Act 2024/1689. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

7. Eylül B, Yueqi C, Sarah G, et al., Predictive Policing or Predictive Prejudice? A Study of the Legal, Historical and Ethical Implications of AI in Policing, 2025. https://www.oxjournal.org/predictive-policing-or-predictive-prejudice/

8. Almasoud, A.S., Idowu, J.A. Algorithmic fairness in predictive policing. AI Ethics, 2025, **5**, 2323–2337. https://doi.org/10.1007/s43681-024-00541-3

9. Chiara G. Predictive policing and predictive justice: Ethics, data protection, and the AI act. Computer Law & Security Review, 2026, 61,106282, 1-9

10. Gültekin-Várkonyi, Dr. Gizem, Predictive Policing and Bias in a Nutshell: Technical and Practical Aspects of Personal Data Processing for Law Enforcement Purposes (June 5, 2022). Digital Criminal Justice: A Studybook Selected Topics for Learners and Researchers. Krisztina Karsai, Adem Sözüer, Liane Wörner (eds.). Istanbul, 2022, https://ssrn.com/abstract=4238774 or http://dx.doi.org/10.2139/ssrn.4238774

11. https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes

12. Tracing The Rise of Predictive Policing in India, Febuary 26, 2026. https:// sflc.in/tracing-the-rise-of-predictive-policing-in-india/

13. Akriti B. India's Tryst With Predictive Policing: The need for safeguards for greater transparency. vidhilegalpolicy.in/blog/indias-tryst-with-predictive-policing/

14. Purushraj P. Algorithmic Policing in India: The Case for a Governing Framework. Centre for Digital Societies at Observer Research Foundation, May 29, 2026, https://www.orfonline.org/expert-speak/algorithmic-policing-in-india-the-case-for-a-governing-framework

15. Hyderabad Police launch AI-powered social media surveillance platform 'SOCEYE'. May 18, 2026, The Hindu Bureau Hyderabad. https://www.thehindu.com/news/national/telangana/hyderabad-police-launch-ai-powered-social-media-surveillance-platform-soceye/article70994663.ece

16. Keerthi L., Suresh C. Data-driven justice: Survey on approaches to identify and mitigate biases to build fair artificial intelligent models for Indian justice system. Intelligent Decision Technologies, Online 2025, 19(2) https://doi.org/10.1177/18724981241301281

17. Soumyarendra B. Tamil Nadu Police nab criminal using face recognition app, questions of safeguards and accountability remain. November 24, 2020. https://www.medianama.com/2020/11/223-tn-police-catch-criminal-using-facial-recognition-issues/

18. Stalin to launch CCTNS 2.0, digital fingerprint collection system on February 26. The Hindu Bureau Chennai, February 25, 2026. https://www.thehindu.com/profile/author/the-hindu-bureau-14355/

19. https://kavalan-sos.en.aptoide.com/app

20. Ashish B., Krishna D.C. The bench has spoken on AI. Now the law school must respond. The Hindu, June 29, 2026. https://www.thehindu.com/education/the-bench-has-spoken-on-ai-now-the-law-school-must-respond/article71154102.ece

21. Supreme Court of India, Draft Regulation for use of Artificial Intelligence in Courts, 2026 https://cdnbbsr.s3waas.gov.in/s3ec0490f1f4972d133619a60c30f3559e/uploads/2026/06/2026060342.pdf

22. https:// www.scconline.com/blog/post/2026/06/05/sc-issues-draft-ai-regulations-for-courts/

23. https://indianexpress.com/article/legal-news/supreme-court-ai-regulations-2026-draft-rules-courts-ban-ai-decisions-10723503/

24. State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

25. United States v. Lopez, 443 F. Supp. 3d 991 (D. Minn. 2020).

26. Legal Aid Society v. New York City Police Department, Index No. 154321/2020 (N.Y. Sup. Ct. 2020).

27. Centro de Estudios Legales y Sociales (CELS) v. Ministerio de Justicia y Seguridad, Expte. 43812/2021-0 (Contencioso Administrativo y Tributario de la CABA, 2022; upheld on appeal, 2023)

28. Lt. Col. Nitisha v. Union of India, (2021) 15 SCC 125.

29. Sadhan Haldar v. The State NCT of Delhi & Ors. W.P. (Crl.) 1560/2017

30. Karthik v. The State of Tamil Nadu MC No. 45/2025/B7 / Criminal Revision Case No. 634 of 2026 (Madras High Court).

31. Digital Rights Foundation v. Union of India & Ors. (NCT of Delhi) WP (C) No. 4812/2024

32. Pooja Ramesh Singh v. Jammu and Kashmir Bank Ltd. and Anr. 2026 INSC 668