
ALGORITHMIC JUSTICE OR ALGORITHMIC BIAS? EXAMINING THE CONSTITUTIONAL LIMITS OF ARTIFICIAL INTELLIGENCE IN INDIA'S CRIMINAL JUSTICE SYSTEM

Aryan Verma, LL.B. (Hons.) University of Allahabad

ABSTRACT

Facial recognition cameras trained on protest sites, a predictive-policing dashboard colouring Delhi's neighbourhoods by projected crime risk, and a High Court bail order that pauses, mid-reasoning, to consult a chatbot India's criminal justice system has begun to acquire artificial intelligence one procurement contract at a time, without a single statute drafted with algorithms in mind. This paper asks whether that acquisition can survive Articles 14 and 21 of the Constitution and the common-law principles of natural justice that give those Articles their operative content. It argues that the doctrinal materials for an answer already exist the anti-arbitrariness principle of *Royappa and Maneka Gandhi*, the proportionality standard of *Puttaswamy*, the reviewability principle of *Anuradha Bhasin*, and the anti-vagueness doctrine of *Shreya Singhal* but that no case has yet welded them into a standard capable of being applied to a machine-generated output. Reading these strands together with the EU AI Act's risk-based architecture and the cautionary lessons of *State v. Loomis* and *Carpenter v. United States*, the paper proposes a four-part constitutional floor transparency, reviewability, accountability and contestability, converging on an enforceable right to explanation that must be satisfied before any AI-generated output is permitted to influence a person's liberty. Its central and, it hopes, its principal claim to originality is this: India does not yet need to resolve whether AI should be used in criminal justice; it needs, first and as a matter of constitutional obligation rather than policy preference, a framework for Explainable AI.

Keywords: Artificial intelligence, Article 14, Article 21, natural justice, facial recognition, predictive policing, algorithmic due process, right to explanation, Explainable AI, EU AI Act, criminal justice, privacy.

I. Introduction

On 27 March 2023, a bench of the Punjab and Haryana High Court, hearing a routine bail application, did something no Indian court had done before: it asked a chatbot what it thought about bail in cases involving cruelty.¹ The reference to ChatGPT in *Jaswinder Singh v. State of Punjab* was explicitly non-precedential the judge took care to record that the exchange was offered only "to present a broader picture on bail jurisprudence" and decided nothing on its own.² But it was a signal. Three years on, the signal has become a pattern. Delhi Police's Crime Mapping, Analytics and Predictive System paints the capital in shades of red and green to tell patrol vans where crime is statistically likely to occur next.³ The National Crime Records Bureau has spent the better part of a decade trying to build a nationwide Automated Facial Recognition System, an ambition first announced by public tender in 2019.⁴ The Supreme Court's own research portal, SUPACE, now sifts case files for judges who once did that work themselves a tool its architects were careful to describe as an assistant rather than an oracle.⁵ None of this required a constitutional amendment, a new criminal code, or, in most instances, even a specific statute. It arrived the way most infrastructure arrives in a large administrative state: through procurement, pilot projects, and the quiet accretion of executive practice.

This is not a paper about whether that arrival is, on balance, good or bad. Both the panic and the enthusiasm that ordinarily attend discussions of "AI in policing" tend to obscure a more tractable legal question, and it is that question this paper takes as its subject: **can artificial intelligence be used in criminal investigations and judicial decision-making in India without violating Articles 14, 21 and the principles of natural justice?** The honest answer, defended over the course of this paper, is conditional rather than binary. Nothing in the constitutional text forbids the police from using software, any more than it forbids them from using fingerprint databases or ballistic comparison. What the Constitution forbids is the deprivation of liberty on the strength of a process that cannot be explained, reviewed, attributed

¹Jaswinder Singh @ Jassi v. State of Punjab, CRM-M-22496-2022 (P&H HC, Mar. 27, 2023) (Chitkara, J.).

²Id. at ¶¶ 9–10. The court expressly clarified that the ChatGPT reference was not "an expression of opinion on the merits of the case" and directed the trial court not to advert to it.

³Delhi Police, Live Crime Mapping, Analytics and Predictive System (CMAPS) (2019); see also Aditi Agrawal, *India's Tryst with Predictive Policing*, Vidhi Centre for Legal Policy (2023).

⁴National Crime Records Bureau, Request for Proposal for National Automated Facial Recognition System (2019, reissued 2020); see Smriti Parsheera, *Adventures in Policing Facial Recognition*, Observer Research Foundation (2022).

⁵Supreme Court of India, SUPACE Supreme Court Portal for Assistance in Courts Efficiency, launched Apr. 2021. Then-Chief Justice S.A. Bobde described the tool as confined to research assistance, with final adjudication remaining "the province of a human mind."

to an accountable actor, or contested by the person it is used against. Indian criminal justice already uses such processes today. The question this paper puts to those uses is not whether they are efficient they very often are but whether they are constitutional under a due-process standard indifferent to the technology through which state power is exercised.

The paper's organising claim, developed fully in Part VIII, is that India does not yet need to settle the prior, more dramatic question of whether AI *should* be used in criminal justice at all. It needs, first, a constitutional floor for *how* it may be used a framework for what computer scientists call Explainable AI (XAI), translated into the vocabulary Articles 14 and 21 already supply. Put as a single sentence: **India requires a constitutional framework for Explainable AI in criminal justice before any AI-generated output is permitted to influence a deprivation of liberty.** That framework does not yet exist, in legislation, in regulation, or in binding precedent. This paper tries to build it not from first principles, but from doctrine India already has.

II. Research Question, Methodology and Scope

The paper proceeds doctrinally and comparatively. Part III maps, as a factual predicate, where artificial intelligence currently sits within the Indian criminal justice pipeline policing, evidence, bail, and adjudication drawing on government tender documents, Right to Information disclosures, parliamentary committee reports and investigative and academic secondary sources, since no comprehensive official inventory exists. Part IV restates the constitutional architecture of Articles 14, 21 and natural justice as it stood before any of these tools appeared. Part V rereads three landmark judgments *Puttaswamy*, *Anuradha Bhasin* and *Shreya Singhal* for what they imply, rather than what they held, about automated state action. Part VI turns comparative, examining the European Union's AI Act and two lines of United States jurisprudence, not to import foreign law wholesale but to test how mature regulatory and judicial systems have handled problems India's law has not yet had to confront. Parts VII and VIII are the paper's constructive core: an account of why algorithmic opacity is a constitutional problem and not merely a technical inconvenience, and a proposed doctrinal and institutional framework the TRACE standard for resolving it. Part IX addresses the strongest objections to that framework, and Part X concludes.

Three limits on scope should be stated at the outset. First, the paper addresses *criminal* justice specifically; it does not directly examine AI in civil adjudication, tax assessment or welfare

administration, though the doctrinal framework developed here is not obviously confined to crime. Second, "artificial intelligence" is used throughout in a deliberately broad, functional sense any system whose output a human decision-maker treats as informative or determinative without being able to fully reconstruct, from first principles, how that output was produced. This includes both large language models of the kind used in *Jaswinder Singh* and older, narrower statistical classifiers of the kind used in facial matching or crime-hotspot prediction; the constitutional analysis in Parts VII and VIII applies to both. Third, the paper's comparative engagement with the European Union and the United States is illustrative rather than exhaustive, chosen because the two jurisdictions represent contrasting regulatory postures *ex ante* statutory risk-tiering in the European Union, *ex post* constitutional adjudication in the United States between which India's own path is still being decided.

III. Mapping the Machine: Artificial Intelligence Across India's Criminal Justice Pipeline

Before the constitutional analysis can proceed, it helps to be concrete about what is actually being deployed, by whom, and on what legal basis because the honest answer to the last question, in almost every instance below, is "none in particular."

A. Predictive Policing and Hotspot Analytics

Delhi Police announced its Crime Mapping, Analytics and Predictive System in 2015: a tool that ingests real-time data from the city's "100" helpline and FIR registrations to generate live spatial "hotspot" maps, on the theory that crime clusters predictably in place and time and that patrol resources should follow the cluster.⁶ Similar tools, in various stages of maturity, have since been piloted by police forces in Punjab, Uttar Pradesh, Rajasthan and Madhya Pradesh, generally for hotspot mapping and, increasingly, in conjunction with facial recognition.⁷ The predictable objection and it is not merely predictable but empirically documented is that a system trained on where police have historically made arrests will recommend policing exactly those places again, regardless of where offending actually occurs; a 2020 academic audit of CMAPS found that its inputs already encode the patrol patterns of the preceding years, so that the tool's "predictions" risk becoming a mechanised restatement of prior enforcement bias

⁶Delhi Police, From the Commissioner's Desk (Technical Report, 2015); Delhi Police, CMAPS (2019).

⁷Vidhi Centre for Legal Policy, *India's Tryst with Predictive Policing* (2023) (noting that Punjab, Uttar Pradesh, Rajasthan and Madhya Pradesh have each piloted predictive or FRT-linked policing tools, and that RTI exemptions for law-enforcement bodies have kept independent audit of CMAPS's accuracy largely impossible).

rather than an independent forecast of crime.⁸ The European Union's AI Act takes this risk seriously enough to prohibit outright any AI system used "to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person" a provision this paper returns to in Part VI, because it marks a considered legislative judgment that some predictive-policing architectures are not merely risky but categorically impermissible.⁹ India has made no comparable judgment, in either direction. What it has done, according to a Ministry of Home Affairs disclosure to a parliamentary standing committee as recently as March 2026, is begin building AI-based predictive-policing and "risk scoring of entities" directly into the next version of the Crime and Criminal Tracking Network System the backbone linking every police station in the country.¹⁰ The pipeline, in other words, is being built faster than the doctrine that would govern it.

B. Facial Recognition Technology and Biometric Surveillance

The National Crime Records Bureau's ambition for a National Automated Facial Recognition System dates to a 2019 public tender, withdrawn after a legal notice from the Internet Freedom Foundation and reissued in substantially similar form the following year; NCRB has pointed to a 2009 Cabinet Note on the Crime and Criminal Tracking Network System as its legal foundation an executive note, not an Act of Parliament, and one that predates the technology it is now stretched to authorise.¹¹ At the state level, Delhi Police's own facial recognition deployment illustrates the pattern well: procured pursuant to a 2018 Delhi High Court direction to help trace missing children, it was subsequently trained on crowds at protest sites, prompting the Internet Freedom Foundation to issue a formal notice questioning both its legal basis and

⁸Kristen Vaccaro et al., *Data in New Delhi's Predictive Policing System*, ACM Conf. on Fairness, Accountability & Transparency 317 (2020) (documenting that CMAPS ingests PCR call and FIR data that itself reflects prior policing patterns, so that neighbourhoods historically over-policed generate more predicted "hotspots," reinforcing the same pattern).

⁹Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence, art. 5(1)(d), 2024 O.J. (L 1689) [hereinafter EU AI Act] (prohibiting "the placing on the market, the putting into service for this specific purpose, or the use of AI systems for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics").

¹⁰Ministry of Home Affairs, evidence before the Parliamentary Standing Committee on Communications and Information Technology, *Impact of Emergence of Artificial Intelligence and Related Issues* (tabled in Lok Sabha, Mar. 30, 2026) (disclosing that NCRB's CCTNS 2.0 architecture envisages AI-based predictive policing and "risk scoring of entities").

¹¹National Crime Records Bureau, Request for Proposal for National Automated Facial Recognition System (2019); the tender was formally withdrawn after a legal notice from the Internet Freedom Foundation but reissued in substantially similar form in 2020. NCRB has relied on a 2009 Cabinet Note authorising the Crime and Criminal Tracking Network System as the legal basis for AFRS an executive instrument, not an enactment of Parliament.

its accuracy.¹² That question of accuracy is not academic. Delhi Police's own disclosures indicate that any match returned below an 80 per cent confidence threshold is treated internally as a "false positive" requiring independent corroboration which is to say that even a reported "match" carries what amounts to an acknowledged, non-trivial error margin, on a technology increasingly used to justify further investigative action against a named individual.¹³ The system's vendor, notably, markets a companion "predictive intelligence" product to the same police force, illustrating how quickly biometric identification and predictive risk-scoring become bundled in practice even where they remain doctrinally distinct.¹⁴ As with predictive policing, no Act of Parliament today authorises, conditions or limits the use of facial recognition by an Indian police force; the closest India has is the certificate regime for electronic evidence in the Bharatiya Sakshya Adhiniyam, which governs how the *output* of an FRT match may later be tendered in court, but says nothing about the conditions under which the match may be generated in the first place.¹⁵

C. AI-Generated and Deepfake Evidence

A different problem arises once synthetic media becomes cheap enough to manufacture and sophisticated enough to survive casual inspection. Two risks, symmetrical and equally corrosive, follow: fabricated audio or video can be introduced as if genuine, and genuine audio or video can be dismissed as fabricated. American scholarship has named the second risk the "liar's dividend" the mere existence of convincing deepfake technology lets a defendant plausibly, and sometimes successfully, contest evidence that is entirely authentic.¹⁶ India's evidentiary architecture is not well positioned for either risk. The Bharatiya Sakshya Adhiniyam's expert-opinion and certification provisions were drafted to authenticate *that* an

¹²Internet Freedom Foundation, Legal Notice to the Delhi Police (Dec. 28, 2019); Internet Freedom Foundation, Project Panoptic (facial recognition tracker, ongoing). The Delhi Police FRT system was procured to trace missing children pursuant to a 2018 direction of the Delhi High Court in *Sadhan Haldar v. NCT of Delhi*, W.P.(CrL) 1290/2015, but was subsequently used for crowd surveillance at protest sites.

¹³Delhi Police affidavit and RTI disclosures, discussed in Vrinda Bhandari & Renuka Sane, *The Regulation of Facial Recognition Technology in India*, Dakshindia (2021) (noting that any match below an 80% confidence threshold is treated as a "false positive" requiring corroboration meaning even a reported "positive" match carries a substantial margin of error).

¹⁴The vendor supplying the Delhi Police FRT system, Innefu Labs Pvt. Ltd., separately markets a predictive-intelligence product; see Internet Freedom Foundation, Legal Notice to the Delhi Police (Dec. 28, 2019).

¹⁵The Bharatiya Sakshya Adhiniyam, 2023, No. 47 of 2023, § 63(1) (deeming a "computer output" a document, admissible without production of the original, if the conditions of the section are satisfied).

¹⁶Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Calif. L. Rev. 1753, 1785–86 (2019) (coining the "liar's dividend" the risk that the mere existence of deepfake technology allows genuine audio-video evidence to be plausibly, and sometimes successfully, dismissed as fabricated).

electronic record was produced by a particular device through a particular process; they were not drafted to authenticate that the *content* of that record depicts events that actually occurred, which is precisely the question a deepfake puts in issue.¹⁷ Nothing in Indian law today requires provenance metadata, content-authentication watermarking, or an affirmative burden on the party tendering audio-video evidence to rule out synthetic generation before a court is asked to rely on it. The problem compounds where AI is used not merely to fabricate evidence but to *evaluate* it—for example, software that scores the likelihood a given statement is truthful, or that flags a given face as matching a suspect—because such tools import exactly the opacity problem this paper addresses in Parts VII and VIII into the heart of the trial process itself.¹⁸

D. Algorithmic Bail and Risk-Assessment Tools

India has not, as of this writing, deployed a proprietary bail- or sentencing-risk instrument comparable to the American COMPAS tool discussed in Part VI.¹⁹ That absence is worth pausing on rather than passing over, for two reasons. First, the pressure that produced such tools elsewhere—severe undertrial overcrowding, a perception that bail decisions vary with the identity of the judge rather than the facts of the case, and a policy appetite for "objective," auditable criteria—is, if anything, more acute in India than in Wisconsin. Second, the *Jaswinder Singh* episode shows that Indian judges are already, on their own initiative and without any governing protocol, reaching for general-purpose AI tools in exactly this context.²⁰ A jurisdiction that permits ad hoc judicial consultation of ChatGPT on bail jurisprudence, while having no framework at all for a purpose-built, India-specific, auditable risk-assessment tool, has arguably chosen the worse of the two available options: informality without accountability, rather than a designed system with accountability built in. This is not an argument for adopting a COMPAS-style tool. It is an argument that the absence of one is not, by itself, evidence of institutional caution—the caution may simply be misdirected.

¹⁷See Bharatiya Sakshya Adhiniyam, 2023, § 39 (expert opinion) read with § 63(4) (certificate). Neither provision was drafted with generative or synthetic media specifically in view, and neither imposes a provenance or watermarking requirement on the media itself.

¹⁸Compare *State v. Loomis*, 2016 WI 68, 881 N.W.2d 749 (Wis. 2016), discussed *infra* Part VI(B).

¹⁹No Indian court or police force has, as of this writing, formally deployed a proprietary bail- or sentence-risk-scoring instrument comparable to COMPAS; the absence is itself notable given the reform proposals discussed by NITI Aayog and the Vidhi Centre for Legal Policy, and given that undertrial overcrowding has repeatedly prompted calls for "objective" bail criteria.

²⁰*Jaswinder Singh @ Jassi v. State of Punjab*, CRM-M-22496-2022 (P&H HC, Mar. 27, 2023).

E. Judicial Deployment of AI: SUPACE, SUVAS and the ChatGPT Order

The judiciary's own encounters with AI have been more self-conscious than the executive's, though not always more coherent. *Jaswinder Singh* remains the paradigmatic instance: Justice Chitkara posed ChatGPT a direct question "What is the jurisprudence on bail when the assailants assaulted with cruelty?" received a generic, jurisdiction-agnostic answer, and then expressly ring-fenced that answer from the operative reasoning of the order. Commentary since has been sharply divided over whether that ring-fencing actually worked, or whether an unreasoned, unattributable AI-generated passage sitting inside an otherwise reasoned judicial order inevitably colours the order's legitimacy regardless of the disclaimer attached to it.²¹ The Delhi High Court took a more cautious line a few months later, declining, in a trademark dispute, to rely on ChatGPT-generated material to resolve any disputed question of fact.²² Caution of this kind is not excessive: a well-known American episode, in which counsel filed a brief containing entirely fictitious, AI-fabricated case citations, is now a byword for what happens when generative AI's fluent confidence is mistaken for legal reliability.²³ At least one other Indian High Court has since acknowledged consulting ChatGPT for supplementary research.²⁴ The Supreme Court's own institutional tools SUVAS, an AI translation system for regional-language judgments, and SUPACE, an AI-assisted research portal were both designed, deliberately, to stop short of anything resembling a recommendation: SUPACE surfaces material for a judge to read, and does not itself suggest an outcome.²⁵ That design choice is sound, and Part VIII argues it should be generalised into a binding principle rather than left to the good sense of individual institutions and individual judges.

IV. The Constitutional Architecture: Articles 14, 21 and Natural Justice

None of the tools surveyed in Part III were contemplated by the framers of the Constitution. But the constitutional standard to which they must answer was not built for them specifically,

²¹Anmol Jain, *ChatGPT-fication of the Judicial Process*, SCC OnLine Blog (Nov. 14, 2023).

²²*Christian Louboutin SAS v. M/s The Shutiq*, CS(COMM) 583/2023 (Del. HC, Aug. 2023) (Singh, J.) (declining to rely on ChatGPT-generated material for adjudicating disputed facts, while permitting counsel to place it before the court for illustrative purposes only).

²³*Mata v. Avianca, Inc.*, No. 22-cv-1461 (S.D.N.Y. June 22, 2023) (sanctioning counsel for filing a brief containing fictitious, AI-generated case citations).

²⁴Anmol Jain, *ChatGPT-fication of the Judicial Process*, SCC OnLine Blog (Nov. 14, 2023) (reporting a Manipur High Court bench's acknowledgment of having consulted ChatGPT 3.5 and Google for supplementary research).

²⁵Supreme Court of India, SUVAS Supreme Court Vidhik Anuvaad Software (launched 2019, an AI-assisted translation tool for rendering judgments into regional languages) and SUPACE (launched Apr. 2021, an AI-assisted research portal). Neither tool is authorised, nor has it been used, to generate a decision or a recommended outcome.

and does not need to be rebuilt from nothing; it needs, principally, to be read across.

A. Article 14 and the Anti-Arbitrariness Principle

Article 14 guarantees equality before the law and the equal protection of the laws.²⁶ Justice Bhagwati's opinion in *E.P. Royappa v. State of Tamil Nadu* gave the guarantee a second, more expansive life: equality, he wrote, is "antithetic to arbitrariness," so that state action which is arbitrary is, for that reason alone, action that offends Article 14, independent of any showing of unequal treatment between comparable persons.²⁷ *Maneka Gandhi v. Union of India* entrenched the point and connected it to Article 21: a "procedure established by law" depriving a person of liberty must itself be fair, just and reasonable, not fanciful, oppressive or arbitrary, and Articles 14, 19 and 21 are to be read not as isolated silos but as a mutually reinforcing whole.^{28,29} Applied to an algorithmic system, the anti-arbitrariness principle does real work. A decision to flag a face as a match, to mark a neighbourhood as a hotspot, to score a defendant as high-risk that cannot be reconstructed, even by the officials who rely on it, is arbitrary in the specific, technical sense *Royappa* gives that word: it is not the product of an identifiable, followable process of reasoning applied consistently to like cases, but an opaque output accepted on faith. Opacity of this kind is not a novel constitutional problem dressed in new technological clothes; it is arbitrariness in its oldest and most literal sense a decision whose "why" nobody can state.

B. Article 21 and the Due Process Turn

Article 21 protects life and personal liberty against deprivation "except according to procedure established by law."³⁰ Since *Maneka Gandhi*, that phrase has been read to import a substantive fairness requirement resembling American due process, even though the Constituent Assembly deliberately chose "procedure established by law" over "due process of law" to avoid exactly that result. *Justice K.S. Puttaswamy (Retd.) v. Union of India* extended the same logic to

²⁶India Const. art. 14 ("The State shall not deny to any person equality before the law or the equal protection of the laws within the territory of India.").

²⁷*E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3, ¶ 85 (Bhagwati, J.) ("Equality is a dynamic concept with many aspects and dimensions and it cannot be cribbed, cabined and confined... From a positivistic point of view, equality is antithetic to arbitrariness.").

²⁸*Maneka Gandhi v. Union of India*, (1978) 1 SCC 248, ¶ 7 (holding that "procedure established by law" under Article 21 must be fair, just and reasonable, and reading Articles 14, 19 and 21 as an interlinked, mutually reinforcing "golden triangle").

²⁹*Id.* ¶ 48.

³⁰India Const. art. 21 ("No person shall be deprived of his life or personal liberty except according to procedure established by law.").

privacy, holding it intrinsic to life and personal liberty under Article 21 and diffused through the freedoms of Article 19.³¹ The extension matters directly here, because most of the tools surveyed in Part III are, first and foremost, surveillance technologies: they work by collecting, matching or inferring information about identified or identifiable people. A facial recognition system that scans a protest crowd, or a predictive-policing tool that profiles a neighbourhood's residents by proxy, is engaged in exactly the kind of informational self-determination question *Puttaswamy* answers and its answer is that such action requires a valid law, a legitimate aim, and a proportionate relationship between the two, not administrative convenience alone.

C. Natural Justice: The Duty to Hear and the Duty to Explain

Natural justice supplies the two principles that do the most direct work in this paper. The first, *audi alteram partem*, requires that a person facing an adverse decision be heard before it is made and to be heard meaningfully, a person must first know, with some particularity, the case being made against them.³² An opaque algorithmic output frustrates this at the root: a defendant cannot meaningfully contest a facial-recognition match or a risk score whose method of arriving at that output is neither disclosed to them nor, in many cases, fully understood by the official relying on it. The second principle, the duty to give reasons, was placed on a firm constitutional footing in *S.N. Mukherjee v. Union of India*, which described reasons as the indispensable "link between the material on which certain conclusions are based and the actual conclusions" a link that allows both the affected person and a reviewing court to test whether a decision-maker's conclusion actually follows from its premises.³³ Natural justice, unlike a specific statute, applies of its own force to public power wherever a statute is silent and does not expressly exclude it³⁴, which is precisely the position of nearly every AI tool surveyed in Part III: each operates in a statutory vacuum, which means natural justice, not a bespoke AI statute, is currently the *only* binding legal constraint most of these tools face. Finally, the right to a fair trial an "triangulation" of interests owed to the accused, the victim and society³⁵

³¹Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1, ¶¶ 3, 325 (nine-judge bench) (holding privacy to be intrinsic to life and personal liberty under Article 21 and to permeate the freedoms in Article 19).

³²*Maneka Gandhi v. Union of India*, (1978) 1 SCC 248, ¶ 84 (Bhagwati, J.) (holding audi alteram partem to be implicit even where a statute is silent, unless expressly or by necessary implication excluded).

³³*S.N. Mukherjee v. Union of India*, (1990) 4 SCC 594, ¶ 34 (holding that, subject to limited exceptions, the recording of reasons is a requirement of natural justice, being "the link between the material on which certain conclusions are based and the actual conclusions").

³⁴*A. K. Kraipak v. Union of India*, (1969) 2 SCC 262, ¶ 20 (collapsing the older rigid distinction between "administrative" and "quasi-judicial" functions for the purpose of applying natural justice).

³⁵*Zahira Habibullah Sheikh v. State of Gujarat*, (2004) 4 SCC 158, ¶ 42 (describing a fair trial as one in which the accused is not prejudiced and the truth is capable of emerging, and holding this a "triangulation" of interests owed to the victim, the accused and society).

supplies the umbrella under which all of these strands sit once a matter reaches adjudication: a trial cannot be fair if a central piece of the evidence against the accused is, by design, un-cross-examinable.

It is useful, before proceeding further, to give this cluster of ideas a name, since American scholarship already has one this paper borrows and adapts: *digital due process* the proposition that the fairness guarantees the Constitution has always demanded of a human decision-maker do not evaporate merely because a machine has been interposed between the evidence and the decision, and may in fact need to be made *more* explicit, precisely because a machine cannot be cross-examined the way a witness can.³⁶ Neither the Bharatiya Nagarik Suraksha Sanhita, which replaced the Code of Criminal Procedure in 2024, nor any other post-2023 criminal statute, contains a single provision addressing automated or algorithmic decision-making.³⁷ The doctrinal resources of Parts IV(A) through IV(C) anti-arbitrariness, substantive fairness, the duty to hear, and the duty to explain are accordingly not merely relevant background to what follows; for now, they are the *entire* Indian law of algorithmic criminal justice, doing work no purpose-built statute yet does.³⁸³⁹⁴⁰

V. Reading the Constitutional Trilogy for the Algorithmic Age

None of the three judgments examined in this Part concerned artificial intelligence. That is exactly the point of reading them here: each supplies a doctrinal tool general enough to reach a problem its authors did not foresee, in the way constitutional text is meant to.

A. Puttaswamy: Proportionality as a Gate for Algorithmic Surveillance

A nine-judge bench decided *Justice K.S. Puttaswamy (Retd.) v. Union of India* unanimously in 2017, holding privacy to be a fundamental right woven through Articles 14, 19 and 21 rather

³⁶This paper borrows and adapts the phrase "digital due process" from the American scholarship on data-driven government decision-making; see Danielle Keats Citron, *Technological Due Process*, 85 Wash. U. L. Rev. 1249 (2008).

³⁷See Bharatiya Nagarik Suraksha Sanhita, 2023, No. 46 of 2023 (the successor to the Code of Criminal Procedure, 1973), which contains no provision addressing algorithmic or automated decision-making in arrest, investigation or bail.

³⁸*Maneka Gandhi v. Union of India*, (1978) 1 SCC 248, ¶ 7.

³⁹*S.N. Mukherjee v. Union of India*, (1990) 4 SCC 594, ¶¶ 24, 34.

⁴⁰*Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1, ¶ 310 (Chandrachud, J., writing for four judges) (formulating a proportionality standard requiring legality, a legitimate state aim, and a proportional relationship between means and ends); *id.* ¶ 638 (Kaul, J., concurring) (adding necessity the absence of a less restrictive alternative and procedural safeguards against abuse as independent limbs).

than a freestanding common-law interest that Parliament was free to recognise or withhold.⁴¹ In doing so, the Court overruled two earlier decisions that had denied or fragmented the right, and closed a doctrinal gap that had stood since the 1950s.⁴² The judgment's lasting operational contribution is its proportionality standard. Justice Chandrachud's opinion requires legality a law must actually exist a legitimate state aim, and a proportionate relationship between the measure and that aim.⁴³ Justice Kaul's concurrence goes further, adding necessity the absence of a less restrictive alternative capable of achieving the same end and procedural safeguards against the abuse of the power in question, as independent requirements rather than mere colour on the balancing exercise.⁴⁴ The Court applied its own test just over a year later to substantially narrow the private-sector use of Aadhaar authentication, confirming that the standard has real, not merely rhetorical, teeth.⁴⁵ Run the facial recognition and predictive-policing tools of Part III through this four-part test and the results are not encouraging for the State. Legality fails at the first hurdle: a Cabinet Note from 2009 is not "law" in the sense *Puttaswamy* requires, and no subsequent Act of Parliament has filled the gap. Necessity fails at the second: a predictive-policing tool trained on the same historically biased arrest data it is meant to correct is not merely imperfect but arguably *unfit for its own stated purpose*, which is a harder proposition to defend than a merely disproportionate but functioning tool would be. And procedural safeguards fail almost by definition, since as Parts VII and VIII develop at length there is currently no mechanism by which an individual profiled, flagged or matched by any of these systems can even learn that it happened, still less contest it.

B. Anuradha Bhasin: Reviewability of Technologically Mediated State Action

If *Puttaswamy* supplies the test, *Anuradha Bhasin v. Union of India* supplies the enforcement mechanism, and for that reason its transplantation to algorithmic criminal justice is, if anything, more direct. The case arose from the prolonged internet shutdown imposed in Jammu and Kashmir following the abrogation of Article 370, and the Court held communication over the

⁴¹Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.

⁴²Id. ¶¶ 3(B), 325 (unanimously overruling *M.P. Sharma v. Satish Chandra*, AIR 1954 SC 300, and *Kharak Singh v. State of Uttar Pradesh*, AIR 1963 SC 1295, to the extent they had denied or fragmented a constitutional right to privacy).

⁴³Id. ¶ 310 (Chandrachud, J.).

⁴⁴Id. ¶ 638 (Kaul, J.).

⁴⁵*K.S. Puttaswamy (Aadhaar-5J.) v. Union of India*, (2019) 1 SCC 1, ¶¶ 316–27 (applying the *Puttaswamy* proportionality standard to uphold the Aadhaar scheme subject to significant reading down, including striking provisions permitting private-party use of Aadhaar authentication).

internet to be protected under Articles 19(1)(a) and 19(1)(g).⁴⁶⁴⁷ Its more important holding, for present purposes, is procedural rather than substantive: an order restricting a fundamental right through a technological measure must satisfy the proportionality standard *Puttaswamy* had already established, and critically must be *published*, because an unpublished restriction forecloses the very judicial review Article 32 exists to make available.⁴⁸⁴⁹ The Court's reasoning on this point translates almost without modification: "if the impugned action or order is not put out in the public domain, it would be difficult for the affected persons to challenge the same before the appropriate forum."⁵⁰ Substitute "algorithmic flag" for "order" and the sentence describes precisely the position of a person identified by facial recognition, scored by a predictive tool, or denied bail after a judge's off-the-record consultation of a generative model: in each case, the person affected has no order to challenge, because nothing was published, logged, or even necessarily recorded in a form the affected person could later obtain. The suspension orders in *Anuradha Bhasin* were, whatever their other flaws, at least in principle discoverable documents. Most algorithmic criminal-justice outputs in India today are not even that.⁵¹

C. Shreya Singhal: Vagueness, Overbreadth and the Problem of the Black Box

The third case in the trilogy struck down Section 66A of the Information Technology Act, 2000, which had criminalised sending information that was "grossly offensive" or had a "menacing character," among other undefined categories.⁵² The Court's reasoning rested on two connected doctrines. Vagueness: terms like "annoying" and "inconvenient" supplied no ascertainable standard by which either a citizen could conform their conduct to the law or an enforcing authority could apply it consistently, leaving the line between the lawful and the unlawful to be drawn arbitrarily, case by case, by whichever official happened to be enforcing it.⁵³ Overbreadth: because the provision's language swept in a "very large amount of protected

⁴⁶*Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637.

⁴⁷*Id.* ¶¶ 26–27 (holding freedom of speech and expression, and the freedom to carry on trade or business, over the medium of the internet to be protected under Articles 19(1)(a) and 19(1)(g)).

⁴⁸*Id.* ¶ 53 (defining proportionality as requiring "the appropriate or least restrictive choice of measures" to achieve the object of the restriction).

⁴⁹*Id.* ¶¶ 26, 154–56 (directing that all orders imposing restrictions be published, since an unpublished order forecloses the very judicial review Article 32 exists to secure).

⁵⁰*Id.* ¶¶ 155–56 (the Court noted that "if the impugned action or order is not put out in the public domain, it would be difficult for the affected persons to challenge the same before the appropriate forum").

⁵¹*Id.* ¶¶ 140, 152 (holding an indefinite suspension of internet services impermissible, and requiring periodic review of suspension orders).

⁵²*Shreya Singhal v. Union of India*, (2015) 5 SCC 1.

⁵³*Id.* ¶¶ 82–83 (holding terms such as "grossly offensive," "annoying" and "inconvenient" in § 66A of the

and innocent speech" alongside anything that might legitimately have been restricted, it could not be saved by the possibility that it might, in a given case, be applied reasonably.⁵⁴⁵⁵ This paper suggests that an opaque machine-learning system exhibits something closely analogous, which might be called *algorithmic vagueness*: a model whose decision boundary cannot be stated in words not merely because no one has bothered to state it, but because the boundary genuinely exists only as a pattern of weighted parameters across a high-dimensional space is, from the standpoint of the person it is applied to, functionally indistinguishable from a law that uses undefined, subjective terms. Both leave the affected person unable to know, in advance or in retrospect, which side of the line they fell on, or why. *Shreya Singhal* held that a legislature cannot escape the vagueness doctrine merely by assuring citizens the vague power will be "administered in a reasonable manner"; the same logic should prevent an executive agency from escaping the same doctrine merely by assuring citizens that an opaque model will be used responsibly. A promise of good faith is not a substitute for an ascertainable standard in either context.

VI. Comparative Perspectives

India is a relative latecomer to regulating algorithmic criminal justice, which is at least one advantage of latecomer status: two very different regulatory experiments have already produced lessons India can examine before choosing its own path.

A. The European Union's Risk-Based Model

The European Union's Artificial Intelligence Act takes an ex ante, risk-tiered approach, and its treatment of law enforcement is the most stringent in the entire Regulation. Article 5 prohibits outright not merely regulates the use of "real-time" remote biometric identification in publicly accessible spaces for law-enforcement purposes, subject only to three narrow, exhaustively listed exceptions: locating specific victims of trafficking or abduction or missing persons; preventing a specific, substantial and imminent threat to life or a terrorist attack; and locating

Information Technology Act, 2000 to be undefined and inherently subjective, leaving both citizens and enforcing authorities without any ascertainable standard).

⁵⁴Id. ¶ 83 (holding that a provision failing to draw a "clearly drawn line" between the innocent and the culpable is void for vagueness, adopting the standard articulated in *Grayned v. City of Rockford*, 408 U.S. 104 (1972)).

⁵⁵Id. ¶¶ 94–95 (holding that § 66A was capable of catching "a very large amount of protected and innocent speech" alongside any speech that might legitimately be restricted, and was accordingly overbroad and unsaved by Article 19(2)).

a suspect of an offence carrying at least a four-year custodial sentence.⁵⁶ The same Article separately prohibits predictive-policing tools that assess an individual's risk of offending "based solely on... profiling," and prohibits the untargeted scraping of facial images to build recognition databases even where any exception applies, deployment is gated behind a mandatory fundamental-rights impact assessment and registration in a public EU database.⁵⁷ Where law-enforcement AI falls short of an outright prohibition, it is almost invariably classified "high-risk" under Annex III individualised risk assessment, evaluation of evidence reliability, and profiling in the course of investigation or prosecution are named categories which triggers mandatory risk-management systems, bias-audited data governance, technical documentation, human oversight and accuracy requirements before the system may be placed on the market at all.⁵⁸ Most significant for the argument developed in Part VIII, the Act creates a freestanding right to explanation: a person subject to a decision based on a high-risk system's output, where that decision produces legal or similarly significant effects, may demand from the deployer a clear and meaningful explanation of the system's role and of the decision's main elements.⁶⁰ Nothing resembling any part of this architecture exists in Indian law.

B. United States Jurisprudence: Loomis and Carpenter

The United States offers a contrasting model: no comprehensive federal AI statute, but two significant judicial encounters with algorithmic and digital-age state power, decided within two years of each other and pulling in opposite directions. *State v. Loomis* concerned COMPAS, a proprietary recidivism-risk tool developed by Northpointe Inc. and used in a Wisconsin

⁵⁶EU AI Act, supra note 9, art. 5(1)(h) (prohibiting "real-time" remote biometric identification in publicly accessible spaces for law-enforcement purposes, subject to three narrow, exhaustively listed exceptions: locating specific victims of trafficking or abduction, or missing persons; preventing a specific, substantial and imminent threat to life or a terrorist attack; and locating a suspect of a serious offence carrying a custodial sentence of at least four years).

⁵⁷Id. art. 5(1)(d) (prohibiting risk-profiling of individuals for predicted offending based solely on profiling or personality traits); art. 5(1)(e) (prohibiting untargeted scraping of facial images to build recognition databases); art. 5(2)–(7) (fundamental-rights impact assessment and EU database registration as preconditions for any exception-based use).

⁵⁸Id. annex III, pt. 6 (classifying as "high-risk" AI systems used by or on behalf of law-enforcement authorities for individualised risk assessment, polygraph-type tools, evaluation of evidence reliability, profiling in the course of detection, investigation or prosecution, and crime-analytics tools).

⁵⁹Id. arts. 9, 10, 13–15, 26 (risk-management systems, data-governance and bias-mitigation obligations, technical documentation, human oversight and accuracy/robustness requirements for high-risk systems).

⁶⁰Id. art. 86 (right of an affected person to obtain from the deployer of specified high-risk systems a clear and meaningful explanation of the role of the AI system in a decision and of the main elements of that decision, where the decision produces legal effects or similarly significantly affects the person).

presentence report.⁶¹⁶² The Wisconsin Supreme Court rejected Eric Loomis's due-process challenge, but only on the condition that future use of such tools be hedged with specific, mandatory cautions: that the tool's proprietary methodology prevents disclosure of how factors are weighed; that its norming sample may not reflect the local population; that independent studies have raised questions about disproportionate risk classification of minority defendants; and that the tool requires continual re-validation.⁶³⁶⁴ Critically, the Court held that a risk score may inform a sentence but must never be its determinative factor.⁶⁵ The independent study the Court felt obliged to acknowledge in its own opinion was ProPublica's influential 2016 investigation, which found that COMPAS misclassified Black defendants as high-risk at roughly twice the rate of white defendants who did not go on to reoffend the single most-cited piece of evidence in the global algorithmic-bias literature, produced not by a regulator but by investigative journalists auditing a system no regulator had thought to audit first.⁶⁶ *Loomis* is best read as a warning about what happens when courts allow opaque tools into the process without first demanding disclosure standards: a set of judicially improvised cautionary instructions, however well-intentioned, is a weaker safeguard than a rule fixed in advance of any individual case, because it depends entirely on each successive court remembering to impose it.

Carpenter v. United States, decided two years later, took the harder line the Fourth Amendment's text arguably always supported. The government had obtained months of historical cell-site location data without a warrant, relying on the "third-party doctrine," under which information voluntarily shared with a third party here, a phone company traditionally carries no reasonable expectation of privacy.⁶⁷ Chief Justice Roberts, writing for the majority, declined to extend that doctrine to cell-site records, reasoning that their "deeply revealing

⁶¹State v. Loomis, 2016 WI 68, 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S. Ct. 2290 (2017).

⁶²Id. ¶ 13 (the Correctional Offender Management Profiling for Alternative Sanctions, or COMPAS, tool, developed by Northpointe Inc.).

⁶³Id. ¶¶ 51–53 (rejecting Loomis's argument that reliance on a proprietary, undisclosed scoring methodology violated his right to be sentenced on accurate, individually verifiable information).

⁶⁴Id. ¶¶ 65–66 (requiring that any presentence report using COMPAS include specific written cautions: that the proprietary nature of the tool prevents disclosure of how factors are weighed; that the norming sample may not reflect the local population; that studies have raised questions about disproportionate classification of minority defendants as high-risk; and that the tool requires continual re-validation).

⁶⁵Id. ¶ 98 (holding that a COMPAS score may be considered as one factor among several but must not be determinative of the sentence, and particularly not of the decision to incarcerate or of its severity).

⁶⁶Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, ProPublica (May 23, 2016) (reporting that COMPAS misclassified Black defendants as high-risk at nearly twice the rate of white defendants who did not subsequently reoffend); the Wisconsin Supreme Court itself acknowledged this study in *Loomis*, 881 N.W.2d at 763–64.

⁶⁷*Carpenter v. United States*, 138 S. Ct. 2206 (2018).

nature," their "depth, breadth, and comprehensive reach," and the fact that they are compiled automatically and near-effortlessly, distinguish them from the narrow categories of business record the third-party doctrine was designed for.⁶⁸ The Court explicitly framed its holding as a response to "seismic shifts in digital technology," declining to let a doctrine built for a pre-digital world silently swallow the privacy protections a digital world requires.⁶⁹ The same instinct animates *Kyllo v. United States*, which had earlier held that a home's interior does not become less constitutionally protected merely because a thermal-imaging device, rather than a human being, does the looking.⁷⁰ Read together, *Carpenter* and *Kyllo* stand for a proposition Indian courts have not yet had occasion to state but should: that a technology's capacity to make a constitutional violation efficient, silent and comprehensive is a reason for *more* judicial scrutiny of its use, not less, and that the automation of a search or a seizure does not, by virtue of being automated, place it outside the constitutional categories developed for human searches and seizures.

C. Lessons for India

Neither model can be transplanted wholesale, and this paper does not propose that either should be. The European Union's risk-tiered statute presupposes an administrative and regulatory capacity a functioning market-surveillance authority, a notified-body conformity-assessment ecosystem, a data-protection regulator with real enforcement teeth that India's Data Protection Board, still in its institutional infancy, does not yet possess. The American model presupposes a steady drip of individual constitutional litigation, over years and across many trial courts, that is a slow and expensive way to build a doctrine when the harm in question is measured in years of undertrial detention. What both models share, however, and what India's law currently lacks entirely, is a settled answer to a single, prior question: *who is entitled to know how an algorithmic output was reached, and by what mechanism can that entitlement be enforced before, not merely after, the output has been used against them?* The European Union answers with an ex ante statutory right; the United States answers, imperfectly and only sometimes,

⁶⁸Id. at 2217 (Roberts, C.J.) (holding that the government's acquisition of historical cell-site location records constitutes a Fourth Amendment "search," given the "deeply revealing nature," "depth, breadth, and comprehensive reach" of the data, and its automatic, near-effortless compilation).

⁶⁹Id. at 2219–20 (declining to extend the third-party doctrine of *Smith v. Maryland*, 442 U.S. 735 (1979), to cell-site records, and describing the decision as addressing "a narrow" but foundational question raised by "seismic shifts in digital technology").

⁷⁰*Kyllo v. United States*, 533 U.S. 27, 34 (2001) (holding that use of a thermal-imaging device to detect heat emanating from a home constitutes a search, since technology that discloses details of a constitutionally protected space that would otherwise be unknowable without physical intrusion cannot be placed outside the Fourth Amendment merely because it is a machine, not a person, doing the looking).

with ex post constitutional litigation. Part VIII argues that India's own constitutional text Articles 14 and 21, read through the natural-justice principles restated in Part IV already supplies the resources for an answer of its own, and that the task remaining is to state that answer as a single, applicable standard rather than leaving it scattered, as it currently is, across doctrines developed for entirely different problems.

VII. The Black Box Problem and the Right to Explanation

A. Three Kinds of Opacity

Computer scientist Jenna Burrell's influential typology distinguishes three reasons a machine-learning system might resist explanation, and the distinction matters because each calls for a different legal response. Intentional opacity is deliberate secrecy, typically commercial a vendor withholding its methodology as a trade secret, as Northpointe did with COMPAS in *Loomis*. Illiterate opacity is the ordinary citizen's, and often the ordinary judge's, inability to read code or statistics, regardless of whether the code is published. Technical opacity is the hardest problem: in a sufficiently large neural network, even the system's own designers cannot fully trace why a particular input produced a particular output, because the "reasoning," if it can be called that, is distributed across millions of weighted parameters in a form no human-scale explanation can faithfully compress.⁷¹ Frank Pasquale's account of the "black box society" adds an institutional dimension: opacity is frequently not an unfortunate by-product of complexity but a *chosen* feature, useful to the institution deploying the system precisely because it forecloses the kind of scrutiny that transparency would invite.⁷² Danielle Citron and Frank Pasquale's earlier, and in some ways more directly applicable, work on automated scoring systems argues that any system materially shaping a person's life chances owes that person, at an irreducible minimum, notice that scoring occurred, a meaningful opportunity to contest the inputs used, and periodic external audit of the system's accuracy a minimum this paper adopts and builds on in Part VIII.⁷³

⁷¹Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 Big Data & Society 1, 1–2 (2016) (distinguishing intentional opacity (deliberate secrecy, e.g. trade secrets), illiterate opacity (the reader's inability to comprehend code or mathematics), and technical opacity (the sheer scale and non-linearity of the model, which resists human-scale explanation even for its own authors)).

⁷²Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* 3, 8 (Harv. Univ. Press 2015).

⁷³Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 Wash. L. Rev. 1, 4–5 (2014) (arguing that scoring systems that shape a person's life chances demand, at minimum, notice, an opportunity to contest inaccurate inputs, and periodic external audit).

B. The Indian Statutory Vacuum: DPDP Act, BNSS and BSA

India's three most relevant recent statutes each had an opportunity to address algorithmic opacity in criminal justice, and each declined it, whether by omission or by design. The Digital Personal Data Protection Act, 2023 exempts from most of its own obligations including, critically, Chapter III, which contains the data principal's access and correction rights any processing "in the interest of prevention, detection, investigation or prosecution of any offence," and separately allows the Central Government to exempt notified state instrumentalities from the Act altogether on security and public-order grounds.⁷⁴⁷⁵ Unlike its European counterpart, the DPDP Act contains no analogue to Article 22 of the General Data Protection Regulation no qualified right against being subject to a solely automated decision with legal or significant effect, and no right to "meaningful information about the logic involved" in such a decision.⁷⁶ The gap is not an oversight so much as an inherited policy choice: the Justice B.N. Srikrishna Committee, whose 2018 report underlies the DPDP Act's structure, explicitly recommended that obligations beyond fair processing and basic security safeguards be left out of the data protection statute for national-security and crime-prevention processing, to be addressed instead through a dedicated law.⁷⁷ Eight years later, that dedicated law has not been written. The Bharatiya Nagarik Suraksha Sanhita, which replaced the Code of Criminal Procedure in mid-2024, contains no provision on algorithmic or automated decision-making at any stage of investigation, arrest or bail. The Bharatiya Sakshya Adhiniyam's certificate regime for electronic evidence discussed further in Part VIII addresses how a computer-generated record may be authenticated once created, but nothing in it requires that the *logic* by which an AI system generated an investigatively significant output (a match, a flag, a score) be disclosed to the person it was used against. The result, across all three statutes, is a system that regulates the packaging of algorithmic output far more carefully than

⁷⁴The Digital Personal Data Protection Act, 2023, No. 22 of 2023, § 17(1)(c) (disapplying most of Chapters II and III, and § 16, where personal data is processed "in the interest of prevention, detection, investigation or prosecution of any offence").

⁷⁵Id. § 17(2)(a) (empowering the Central Government to exempt notified state instrumentalities from the entire Act on grounds including sovereignty, security of the State, public order and prevention of a cognizable offence). See generally Digital Personal Data Protection Rules, 2025 (notified Nov. 13, 2025).

⁷⁶Compare Regulation (EU) 2016/679 (General Data Protection Regulation), art. 22 (a qualified right not to be subject to a decision based solely on automated processing, including profiling, which produces legal or similarly significant effects) and art. 15(1)(h) (a right to "meaningful information about the logic involved" in such processing). The DPDP Act contains no analogue to either provision.

⁷⁷Justice B.N. Srikrishna Committee, A Free and Fair Digital Economy: Protecting Privacy, Empowering Indians (2018), at 87–90 (recommending that obligations beyond fair, reasonable processing and security safeguards not apply to processing for national security or crime prevention, to be addressed instead through a dedicated law a dedicated law that, as of this writing, does not exist).

it regulates the process that produced it.

C. Locating a Right to Explanation Within Article 21

The absence of a statutory right to explanation does not mean India lacks the constitutional resources to build one. This paper argues the resources already sit inside Article 21, in combination with the natural-justice reasoned-decision doctrine restated in Part IV. *S.N. Mukherjee* held reasons to be the necessary "link" between a decision-maker's material and its conclusion a link owed to litigants as a matter of basic fairness, not judicial courtesy.⁷⁸ If that link must exist when a human tribunal decides a licensing dispute, it is difficult to construct a principled reason why the same link should be dispensable when the "decision-maker" is a facial-recognition system whose output leads directly to an arrest, or a predictive-policing tool whose output leads to intensified surveillance of a named community. The interest protected the ability of an affected person and a reviewing court to test whether a conclusion actually follows from stated premises is identical in both cases; only the identity of the decision-maker has changed, and Article 21's guarantee has never been understood to turn on that identity. This is the doctrinal seed from which Part VIII grows a full standard: not a wholly new constitutional right invented for the occasion, but the reasoned-decision doctrine India already has, applied without exception to a category of decision-maker the doctrine's original authors did not anticipate but also never excluded.

VIII. Towards a Constitutional Framework for Explainable AI: The TRACE Standard

A. The Case for a Constitutional, Not Merely Statutory, Floor

It would be open to Parliament to legislate a right to algorithmic explanation tomorrow, and Part VIII(C) urges exactly that. But this paper's central claim is stronger, and more useful precisely because it does not depend on Parliament acting: the floor proposed here already exists in the Constitution, in the same sense that a right to privacy already existed in Article 21 before nine judges said so in 2017. *Puttaswamy* did not invent the right to privacy; it *synthesised* decades of scattered, sometimes contradictory dicta into a single, applicable doctrine, and in doing so vindicated Dr. Ambedkar's and the Constituent Assembly's choice of open-textured guarantees over a closed, static catalogue of rights.⁷⁹ This paper suggests the

⁷⁸*S.N. Mukherjee v. Union of India*, (1990) 4 SCC 594, ¶ 34.

⁷⁹See *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1, ¶¶ 3(B)–(F) (synthesising decades of

same synthesis is now due, and overdue, for algorithmic due process. The materials anti-arbitrariness, proportionality, reviewability, the anti-vagueness doctrine, the reasoned-decision requirement already sit inside Articles 14 and 21, established across a dozen unrelated judgments. What has never happened is a case that squarely presents a court with an opaque, liberty-affecting algorithmic decision and asks it to weld those materials together. When that case arrives a habeas petition contesting an arrest founded on an unexplained facial-recognition match would be an obvious vehicle this paper argues the Court should hold that Articles 14 and 21, read with the natural-justice principles that give them content, already impose an enforceable floor of algorithmic explainability, and that Parliament's failure to legislate one does not suspend a constitutional guarantee that does not depend on legislation to exist.

B. The TRACE Standard Defined

This paper proposes that the floor be stated, and tested, as a five-part standard TRACE because a decision that fails any one of its elements cannot, consistently with Articles 14 and 21, be permitted to influence a deprivation of liberty. Each element is drawn directly from doctrine already examined in this paper; none is imported wholesale from outside Indian constitutional law.

Transparency of design. The existence, general purpose, and general logic of any AI system used in investigation, evidence evaluation or bail must be a matter of public record before deployment not the source code itself, which may legitimately remain a trade secret, but the kind of "meta-transparency" that lets the public and defence counsel know a system exists, what it is meant to do, and on what category of data it was trained. A system procured and deployed the way Delhi Police's FRT system was without a published notification, without a stated legal basis beyond an executive note fails this element regardless of how accurate the system later proves to be, for the same reason a law failing *Puttaswamy's* legality requirement fails regardless of how wise its object.⁸⁰ (see also Part III(B).)

Reviewability by a human authority. No AI output may be self-executing. A human official, with genuine and not merely rubber-stamped discretion, must review the output before it produces any liberty-affecting consequence, and that official must be identifiable

scattered dicta on privacy from Kharak Singh's dissent onward into a single, doctrinally settled right).

and answerable for the decision to proceed. *Loomis*'s insistence that a risk score inform but never determine a sentence is this element's clearest judicial statement to date, and SUPACE's own design a research aid, not a recommender shows the principle is already, if unevenly, understood within the Indian judiciary itself.⁸¹

Accountability for error. A clearly identified natural or legal person the procuring department, the vendor, or both under a contractually and statutorily fixed allocation must be answerable, including in damages, for a system's errors, and an audit trail sufficient to identify what went wrong and why must be generated and retained for every liberty-affecting use. A system procured with no public tender terms, no vendor liability clause and no logging requirement, as several tools surveyed in Part III appear to have been, fails this element by design rather than by accident.

Contestability before a court or tribunal. The affected person must be able to challenge the specific output applied to them, with access to enough information not necessarily the full model, but enough to mount a genuine, not merely notional, contest before the liberty-affecting consequence becomes final. This element operationalises *audi alteram partem* for a non-human decision-maker: it asks not whether a hearing occurred, but whether the hearing was capable of actually testing the thing that produced the adverse outcome.

Explainability of the individual outcome. Finally, and building on all four preceding elements, the person affected is entitled to a reasoned, intelligible account in ordinary language, not source code of why the system produced this output for them specifically. This is *S.N. Mukherjee*'s reasoned-decision doctrine, applied without dilution to an automated decision-maker: the "link" between material and conclusion that the doctrine requires of a human tribunal is required in identical terms here, and a system architecturally incapable of supplying that link what Part VII called technical opacity is, for that reason, architecturally incapable of lawfully producing a liberty-affecting output in an Indian criminal proceeding.⁸²

A decision that satisfies all five elements is not thereby immune from further scrutiny it must still, independently, satisfy *Puttaswamy*'s proportionality test if it burdens privacy, and

⁸¹State v. Loomis, 2016 WI 68, ¶ 98, 881 N.W.2d 749 (Wis. 2016) (a risk score may inform, but must not determine, the ultimate decision).

⁸²S.N. Mukherjee v. Union of India, (1990) 4 SCC 594, ¶ 34.

Anuradha Bhasin's publication requirement if it rests on an executive order. TRACE is a floor, not a ceiling: a necessary condition for constitutional validity, not a sufficient one. But it is a floor Indian criminal justice does not currently clear for a single one of the tools surveyed in Part III, which is precisely the paper's point.

C. Institutional Architecture: Legislation, Evidence Law and an Algorithmic Oversight Board

A constitutional floor identified by scholarship is of limited use until a court applies it or Parliament codifies it, and this paper closes with concrete proposals for both paths.

First, Parliament should enact a dedicated Artificial Intelligence in Criminal Justice (Regulation) Act, built around a mandatory, pre-deployment Algorithmic Impact Assessment for any AI system used in investigation, evidence evaluation, bail or sentencing a structure with a ready domestic precedent in the environmental clearance process, under which certain categories of project cannot lawfully proceed without a prior, documented, publicly consultable assessment, and a ready comparative precedent in the European Union's fundamental-rights impact assessment and conformity-assessment requirements.⁸³⁸⁴ The Assessment should mandate disaggregated accuracy testing false-positive and false-negative rates reported separately by relevant demographic categories, precisely the kind of testing Delhi Police's own 80 per cent confidence threshold shows is technically feasible but is not currently made public and a sunset clause requiring re-certification on a fixed cycle rather than a one-time approval.

Second, the Bharatiya Sakshya Adhiniyam should be amended to add an Algorithmic Evidence Certificate, modelled directly on the existing Section 63(4) certificate for electronic records which already requires dual sign-off from a device custodian and an independent expert but adapted to require the certifying expert to attest to the system's tested accuracy rates, its training-data provenance, and its TRACE compliance, before any AI-generated or AI-evaluated evidence may be tendered.⁸⁵ This is a modest extension of a regime the Supreme

⁸³EU AI Act, *supra* note 9, art. 27 (fundamental rights impact assessment) and art. 43 (conformity assessment), read together as the closest comparative analogue to the impact-assessment model proposed here.

⁸⁴Cf. the environmental clearance architecture under the Environment (Protection) Act, 1986 and the Environmental Impact Assessment Notification, 2006, which conditions certain categories of project on a prior, documented, publicly consultable assessment a structural template this paper adapts for algorithmic deployment rather than land use.

⁸⁵Bharatiya Sakshya Adhiniyam, 2023, § 63(4) (certificate regime for electronic records), read with the Schedule (Parts A and B, requiring both the custodian of the device and a named expert to certify), on which this paper models the proposed "Algorithmic Evidence Certificate."

Court has already twice insisted must be strictly, not loosely, complied with: *Anvar P.V.* held the predecessor certificate to be a mandatory condition precedent to admissibility rather than a mere matter of weight, and *Arjun Panditrao Khotkar* reaffirmed that position while clarifying only the narrow circumstances in which the certificate may be dispensed with.⁸⁶ There is no principled reason evidentiary rigour should be higher for a hard drive than for the algorithm that decided which hard drive to seize in the first place.

Third, India should establish an independent Algorithmic Oversight Board, combining technical and judicial membership on the model already used for the Central Information Commission and, more recently, the Data Protection Board of India bodies that pair adjudicatory authority with domain expertise rather than leaving either to develop in isolation.⁸⁸ The Board's function would be to receive Algorithmic Impact Assessments, conduct or commission independent bias audits, certify TRACE compliance, and hear individual complaints from persons affected by a non-compliant system giving NITI Aayog's existing, currently non-binding Responsible AI principles the institutional teeth they have lacked since 2021.⁸⁹

Fourth, pending legislation, courts confronted with an appropriate case should recognise TRACE as the operative content of Articles 14 and 21 in this domain, exactly as this paper argues *Puttaswamy* recognised privacy as the operative content of the same Articles in 2017 not as an act of judicial invention, but as the overdue synthesis of doctrine the Court had already built for other purposes.

IX. Anticipated Objections

Three objections recur whenever a framework like TRACE is proposed, and each deserves a

⁸⁶*Anvar P.V. v. P.K. Basheer*, (2014) 10 SCC 473, ¶ 24 (holding the certificate under the predecessor provision, § 65B of the Indian Evidence Act, 1872, to be a mandatory condition precedent to admissibility, not a mere matter of weight).

⁸⁷*Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*, (2020) 7 SCC 1, ¶¶ 71–73 (affirming *Anvar P.V.*, while clarifying that the certificate may be dispensed with only where it is shown to be genuinely unobtainable despite diligent effort, and that courts may in such cases summon it under their own powers).

⁸⁸On the model of the Central Information Commission and State Information Commissions under the Right to Information Act, 2005, and of the Data Protection Board of India under the Digital Personal Data Protection Act, 2023, § 18 each an independent, statutorily created body combining technical and adjudicatory competence.

⁸⁹NITI Aayog, Approach Document for India Part 1 Principles for Responsible AI (Feb. 2021); NITI Aayog, Approach Document for India Part 2 Operationalizing Principles for Responsible AI (Aug. 2021); NITI Aayog, Responsible AI for All: Adopting the Framework A Use-Case Approach on Facial Recognition Technology (discussion paper, Nov. 2022). All three remain administrative guidance without the force of law.

direct answer rather than a dismissal.

The efficiency objection holds that explainability requirements will cripple an already under-resourced police force, trading marginal fairness gains for a real loss in investigative speed. The answer is that TRACE is not a uniform, maximal standard applied identically regardless of stakes. A tiered model, in which the rigour of the five elements scales with the severity of the liberty interest at stake, is both doctrinally sound *Puttaswamy*'s own proportionality test already varies with the gravity of the intrusion and administratively realistic. An AI tool used only to prioritise which of several already-filed FIRs a constable reviews first need not clear the same bar as a facial-recognition match used to justify an arrest; the European Union's own severity-linked gate for real-time biometric identification, keyed to offences carrying at least four years' imprisonment, is one workable model for where such a line might sit in an Indian statute.⁹⁰

The trade-secrecy objection holds that mandatory disclosure will deter private vendors from selling to Indian law enforcement at all, or will force disclosure of genuinely competitive intellectual property. TRACE does not require source-code disclosure to the public, or even to the defence in the ordinary case; "transparency of design" is meta-transparency, and "contestability" requires only enough information to mount a genuine challenge, not a complete audit. The European Union's model regulatory and independent-auditor access under confidentiality obligations, with only outcome-level disclosure made public shows the tension between commercial confidentiality and constitutional accountability is manageable rather than fundamental, and Part VIII(C)'s proposed Algorithmic Oversight Board is designed specifically to hold that line.

The institutional-capacity objection holds that Indian courts, already burdened with enormous pendency, lack the technical capacity to evaluate algorithmic evidence even if a right to explanation is recognised. This is the most serious of the three objections, and this paper does not minimise it but Indian procedural law already has a tool built for exactly this problem. Order XXVI of the Code of Civil Procedure allows a court to appoint a commissioner to investigate and report on matters beyond judicial expertise; nothing prevents an analogous, statutorily empowered technical assessor from being appointed in criminal proceedings turning

⁹⁰Cf. EU AI Act, *supra* note 9, art. 5(1)(h)(iii) (permitting real-time biometric identification for offences carrying at least four years' custodial sentence) a severity-linked gate this paper adapts as one axis of the proposed tiered model.

on algorithmic evidence, drawing on the same empanelled experts the proposed Algorithmic Oversight Board would in any event need to maintain.⁹¹ Capacity is a real constraint on *how quickly* TRACE can be operationalised. It is not a reason to deny that the underlying constitutional entitlement exists.

X. Conclusion

Return to the question this paper opened with: can artificial intelligence be used in Indian criminal investigation and adjudication without violating Articles 14, 21 and natural justice? The answer defended here is neither the reflexive "no" of technological anxiety nor the reflexive "yes" of administrative convenience, but a conditional "not yet, not like this." Every tool surveyed in Part III—predictive policing, facial recognition, AI-evaluated evidence, and ad hoc judicial consultation of generative models—currently operates in a constitutional gap: authorised by nothing more than an executive note or an unstated assumption of implied power, reviewable by no one in particular, and explicable to the person it is used against not at all. That gap is not a reason to prohibit these tools outright, and this paper has not argued for prohibition. It is a reason to insist, as a matter of existing constitutional obligation rather than aspirational policy, that the gap be closed before, not after, the next arrest an opaque algorithm helps to justify.

The paper's claim to originality is narrow and, it hopes, useful precisely because it is narrow: not a new theory of privacy, and not a wholesale import of European or American law, but the proposition that Articles 14 and 21, read together with the natural-justice doctrines India has applied to human decision-makers for seventy years, already supply a five-part floor—transparency, reviewability, accountability, contestability and explainability—that no algorithmic criminal-justice tool in India today fully clears. *Puttaswamy* took forty years, and nine judges, to say explicitly what the Constitution had arguably always implied about privacy. Algorithmic due process should not have to wait as long, if only because the harm in the interim—an unexplained facial-recognition match, an unreviewed risk score, an unpublished flag—accrues one liberty at a time, in cases that rarely reach a court capable of saying so.⁹²

⁹¹Order XXVI, Code of Civil Procedure, 1908 (commissions to make local investigations or to examine accounts, appointing a qualified outsider to inform the court on matters beyond judicial expertise) offers a ready domestic template for court-appointed technical assessors in cases turning on algorithmic evidence.

⁹²Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1, ¶ 297 (Chandrachud, J.) ("Formalised constitutional morality... constitutes the soul of the Constitution.").

BIBLIOGRAPHY

Books

Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harv. Univ. Press 2015).

Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016).

Articles

Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, ProPublica (May 23, 2016).

Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 *Big Data & Society* 1 (2016).

Danielle Keats Citron, *Technological Due Process*, 85 *Wash. U. L. Rev.* 1249 (2008).

Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 *Wash. L. Rev.* 1 (2014).

Danielle Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 *Calif. L. Rev.* 1753 (2019).

Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"*, 38 *AI Magazine* 50 (2017).

Anmol Jain, *ChatGPT-fication of the Judicial Process*, SCC OnLine Blog (Nov. 14, 2023).

Kristen Vaccaro et al., *Data in New Delhi's Predictive Policing System*, *Proc. 2020 ACM Conf. on Fairness, Accountability & Transparency* 317 (2020).

Sandra Wachter, Brent Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 *Harv. J.L. & Tech.* 841 (2018).

Reports and Working Papers

Justice B.N. Srikrishna Committee, A Free and Fair Digital Economy: Protecting Privacy, Empowering Indians (2018).

NITI Aayog, Approach Document for India Part 1 Principles for Responsible AI (Feb. 2021).

NITI Aayog, Approach Document for India Part 2 Operationalizing Principles for Responsible AI (Aug. 2021).

NITI Aayog, Responsible AI for All: Adopting the Framework A Use-Case Approach on Facial Recognition Technology (discussion paper, Nov. 2022).

Aditi Agrawal, India's Tryst with Predictive Policing, Vidhi Centre for Legal Policy (2023).

Smriti Parsheera, Adventures in Policing Facial Recognition, Observer Research Foundation (2022).

Vrinda Bhandari & Renuka Sane, The Regulation of Facial Recognition Technology in India, Dakshindia (2021).

Internet Freedom Foundation, Project Panoptic (facial recognition tracker, ongoing).

Parliamentary Standing Committee on Communications and Information Technology, Impact of Emergence of Artificial Intelligence and Related Issues (tabled in Lok Sabha, Mar. 30, 2026).