SUPREMO AMICUS

INDIA'S FIRST AI INTEGRATED LAW JOURNAL

Peer Reviewed, Refereed and Open access Journal

- Available in 331+ International Libraries
- Indexed at 32 Databases

ISSN NO. 2456-9704
Volume 10 Issue 2
www.supremoamicus.org

DISCLAIMER

The information presented in this article is intended for general informational and educational purposes only. While every effort has been made to ensure that the content is accurate, up-to-date, and reliable at the time of publication, the editorial board and publisher make no representations or warranties of any kind, express or implied, regarding the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

The views and opinions expressed in this article are those of the author and are based on personal research, experience, and interpretation. They do not necessarily reflect the official policy, position, or opinions of any affiliated organization, institution, or entity.

This article is not intended to serve as professional advice of any kind. The editorial board and publisher shall not be held liable for any errors or omissions in the content, nor for any losses, injuries, or damages arising from the use of or reliance on this information.

ABOUT THE JOURNAL

Supremo Amicus is an online, peer-reviewed international journal devoted to the interdisciplinary fields of law and science. In an era marked by rapid technological progress and evolving legal frameworks, the journal seeks to bridge the gap between these two dynamic domains by offering comprehensive and critical insights into their various aspects. The journal places a strong emphasis on contemporary advancements, emerging trends, and the complex challenges faced by both the legal community.

The primary objective of the journal is to encourage and promote original, high-quality research. It is committed to publishing well-researched, analytically sound, and thought-provoking articles that adhere to rigorous academic standards. Each submission undergoes a thorough peer-review process to ensure authenticity, relevance, and scholarly integrity. In doing so, the journal maintains its commitment to excellence and credibility.

In addition to fostering research, the journal aims to make complex ideas accessible and engaging for a diverse readership. It strives to present content that is not only intellectually enriching but also clearly written and reader friendly.

Furthermore, the journal is committed to promoting interdisciplinary collaboration and global engagement. It welcomes diverse perspectives from contributors across different regions and backgrounds, thereby enriching the quality and scope of discussions presented within its pages.

With this vision we proudly present Supremo Amicus to our readers.

-Editorial Team
Supremo Amicus

REIMAGINING CRIMINAL JUSTICE: THE ROLE OF AI IN FORENSIC INVESTIGATION AND LEGAL DECISION-MAKING INDIAN CRIMINAL JURISPRUDENCE

By *Vishvendra Raj Malik*
Assistant Professor *at COER University, Roorkee, Uttarakhand*

By *Prof. (Dr.) Jageshwar Nath Singh*
Professor *at COER University, Roorkee, Uttarakhand*

By *Dr. Sandeep Kumar*
Associate Professor *at COER University, Roorkee, Uttarakhand*

Abstract

Full-fledged artificial intelligence (AI) research is no longer some ancillary accessory for criminal justice administration; it is rapidly growing into the central element of how evidence gets generated, risk being assessed and judicial or quasi-judicial decisions made. In this article, I sort of critically review the increasing footprints of AI across two closely linked spaces evidence generation in forensic investigation and courtroom decision-making. It also asks whether AI's promise of better efficiency, and this alleged objectivity in particular, can actually be made compatible with constitutional guarantees like due process, equality, and a fair trial. It draws from comparative developments in the United States, the European Union and India, and it places the discussion alongside jurisprudence that is still taking shape from[1] the paper argues that 'intelligent justice' must be understood not as automation of judgment but as an evidentiary and decisional aid subordinate to human oversight. It traces the doctrinal challenges posed by algorithmic opacity, examines landmark controversies such a[2] and the ProPublica investigation on recidivism scoring, kind checks and then evaluates India's tech-forward criminal procedure reforms under "the Bharatiya Nagarik Suraksha Sanhita 2023 as well as the Bharatiya Sakshya Adhiniyam 2023". The piece kind of wraps up with a reform program that stays focused on explainability, algorithmic audit and then a legally embedded right to something like meaningful human review, which is, in effect, the core point.[3].

Keywords
Artificial intelligence; forensic science; criminal procedure; due process; algorithmic bias; predictive policing; digital evidence; Bharatiya Nagarik Suraksha Sanhita; explainable AI.

1. Introduction

The criminal justice system has depended on a fairly consistent structure of evidence for over one hundred years, so you know eyewitness testimony and documentary evidence and forensic science all kind of run through peer reviewed methodology which is then tested by opponents through cross examination. So, when you introduce artificial intelligence into that mess it doesn't feel like some little incremental enhancement, more as if the whole thing kind of shifts ever so slightly based on structural changes. AI tools now assist with matching latent fingerprints, reconstructing mutilated faces and sifting through terabytes of digital evidence even forecasting the geography of future crime; they also produce recidivism scores that drive bail, sentencing and parole decisions directly. More recent, interdisciplinary scholarship has called this convergence of data driven forensic tooling with judicial reasoning "Intelligent Justice"; it almost seems to be a paradigm by which human error; bias is curbed and hence minimised, and speed up case disposal.[4] However, the same characteristics that make AI seem so appealing scale with velocity and an

[1] "Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms" (2026) 84 *International Journal of Law, Crime and Justice* 100832".

[2] "*State v. Loomis*, 881 N.W. 2d 749 (Wis. 2016)".
[3] "*Loomis v. Wisconsin*, 137 S. Ct. 2290 (2017)".
[4] *supra* note 1.

ability to scan for patterns far beyond what any human brain could realistically manage are also ones that render it a bit constitutionally hazardous. Machine learning models are already often constructed from something like opaque black boxes, even to the people who built them (a recent programmer controversy), and train on old data that may be decades of discriminatory policing already encoded there as well (an "algorithm problem") so it can all get a little… messy, more or less. And then the output gets delivered to courts like it is authoritative probabilities, not like contestable hunches you can actually challenge.

The article is divided into three parts which highlight some aspects of the use of AI in forensic practices and decision-making process. It touches upon the concerns regarding various legal and moral problems which arise out of the usage of the AI, such as the issues of admissibility, lack of transparency in the decision-making mechanism, and discrimination in the algorithms, providing a number of comparisons between the case laws of the United States, EU and India. Furthermore, it suggests some reforms with regard to the AI, such as introduction of explainability and involvement of people in the decision-making process.

2. Artificial Intelligence in Forensic Investigation

Forensic science has historically depended on the trained perception of human examiners fingerprint analysts, forensic odontologists, and DNA technicians whose conclusions were subject to inter-examiner variability. Machine-learning models, particularly convolutional neural networks, now perform pattern-matching tasks such as fingerprint recognition and bloodstain pattern analysis with reported accuracy gains over manual methods. AI-assisted DNA mixture analysis similarly improves the precision with which complex, multi-contributor biological samples are resolved into individual donor profiles, a task that traditional probabilistic genotyping struggled to perform at scale. In digital forensics, AI-driven triage tools prioritise and cluster vast volumes of seized data messages, metadata, geolocation logs surfacing anomalies that a manual review would take investigators weeks to identify.

Facial and voice recognition systems have moved from experimental deployment to routine investigative use, assisting in suspect identification, missing-person tracing, and the reconstruction of facial structures from decomposed or disfigured remains. Crime-scene reconstruction software builds probabilistic timelines from CCTV, sensor, and forensic data, while computer-vision tools scan baggage and cargo imagery for concealed contraband.[5] These capabilities are not confined to the investigative stage; predictive analytics increasingly precede it, generating geospatial 'hotspot' maps that direct patrol deployment before any offence has occurred.

It is precisely this pre-emptive capacity that forces a return to first constitutional principles. The Latin maxim nemo tenetur seipsum accusare that no person is bound to incriminate himself underlies the constitutional architecture of most criminal justice systems and constrains how invasive biometric and neuro-investigative techniques may be deployed. India's Supreme Court confronted an analogous question in Selvi v State of Karnataka, holding that the involuntary administration of narcoanalysis, polygraph, and Brain Electrical Activation Profile tests amounts to testimonial compulsion and therefore violates "the guarantee against self-incrimination under Article 20(3), read with the right to personal liberty under Article 21, of the Constitution".[6] The Court permitted such techniques only with the subject's free and informed consent, and confined any resulting material to a corroborative rather than a primary evidentiary role. The reasoning in Selvi anticipates precisely the anxiety that AI-driven biometric and neuro-forensic tools now revive: that scientific sophistication cannot, by itself, dispense

[5] *Infra* note 6.

[6] "Selvi v State of Karnataka (2010) 7 SCC 263".

with the voluntariness on which the reliability of testimonial evidence depends.

3. Predictive Policing and Algorithmic Risk Assessment

Predictive policing tools use historical crime data, weighted by spatial and temporal variables, to forecast where and when future offences are statistically likely to occur, thereby guiding the allocation of patrol resources. A parallel application actuarial risk assessment uses similar techniques to forecast an individual's likelihood of reoffending, informing decisions on bail, sentencing, and parole. "The most litigated example of this second category is the Correctional Offender Management Profiling for Alternative Sanctions tool, known as COMPAS, developed by the company Northpointe (later Equivant) and used widely across US state courts.[7] COMPAS generates a recidivism-risk score, on a scale of one to ten, derived from a defendant's responses to a lengthy questionnaire and from criminal-history records, without disclosing the internal weighting of factors, which its developer treats as a trade secret".[8]

In State v. Loomis, the constitutionality of using COMPAS at sentencing was examined. The defendant claimed that relying on a proprietary risk score violated due process because it prevented him from confirming the assessment's scientific validity, denied him an individualized sentence, and might have taken gender into account.[9] The Wisconsin Supreme Court rejected each argument but circumscribed the tool's use: any presentence report containing a COMPAS score must include a written advisement of the tool's limitations, the score cannot be the determinative factor in a sentencing decision, and courts must be alert to the fact that Wisconsin had not cross-validated the instrument against its own offender population.[10] The United States Supreme Court denied certiorari in 2017, leaving the Wisconsin approach as the operative, if contested, precedent.[11] Commentators have observed that the ruling effectively asks trial judges to self-regulate their reliance on a score whose internal logic they cannot inspect an arrangement that places considerable faith in a judicial capacity to discount 'anchoring' effects that behavioural research suggests are difficult to resist in practice.

The empirical foundation for judicial caution was supplied by ProPublica's investigative analysis of over ten thousand COMPAS scores in Broward County, Florida, which found that the tool was substantially more likely to mislabel Black defendants who did not subsequently reoffend as high-risk, while white defendants who did reoffend were disproportionately mislabelled as low-risk. Northpointe disputed the methodology, and subsequent statistical analysis demonstrated that the disagreement reflects a genuine mathematical incompatibility between competing definitions of fairness a tool cannot simultaneously achieve equal predictive accuracy across racial groups and equal false-positive rates unless base recidivism rates are themselves equal, a tension formalised in the algorithmic-fairness literature.[12] This is not merely a technical curiosity: it means that the choice of which fairness metric to prioritise is itself a normative, and ultimately legal, decision that cannot be outsourced to engineers.

[7] "Julia Angwin and others, 'Machine Bias' ProPublica (23 May 2016)".
[8] Ellora Israni, 'Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing' (2017) UCLA Law Review (Online).
[9] *State v. Loomis*, 881 N.W. 2d 749 (Wis. 2016).
[10] "ibid.; see also 'State v. Loomis' (2017) 130 Harvard Law Review 1530 (case comment)".
[11] "Loomis v Wisconsin, 137 S Ct 2290 (2017)".
[12] "Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms" (2026) 84 *International Journal of Law, Crime and Justice* 100832".

4. AI in Legal Research, Case Management, and Judicial Decision-Making

Beyond sentencing, AI has slipped into the wider machinery of legal decision making. Natural language processing tools now scan case law and legal documents at a scale that is basically impossible to do by hand, they supplement legal research and, in a kind of practical turn, enable predictive legal analytics that forecast case outcomes and help shape litigation strategy.[13] Case-management systems use AI to automate administrative processing of evidence and documentation, freeing judicial and prosecutorial time for substantive determinations. Virtual-assistant and online-dispute-resolution platforms extend a measure of legal access to litigants who cannot otherwise afford representation, a development framed within the broader access-to-justice literature as democratising, if imperfect.

The more contested frontier is in adjudicative decision support: tools that help judges and parole boards, weighing bail, sentence length, or release eligibility. Advocates say that more structured, data informed judgments reduce the variance that used to creep in from individual judicial temperament, and maybe even fix unfair outcomes rather than harden them, if these systems are properly audited. In other words, they claim the tech can be a kind of correction instead of a new kind of lock in. Sceptics respond that risk scores tend to import, rather than correct, the racial and socio-economic skew embedded in the arrest and conviction records on which they are trained, and that the appearance of mathematical objectivity can mask rather than resolve contested value judgments about the purposes of punishment.[14] The debate over psychiatric history and parole where hospitalisation records have empirically been shown to reduce parole prospects even absent any adjudicated dangerousness illustrates how proxy variables, whether selected by a human parole board or an algorithm, can produce outcomes untethered from an individual's actual conduct.[15]

5. Comparative Regulatory Frameworks

5.1 The United States: evidentiary gatekeeping under Rule 702

US federal courts assess the admissibility of AI-generated forensic evidence through the existing gatekeeping architecture of the Federal Rules of Evidence, principally Rule 702, which requires that expert or machine-generated evidence rest on reliable principles and methods reliably applied to the facts of the case.[16] Courts differentiate between purely machine-generated records, like GPS coordinates or server timestamps, which are usually authenticated as conventional computer records, and AI "analytical outputs," like predictive risk scores or facial recognition matches, which require the more thorough Daubert-style reliability inquiry before admission. No federal rule categorically excludes AI-derived evidence.[17] The absence of a uniform, mandatory pre-admission disclosure standard for the underlying model as opposed to, for instance, the validation protocols required of pharmaceuticals has been repeatedly flagged as a structural gap, with practice varying widely across the ninety-four federal judicial districts and comparable variance across state courts.[18]

[13] "ibid. Srivastav and others."

[14] "Julia Angwin and others, "Machine Bias" ProPublica (23 May 2016), available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (last accessed on 1 July 2026)".

[15] Srivastav and others *supra* note 12.

[16] "Admissibility of AI-Generated Evidence in U.S. Courts", *AI Legal Authority* (2020), available at: https://www.ailegalauthority.org/admissibility-pathway-702 (last accessed on 2 July 2026)".

[17] *ibid.*

[18] *Ibid.*; see also Sahibpreet Singh and Lalita Devi, "Reliability and Admissibility of AI-Generated Forensic Evidence in Criminal Trials" (*SSRN Working Paper*, 2025).

5.2 The European Union: a risk-tiered regulatory model

Regulation (EU) 2024/1689 (the Artificial Intelligence Act), which went into effect on August 1, 2024, is the first complete legal framework for AI in the world. It divides AI systems into four risk categories: unacceptable, high, limited, and low.[19] In addition to prohibiting the untargeted scraping of facial images from the internet or CCTV footage to create recognition databases, Article 5 of the Act forbids the use of AI systems to evaluate or predict the likelihood of an individual committing a criminal offense based only on profiling or personality-trait assessment, absent objective and verifiable facts directly linked to criminal activity.[20] It is generally forbidden for law enforcement to use real-time remote biometric identification in public areas, with a few specific exceptions that call for prior judicial or independent administrative authorization and a fundamental-rights impact assessment, such as the search for missing persons or the prevention of an impending terrorist threat.[21] AI systems used in the administration of justice including tools that assist judges in researching or interpreting facts and law, or that evaluate the reliability of evidence are separately classified as high-risk, triggering obligations of documentation, human oversight, and conformity assessment before deployment.[22] The EU model therefore does not wait for harm to materialise and be litigated after the fact, as the American gatekeeping model largely does; it pre-emptively forecloses categories of use it judges incompatible with fundamental rights.

5.3 India: technology-forward criminal procedure

"India's three new criminal statutes the Bharatiya Nyaya Sanhita 2023, the Bharatiya Nagarik Suraksha Sanhita 2023 (BNSS), and the Bharatiya Sakshya Adhiniyam 2023 (BSA) replaced the colonial-era Indian Penal Code, Code of Criminal Procedure, and Evidence Act with effect from 1 July 2024", embedding technology at the core of investigation and trial.[23] Section 176(3) BNSS makes forensic examination mandatory, rather than discretionary, for offences punishable by seven years' imprisonment or more, requiring a forensic team to visit the crime scene, collect samples, and videography the entire process.[24] Section 185 similarly mandates audio-video recording, preferably by mobile device, of every search and seizure, while section 193(2)(i) requires the investigating officer's report to record a documented chain of custody for electronic devices.[25] The BSA correspondingly modernises the evidentiary treatment of digital and electronic records, building upon and extending the certificate-based regime for computer output that Indian courts had developed

[19] "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Official Journal of the European Union, OJ L 1689 (2024)".

[20] "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Official Journal of the European Union, OJ L 1689 (2024), art 5(1)".

[21]" Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), arts. 5(2)– (7)".

[22] "European Commission, 'AI Act', *digital-strategy.ec.europa.eu* (2024), available at: https://digital-strategy.ec.europa.eu (last visited on July 3, 2026); *see also* Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), Annex III".

[23] "Bharatiya Nagarik Suraksha Sanhita, 2023; Bharatiya Sakshya Adhiniyam, 2023; PRS Legislative Research, 'The Bharatiya Nagarik Suraksha (Second) Sanhita, 2023', *Bill Track Summary* (2023), available at: https://prsindia.org (last visited on July 3, 2026)".

[24] "Obhan & Associates, 'Introduction of the Bharatiya Nagarik Suraksha Sanhita, 2023: A Step Towards Digital India' (2023), available at: https://www.obhanandassociates.com (last visited on July 3, 2026)".

[25] "Bharatiya Nagarik Suraksha Sanhita, 2023, ss. 185, 193(2)(i); *see also* 'Role of Forensic Evidence under BNSS 2023: Enhancing Credibility in Criminal Investigations', 8(3) *International Journal of Law Management and Humanities* 547 (2025)".

under the erstwhile section 65B of the Indian Evidence Act 1872.[26]

This legislative architecture is procedural rather than substantive: it mandates the recording and preservation of evidence gathered with technological assistance, but it does not yet regulate the use of predictive, biometric, or risk-assessment AI tools as such, nor does it establish an equivalent to the EU's prohibited-practices list or the US's Daubert-style reliability gate specifically calibrated to algorithmic evidence. India's Digital Personal Data Protection Act 2023 supplies a general data-fiduciary framework applicable to AI systems that process citizens' personal data within the justice system, but a dedicated AI-in-criminal-justice statute remains absent, a gap that NITI Aayog's National Strategy for Artificial Intelligence had flagged as a policy priority without translating it into binding law.[27] Given that the BNSS already normalises AI-adjacent tools such as automated evidence mapping and video graphed forensic collection, the absence of a parallel admissibility and bias-audit framework for algorithmic outputs is a conspicuous lacuna that Indian legislators would do well to close before, rather than after, contested convictions built on unregulated AI evidence reach appellate courts.

6. Due Process, Admissibility, and the Black-Box Problem

The recurring doctrinal difficulty across jurisdictions is that machine-learning systems, particularly deep neural networks, do not yield an auditable chain of reasoning in the way a forensic examiner's testimony can be cross-examined. This opacity threatens the reliability and auditability that due process requires of evidence tendered against an accused.[28] Four sequential evidentiary hurdles recur in the comparative literature: authentication of the system and the integrity of its input data; a reliability inquiry into the scientific validity of the underlying model; disclosure sufficient to permit meaningful cross-examination; and a weighing of probative value against the risk of undue prejudice arising from the jury or judge over-trusting a quantified output.[29] Comparative doctrinal analysis across the United States, the European Union, and the United Kingdom confirms that each jurisdiction is straining existing evidentiary doctrines built around human testimony and deterministic instruments to accommodate probabilistic, adaptive, and often proprietary AI outputs, without any of the three having yet settled a stable, generally applicable standard.[30]

The trade-secret defence that manufacturers routinely invoke as Northpointe did in Loomis compounds the problem: an accused person may be permitted to review the inputs supplied to a model, such as the answers on a questionnaire, without any corresponding ability to interrogate how those inputs were weighted to produce the score used against them.[31] Courts have thus far navigated this tension through procedural compromise mandatory limitation warnings, prohibitions on scores being the sole or determinative factor, and continuous re-validation requirements rather than through a substantive resolution of whether trade-secret confidentiality can properly be allowed to prevail over a defendant's right to test the evidence against them.[32] Reproducibility deficits compound the difficulty: unlike a fingerprint

[26] "'Digital Evidence in India: BSA 2023 & BNSS', *Legal Vidur* (2025), available at: https://legalvidur.com (last visited on July 3, 2026)".
[27] Srivastav, *supra* note 12, Table 1 (comparative legislative ecosystem, entry on India's National Strategy for Artificial Intelligence).
[28] *ibid*. Srivastav and others.
[29] "Admissibility of AI-Generated Evidence in U.S. Courts', *supra* note 22; *see also* Sahibpreet Singh and Lalita Devi, *supra* note 22".
[30] 'Research Paper on Admissibility of AI-Generated Evidence: A Comparative Study of U.S., EU, and U.K. Legal Frameworks', 12(2) *International Education and Research Journal* (2026); *State v. Loomis*, *supra* note 2; Israni, *supra* note 15".
[31] *State v. Loomis*, *supra* note 2; Israni, *supra* note 9.
[32] *State v. Loomis*, *supra* note 2.

comparison or a DNA match, which can in principle be independently re-run by a second examiner, many proprietary risk-assessment and predictive-policing tools cannot be replicated by defence experts, leaving courts to accept vendor-supplied validation studies largely at face value.[33]

7. Algorithmic Bias and the Perpetuation of Structural Inequality

Bias in forensic and predictive AI arises principally from three sources: skewed or unrepresentative training data, which encodes historical patterns of over-policing in particular communities; algorithmic decision-making bias introduced through feature selection or model architecture; and miscalibration between demographic groups in the resulting predictions.[34] The most empirically documented illustration outside the sentencing context is facial recognition. In the groundbreaking Gender Shades audit, Joy Buolamwini and Timnit Gebru assessed three commercial systems designed for gender classification and discovered that error rates were less than one percent for lighter-skinned men, while they reached as high as 34.7 percent for darker-skinned women. This disparity, which intersects along race and gender lines, is primarily attributed to the insufficient representation of darker-skinned women in the training datasets used to develop these systems.[35] The study's naming of specific vendors catalysed measurable industry change, with several companies subsequently narrowing the accuracy gap after the disparities were publicised, and it has since become a reference point for regulatory instruments including the EU AI Act's prohibition on biometric categorisation by sensitive attribute that treat unaudited biometric deployment as inherently high-risk.[36]

For criminal investigation specifically, the implication is that a facial-recognition match used to generate a suspect lead carries a materially higher risk of misidentifying a woman of colour than a light-skinned man, even before any human investigator introduces further confirmation bias by treating an algorithmic match as inherently more objective than an eyewitness identification.[37] The ProPublica findings on COMPAS demonstrate the same structural dynamic in the sentencing context: the tool did not need to ask about race to reproduce racially disparate outcomes, because proxy variables correlated with race family arrest history, neighbourhood, and employment status among them performed that function facially neutrally. The lesson that recurs across both domains is that algorithmic fairness cannot be secured through purely technical de-biasing; because competing fairness metrics are often mathematically incompatible with one another, the selection of which metric to prioritise in a given criminal-justice application is an irreducibly legal and political choice that legislatures, not vendors, must make explicit.

8. Explainable AI and the Case for Meaningful Human Oversight

Explainable AI (XAI) techniques including SHAP, LIME, and counterfactual explanation methods attempt to render otherwise opaque model outputs interpretable to legal and forensic audiences without requiring wholesale disclosure of proprietary source code.[38] Early assumptions that explainability could only be bought at the expense of predictive accuracy

[33] "Singh S and Devi L, 'Reliability and Admissibility of AI-Generated Forensic Evidence in Criminal Trials', *SSRN Working Paper* (2025), available at: https://www.ssrn.com (last visited on July 3, 2026)".

[34] Srivastav and others *supra* note 12.

[35] "Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', 81 *Proceedings of Machine Learning Research* 1 (2018), pp. 77–91".

[36] "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), art. 5(1)(g)".

[37] "Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', 81 *Proceedings of Machine Learning Research* 1 (2018); Digital Personal Data Protection Act, 2023".

[38] Srivastav and others *supra* note 12.

have been challenged by newer research, suggesting that carefully designed models can reach both strong accuracy and real interpretability in other words it kind of breaks the idea that transparency is technically unaffordable in high stakes legal deployment.[39] The doctrinal significance of XAI is that it converts a black-box probability into a reasoned, contestable proposition one that a defence expert can interrogate and a court can weigh rather than a number to be accepted or rejected wholesale.

While human oversight can't function properly without being part of a system, that does not mean that it is a self-sustaining system. Automation bias research has pointed to an effect called "out-of-the-loop," whereby the human operator tends to overly rely on the system's output and therefore lacks the level of engagement required to pick up mistakes in the output. There is also the danger of too much manual intervention resulting in a loss of the benefits brought by responsible automation.[40] According to the Loomis ruling, a written notice attached to the presentence report illustrates certain limits imposed on the disclosure-only measures. While the scores presented may be said to be inaccurate, they may nonetheless be given precedence in a particular case in the absence of any opposing statistics. Such a situation calls for a far more thorough oversight of the score in order to effectively protect the rights of the defendant.

9. Reimagining Criminal Justice: Toward a Rights-Based Governance Framework

A salient aspect that we implied from the materials analysed above is that there is currently no one regulatory regime that strikes that fine balance between innovation and rights protection. The American pathway comprises ex post evidentiary gatekeeping, where AI products are introduced into the market and subjected to investigation only when an offense is being disputed. Conversely, the European model prohibits certain types of usage of AI beforehand, but this exposes them to some possibilities of entrenching beneficial usages in an inflexible risk taxonomy. Other countries such as India are still doing a lot regarding the governance behind the use of predictive and biometric AI systems.

The explainability requirement shall depend on the significance of the matter involved. When it comes to the biometric identification process and sentencing risk scores, one would need to provide a thorough enough explanation to facilitate proper scrutiny of the decision. Whenever trade secret issues arise, these should not intervene with the duty to disclose criminal details.[41] Second, mandatory jurisdiction-specific validation: a tool trained and normed on one population should not be deployed against another without independent cross-validation, addressing precisely the gap the Wisconsin Supreme Court flagged but did not itself require as a precondition of use.[42] Third, a categorical prohibition, modelled on Article 5 of the EU AI Act, on purely profile-based criminal-risk prediction absent objective, verifiable facts directly tied to conduct, and on real-time public biometric surveillance absent prior judicial authorisation.[43]

Fourth, periodic, independently conducted bias audits, published in disaggregated form by demographic category, building on the audit methodology that proved effective in narrowing commercial facial-recognition disparities after the "Gender Shades findings were made public.[44] Fifth, and most fundamentally, a non-dirigible rule that no AI output

[39] ibid.
[40] ibid., discussing variable-autonomy design and the out-of-the-loop syndrome in human–AI oversight.
[41] *State v. Loomis*, *supra* note 2
[42] ibid.
[43] "Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), art. 5(1)(h)".
[44] "Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', 81 *Proceedings of Machine Learning Research* 1 (2018)".

may constitute the sole or determinative basis for an adverse liberty determination arrest, bail denial, conviction, or sentence enhancement without an independent, documented human assessment of the underlying facts, a principle already implicit in Selvi's insistence that scientifically derived material play only a corroborative role and in Loomis's insistence that a risk score cannot be determinative.[45] For India specifically, the BNSS's existing mandate for video graphed, chain-of-custody-documented forensic collection offers a ready procedural chassis onto which a parallel algorithmic-transparency and audit obligation could be grafted through subordinate legislation or judicially developed guidelines, without requiring wholesale statutory redrafting.

10. Conclusion

Artificial intelligence has already reshaped the practical texture of criminal investigation and is steadily encroaching upon the adjudicative core of criminal justice. Fingerprint and bloodstain analysis, DNA mixture resolution, facial and voice matching, crime-scene reconstruction, and the geospatial forecasting of offending have all been rendered faster, more consistent, and, on the available evidence, frequently more accurate than their purely manual predecessors. At the same time, actuarial risk assessment, predictive policing, and algorithmically assisted case management have begun to condition decisions bail, sentence length, parole eligibility that go to the core of an individual's liberty. The case for continued, carefully bounded use of these tools is genuine and should not be understated: faster forensic turnaround eases the backlog that itself produces injustice through delay; more consistent risk categorisation can, in principle, correct for the idiosyncrasy and fatigue that afflict individual human decision-makers; and expanded access to legal information and virtual assistance meaningfully benefits litigants who could not otherwise afford representation.

Yet the doctrinal and empirical record, surveyed in this article counsels against any kind of triumphalist reading of that promise. Loomis, the ProPublica investigation into COMPAS, the Gender Shades audit, and India's own constitutional jurisprudence in Selvi converge (in a kind of recurring, same note) on a single caution: statistical accuracy measured at the level of a population is not the same guarantee as individually tuned justice measured at the level of a defendant, and mathematical objectivity is not identical to legal fairness. A recidivism-risk score that looks good in the aggregate may still be systematically off about the particular person standing before the court, a facial-recognition system with a seemingly impressive average accuracy rate may still misidentify a specific woman of colour with troubling regularity, and a forensic technique that is scientifically sophisticated is not, for that reason alone, automatically something to which an accused person's participation may be compelled. Each of these points was hard won through litigation, investigative journalism, and academic audit not really volunteered by the industry that built the tools in the first place. That overall pattern is instructive too: the burden of exposing algorithmic shortcomings in criminal justice has, so far, fallen disproportionately on defendants, journalists, and independent researchers rather than on the regulatory state, and a mature governance framework can't keep leaning on this ad hoc split of labour.

Nor is the difficulty really stuck inside just one jurisdiction or even one legal tradition. In the United States, the evidentiary gatekeeping model, in the European Union, the risk tiered statutory scheme, and in India, the technologically ambitious but substantively incomplete criminal procedure reforms each snag a piece of the answer then leave another piece hanging. What ties them together, if you read it closely, is not a shared solution, no, rather a shared predicament: legal systems, built around testimony that can be cross examined, and tools whose operating principles are deterministic and disclosable are now being pushed to fit probabilistic, adaptive, and often

[45] *Selvi v. State of Karnataka*, *supra* note 6.

proprietary software. That software outputs arrive with the look of certainty, but not with the actual substance, like a well-made mask. So comparative learning across these three systems and later others too is not just academic curiosity. It is a practical necessity, because no jurisdiction, acting alone, is likely to reach any durable resolution of these questions using only its own doctrinal materials.

Reimagining criminal justice for the age of artificial intelligence does not, on this account, mean building faster or more powerful algorithms, nor does it mean rejecting them out of an excess of caution. It means building the explainability, validation, audit, and human-oversight architecture within which those algorithms may be trusted only so far as they have actually earned that trust and no further. It means treating the choice of fairness metric embedded in a risk-assessment tool as the legal and political decision that it is, rather than allowing it to be made silently by a vendor's engineering team. It means insisting that no person's liberty be curtailed on the strength of an output that cannot, in some meaningful form, be explained to the person it affects and tested by the person defending them. And for India in particular, it means recognising that the procedural sophistication already embedded in the Bharatiya Nagarik Suraksha Sanhita and the Bharatiya Sakshya Adhiniyam mandatory forensic visits, video graphed search and seizure, documented chains of custody is a foundation on which a substantive algorithmic-governance framework can and should now be built, rather than a substitute for one. The future of criminal justice will, in all likelihood, continue to be written partly in code. Whether that future is a just one will depend not on the sophistication of that code, but on the willingness of legislatures, courts, and the legal academy to keep human judgment accountable, reasoned, and reviewable at its centre.

*****

Bibliography

Statutes and Legislative Materials

- The Bharatiya Nagarik Suraksha Sanhita, 2023 (India).
- The Bharatiya Sakshya Adhiniyam, 2023 (India).
- The Digital Personal Data Protection Act, 2023 (India).
- The Federal Rules of Evidence, 2020 (U.S.).
- NITI Aayog, National Strategy for Artificial Intelligence (Government of India, 2020).
- PRS Legislative Research, "The Bharatiya Nagarik Suraksha (Second) Sanhita, 2023" (Bill Summary), available at https://prsindia.org/billtrack/the-bharatiya-nagarik-suraksha-second-sanhita-2023.
- Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 2024 O.J. (L 1689).

Cases & Articles, Reports and Online Sources

- Loomis v. Wisconsin, 137 S. Ct. 2290 (2017).
- Selvi v. State of Karnataka, (2010) 7 S.C.C. 263.
- State v. Loomis, 881 N.W. 2d 749 (Wis. 2016).
- "Admissibility of AI-Generated Evidence in U.S. Courts", AI Legal Authority, available at https://ailegalauthority.com/ai-evidence-admissibility.
- "Digital Evidence in India: BSA 2023 & BNSS", Legal Vidur (2025), available at https://www.legalvidur.com/blog/how-digital-evidence-is-changing-indian-courts.
- "Research Paper on Admissibility of AI-Generated Evidence: A Comparative Study of U.S., EU, and U.K. Legal Frameworks", Int'l Educ. & Res. J. (2026), available at https://ierj.in/journal/index.php/ierj/article/view/5097.
- "Role of Forensic Evidence under BNSS 2023", Int'l J.L. Mgmt. & Human. (2025), available at https://ijlmh.com/wp-content/uploads/Role-of-Forensic-Evidence-under-BNSS-2023.pdf.

- "State v. Loomis", Case Comment, 130 Harv. L. Rev. (2017), available at https://harvardlawreview.org/print/vol-130/state-v-loomis/.
- Angwin, Julia, Jeff Larson, Surya Mattu & Lauren Kirchner, "Machine Bias", ProPublica, May 23, 2016.
- Buolamwini, Joy & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", 81 Proc. Mach. Learning Res. 77 (2018).
- European Commission, "AI Act", Shaping Europe's Digital Future, available at https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.
- European Digital Rights (EDRi), "How to Fight Biometric Mass Surveillance after the AI Act: A Legal and Practical Guide" (2024), available at https://edri.org/our-work/how-to-fight-biometric-mass-surveillance-after-the-ai-act-a-legal-and-practical-guide/.
- European Parliament, "EU AI Act: First Regulation on Artificial Intelligence", available at https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
- Israni, Ellora, "Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing", UCLA Law Review (Sept. 20, 2019), available at https://www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/.
- Obhan & Associates, "Introduction of the Bharatiya Nagarik Suraksha Sanhita, 2023: A Step Towards Digital India" (2023), available at https://www.obhanandassociates.com/blog/introduction-of-the-bharatiya-nagarik-suraksha-sanhita-2023-a-step-towards-digital-india/.
- Singh, Sahibpreet & Lalita Devi, "Reliability and Admissibility of AI-Generated Forensic Evidence in Criminal Trials" (SSRN Working Paper, 2025), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5917402.

Srivastav, Satyam, K.M. Shruti, Dalima Agarwal, Priyanka Soni & Rajat Singh, "Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms", 84 Int'l J.L. Crime & Just. 100832 (2026).