

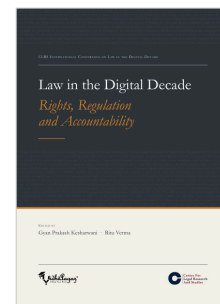
CHAPTER 12

Artificial Intelligence in Judicial and Administrative Decision-Making: Reassessing Rights, Regulation, and Justice

Saudamini Gupta and Priyanshu

in *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma (VidhiAagaz 2026), pp. 137–149

DOI <https://doi.org/10.63108/VAB.LDD.1.12>



CITATIONS

BLUEBOOK (21ST ED.)

Saudamini Gupta & Priyanshu, *Artificial Intelligence in Judicial and Administrative Decision-Making: Reassessing Rights, Regulation, and Justice*, in *LAW IN THE DIGITAL DECADE: RIGHTS, REGULATION AND ACCOUNTABILITY* 137 (Gyan Prakash Kesharwani & Ritu Verma eds., VidhiAagaz 2026).

OSCOLA (4TH ED.)

Saudamini Gupta and Priyanshu, ‘Artificial Intelligence in Judicial and Administrative Decision-Making: Reassessing Rights, Regulation, and Justice’ in Gyan Prakash Kesharwani and Ritu Verma (eds), *Law in the Digital Decade: Rights, Regulation and Accountability* (VidhiAagaz 2026) 137.

CHICAGO (17TH ED.)

Gupta, Saudamini, and Priyanshu. “Artificial Intelligence in Judicial and Administrative Decision-Making: Reassessing Rights, Regulation, and Justice.” In *Law in the Digital Decade: Rights, Regulation and Accountability*, edited by Gyan Prakash Kesharwani and Ritu Verma, 137–149. Maihar: VidhiAagaz, 2026. <https://doi.org/10.63108/VAB.LDD.1.12>.

APA (7TH ED.)

Gupta, S., & Priyanshu. (2026). Artificial intelligence in judicial and administrative decision-making: Reassessing rights, regulation, and justice. In G. P. Kesharwani & R. Verma (Eds.), *Law in the digital decade: Rights, regulation and accountability* (pp. 137–149). VidhiAagaz. <https://doi.org/10.63108/VAB.LDD.1.12>

SOURCE

<https://doi.org/10.63108/VAB.LDD.1.12>

LICENCE

Open access under the Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0): it may be copied, shared and adapted for non-commercial purposes, with credit to the authors and the source.

Citations are given as a guide; check them against the style manual you follow. This page was added by VidhiAagaz to the downloaded file. The chapter follows, with the page numbers of the printed book.

CHAPTER 12

Artificial Intelligence in Judicial and Administrative Decision-Making: Reassessing Rights, Regulation, and Justice

SAUDAMINI GUPTA* AND PRIYANSHU**

*Artificial intelligence has entered courts and administrative agencies more quietly, and a good deal more widely, than the public argument about robot judges would suggest. It is at work in bail and sentencing, in tax assessment and immigration screening, and in the daily administration of welfare. This article asks what becomes of due process once statistical inference begins to supplement, and then to shape, official discretion. There are a few recurring problems. Model reasoning is opaque, which sits badly with the principle of reasoned decisions. Training data carries the imprint of past discrimination, which then reappears wearing the borrowed authority of mathematics. And automation bias hollows out individualized judgment even where nobody intends it to. Drawing on *State v. Loomis*, the Dutch SyRI litigation and the childcare benefits scandal that followed it, Australia's Robodebt scheme, and the framework now emerging in India such as the Draft Regulations for Use of Artificial Intelligence in Courts, 2026, I argue for interpretability calibrated to what is at stake, independent audit, contestability designed in rather than bolted on, and human oversight built to survive a heavy docket.*

Keywords: algorithmic decision-making; due process; natural justice; administrative law; explainability; automation bias; risk assessment; AI regulation

1 INTRODUCTION

The debate on the use of artificial intelligence in the justice system is often found in the future tense. People went on envisioning that one day robot judges would start pronouncing judgments. To delve into this, we must first look at the factors that bring forth such ideas. In contemporary times, the citizens have a case against the slow response of the judicial system. Artificial intelligence is seen as an unbiased and effective alternative to bring in reforms to the issue of judicial delays. The use of artificial intelligence as per popular imagination has a shock value attached to it. The underlying reality, however, is less dramatic and considerably harder to regulate. The basis of all machine learning is the raw data that it builds its model upon.

Use of artificial intelligence isn't detached from human intervention. When artificial intelligence is applied to create filtering mechanisms, it would have practical implications. It would mean that the magistrate would read a risk score before ruling on bail. A revenue officer might examine a file because an artificial intelligence classifier flagged it. A social welfare employee may suspend a benefit because the fraud-detection model classified a household beyond the threshold. Each of these situations would have a machine learning model flagging a case and in each such flagging, the final decision would be attributed

* Junior Standing Counsel at Income Tax Department.

** Junior Research Fellow at National Law University Odisha, Cuttack, Odisha, India.

to a human. Every resultant human action has several steps wherein artificial intelligence intervened and changed the course of decision making. The interventions impacted the attending, framing and default position of the authority. When the end result attains finality, neither the certifying human authority nor the affected person can inspect the parameters of decision.

In the traditional sense, administrative and judicial laws are grounded in the principles of natural justice. This creates a duty upon the decision maker to apply her own mind to the case brought before her. She is required to make a reasoned decision and spell out the basis of her decision. A reasoned decision allows the affected person to know and contest those reasons. These reasons can be tested against the record by a reviewing court. This basic rule is disrupted by the introduction of a predictive system. Under the new arrangement, the order may still state the reasons but there may be no visible connection between the order and the outcome.

It is essential to have an effective legal system. In India, unfortunately, we have a large backlog of cases especially at the lower courts. There are times where a petitioner may die before their claim is decided. In such a scenario, automation systems can serve as an assistive mechanism. They can be helpful to translate judgments, transcribe proceedings, summarise records and triage filings. However, there is a difference in using assistive technology in court administration and using predictive technology for adjudication. It becomes pertinent to distinguish between a system that expands a decision-maker's capacity to reason about a case and a system that applies the general population level inferences to an individual case. The difference lies in the function performed. Any deployment of artificial intelligence must begin with the demarcation where its impact as a productivity tool ends and the role of adjudication begins. In each such case, the principles of natural justice that are applicable to adjudicatory acts must be applied to automated systems.

2 DEPLOYMENT AND RISK ASSESSMENT

As highlighted earlier, the use of artificial intelligence has begun in some areas of the justice system across jurisdictions. In this light, it becomes necessary to understand the level of risks across deployment.

Case Management and Drafting Support – One of the most extensive uses of automated systems is seen for functions such as to index filings, propose cause lists, transcribe hearings and translate judgments. India's SUVAS and SUPACE¹ tools utilise such deployments to further procedural convenience rather than impacting substantive outcomes. However, there lies a risk of mistranslation. For example, a paragraph may have been relied upon by the defense and can be dropped by the summariser. Such a problem may be classified as a quality-control problem.

Triage and Resource Allocation – One of the deployments of automated systems is for the purpose of tax assessment. In this area, while automated systems may not decide the case, they can still decide the people whose files would be examined, returns be audited and claims be investigated. This exercise may impact millions of people while being subjected to negligible checks and balances. When a particular trade or region is flagged disproportionately it marks the onset of an enforcement pattern that would not be revealed in any individual assessment order.

¹ Supreme Court Vidhik Anuvaad Software (SUVAS) and Supreme Court Portal for Assistance in Court's Efficiency (SUPACE); see Supreme Court of India, *Annual Report* (2021).

Risk and Needs Assessment – Automated systems are also used to predict recidivism or absconding possibilities. Such predictive inputs end up impacting bail, sentencing, parole and prison classification. At times due to bias in training data, the system attaches a probabilistic statement about a reference class. When a member of such a reference class seeks a remedy, a general probabilistic statement would be applied ignoring the individual differences of the particular case. This has a counterproductive result and ends up increasing the volume of litigation.

Eligibility Determination – There is a growing trend of using automated systems in welfare, subsidy, pension and tax administration for matching, scoring and calculation. Similarly, automated systems are used in visa triage and immigration risk tiering. The key concern lies in the profiles of subjects to such scrutiny. The beneficiaries of these programs are often those who have access to limited resources. In the event of any adverse decision, such people do not have the means to go for an appeal.

The underlying issue is that the harm caused cannot be mitigated by introducing sophisticated systems. For instance, the Australian government came up with the Robodebt scheme for automated debt assessment and recovery. The scheme did not use any machine learning. Yet, it resulted in one of the largest administrative disasters in Australian history. The problem lies with the automation of judgment on a large scale and the reversal of onus of proof. In the real world, the entire process is riddled with a huge power asymmetry. It is viable to see sentencing regulations being litigated because defendants have counsel and appeal rights. On the other hand, welfare based classifications are less likely to be litigated.

3 STRUCTURAL PROBLEMS

3.1 Dichotomy of Natural Justice and Black Box

The idea of natural justice requires the presence of reasoned decisions. This requirement serves a larger purpose by reducing the presence of bias and prejudices and inducing accountability. When decision makers are required to give reasons behind the outcomes, they are subjected to scrutiny and any irrelevant considerations are exposed. They let the affected party know the case against her. The requirement of reasoned orders makes supervisory review possible. The reasons stated in an order are assessed by the higher courts at the stage of appeal. Indian Courts have held that recording reasons is a facet of natural justice in quasi-judicial and administrative decision-making. An order stands and falls on the reasons stated in it.²

The introduction of predictive models disrupts the mechanism of checks and balances. It happens in two ways that run parallel. The initial issue is that of technical opacity. The mechanism of deep-network/gradient-boosted ensembles is based on weights. This violates the reasoned decision requirement of administrative law. This works on the idea of post hoc explanation which follows the idea that if the variables are changed, the output would change too. At the outset, the explanation resembles reason but in reality, it is an approximation of one model generated by the other. Rudin argues that such approximations can be unfaithfully precise where the stakes are highest. It is better to use an inherently interpretable model to reach consequential decisions rather than explaining an opaque one after the fact.³

² *Kranti Associates Pvt. Ltd. v. Masood Ahmed Khan*, (2010) 9 S.C.C. 496 (India); *Mobinder Singh Gill v. Chief Election Comm'r*, (1978) 1 S.C.C. 405 (India).

³ Cynthia Rudin, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 Nature Mach. Intelligence 206 (2019).

The other issue is that of legal opacity. The first step in deploying an automated model involves a contract between the vendor and the justice department. The vendors want to preserve their trade secrets and resist disclosing training data, feature sets and validation results. As seen in *Loomis*' case, the defendant could not examine how COMPAS produced his score because the methodology was proprietary.⁴ In the *Houston Federation of Teachers*⁵ case, a federal court recognised the impact of secrecy on procedural due process. The dismissed teachers had alleged that procedural due process was not followed because they were dismissed on the basis of proprietary value-added scores and they could not verify or contest their score. The federal court upheld the claim of violation of procedural due process. This reasoning deserves wider adoption since a person would be unable to check the validity of a score even if the most mathematically accurate model is deployed.

The duty does *not* require a complete causal account of the decision-maker's mental processes since human reasons are themselves post-hoc rationalisations to an unknown degree.⁶ It requires grounds specific enough to be tested for legal permissibility and rational connection to the outcome. For an algorithmic system this means that the input variables, the logic of their combination, the reference population, and the validation record are disclosable.

3.2 Training Data Bias and Cyclic Discrimination

A supervised machine model needs training data to get the output. With time the model ends up reproducing the patterns gathered through training data. When the input data itself suffers from bias and prejudices, it creates a discriminatory training database. The model reproduces this cycle of discrimination and confers it the authority of statistics. This is usually an indirect mechanism. One may try to resolve it through formally excluding protected characteristics. Such exclusions, however, serve little benefit because characteristics may be constructed from information such as postcode, occupation and prior interaction with police. The basis of the recidivism model becomes rearrest rather than reoffending. In an unevenly distributed policing system, the attention cannot be uniform. The model learns the distribution of police attention as much as the distribution of criminality.

This problem was practically experienced in the COMPAS controversy. There was a ProPublica analysis on trends of risk classification conducted in 2016. It looked at the cases where defendants did not end up becoming repeat offenders and highlighted that in this category, the black defendants were more likely to be classified as high risk. On the other hand, while examining the cases where defendants did end up becoming repeat offenders, it found that white defendants were more likely to be classified as low risk.⁷ The vendor defended the model by stating that it was calibrated. This meant that irrespective of which race a person belonged to a given score carried the same empirical reoffending rate. This was used to establish the relevant fairness criterion. The subsequent literature shows that both were right about their own metrics. However, the metrics are irreconcilable. Where base rates differ, calibration within groups and equality of false positive and false negative rates cannot both be satisfied except in degenerate

⁴ *State v. Loomis*, 881 N.W.2d 749, 761 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017).

⁵ *Hous. Fed'n of Teachers, Local 2415 v. Hous. Indep. Sch. Dist.*, 251 F. Supp. 3d 1168 (S.D. Tex. 2017).

⁶ Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1085 (2018); Sandra Wachter, Brent Mittelstadt & Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the GDPR*, 7 Int'l Data Privacy L. 76 (2017).

⁷ Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, ProPublica (May 23, 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

cases.⁸ The selection of fairness criterion is a normative act. It ends up determining how the burden of predictive error is distributed between social groups. When a vendor's engineering meeting decides this normative question, it ends up circumventing the question of constitutional equality.

This issue can be looked at from the lens of Indian equality doctrine. Article 14 of the Indian Constitution prohibits arbitrariness. In case a classification is to be made, it is permissible only after establishing an intelligible differentia. Such an intelligible differentia must have a rational nexus to the object sought to be achieved. An automated system that is built upon discriminatory data cannot become an intelligible differentia. When cases of violation of equality doctrine are brought in, in absence of disaggregated performance data, a petitioner cannot establish the discriminatory nature of the output made by an opaque process.

An extension of this argument that is often overlooked is the nuanced nature of adjudication. When judges are adjudicating cases, they do not design outcomes based solely on precedents. They perform a normative task that is constantly evolving. They overrule precedents, read down colonial era statutes and reconcile the conception of rights to the changing social realities. An automated model that uses historical data to perform a predictive task is structurally incapable of doing this. It is optimised to reproduce the past distribution of outcomes. Its margin of errors lies precisely in cases where the law ought to correct itself. Allowing automated systems to perform adjudicatory functions does not simply encode the prejudices of the past. It also removes the mechanism through which the justice system can evolve.

3.3 Automation Bias and the Erosion of Individualized Discretion

The subtlest risk in automation lies in the focus on human oversight in frameworks across jurisdictions such as the European Union's, Council of Europe and India's draft judicial regulations. This may end up defeating the safeguards currently being drafted. The empirical literature on human oversight offers very little comfort.

Research on different arenas of automation bias be it cockpit studies, clinical decision support or criminal justice risk management has led to some common observations. Operators tend to accept incorrect recommendations which they would have rejected had they not involved automated systems. They also fail to detect the issues that were not flagged by the system.⁹ Both these issues become manifold under time pressure and heavy caseloads. Supervisors are poor at identifying where a score is wrong.¹⁰ This results in a wrong score distorting judgments while the supervisor remains unaware. The oversight requirements are often satisfied on paper without offering any functioning protection in practice. The system is legitimised as being human control and the automation is not put for scrutiny.¹¹

The use of automation also creates an accountability asymmetry independent of psychology. If a decision maker departs from a recommendation and is wrong, he is exposed. On the other hand, a decision maker who follows it and is wrong can shift the onus on the system. When an institution audits only

⁸ Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, arXiv:1609.05807 [cs, stat] (2016), <https://arxiv.org/abs/1609.05807>; Alexandra Chouldechova, *Fair Prediction with Disparate Impact*, 5 Big Data 153 (2017).

⁹ Linda J. Skitka, Kathleen L. Mosier & Mark Burdick, *Does Automation Bias Decision-Making?*, 51 Int'l J. Hum.-Computer Stud. 991 (1999); see also Danielle Keats Citron, *Technological Due Process*, 85 Wash. U. L. Rev. 1249 (2008).

¹⁰ Ben Green & Yiling Chen, *Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments*, in Proceedings of the Conference on Fairness, Accountability, and Transparency 90 (2019).

¹¹ Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 Computer L. & Sec. Rev. 105681 (2022).

deviances and not concurrences, it elevates an advisory tool to the position of a de facto rule without any formal expression.

This problem is addressed in administrative law though it is rarely applied to software. The rule against acting under dictation holds that an authority vested with discretion must exercise it itself and may not surrender it to another's direction.¹² The principle that "he who decides must hear" condemns dividing the adjudicative function between the entity that receives the evidence and the entity that reaches the conclusion.¹³ An authority that in substance ratifies a vendor's output performs both the functions. The accountability cannot be diluted merely because the dictating authority is a software. On the contrary, the need for accountability is aggravated because a superior officer can at least be asked to state reasons.

4 COMPARATIVE LENS

4.1 United States: The Loomis Case and the Limits of Cautionary Instructions

Eric Loomis was sentenced in Wisconsin in a case of drive-by shooting. The sentencing court had referred to a COMPAS risk assessment in the presentence report. On appeal to the Wisconsin Supreme Court, he argued that due process was violated since he was unable to examine the proprietary methodology. He alleged that the instrument had used gender as an input and produced group based predictions¹⁴.

The court held that a COMPAS score could be considered but could not be used to determine whether an offender should be incarcerated or fix the severity of a sentence. It introduced procedural safeguards wherein any presentence report containing a COMPAS assessment should be accompanied by written cautions such as that the methodology is proprietary and undisclosed, that the instrument was normed on national rather than Wisconsin populations, that studies had questioned whether it disadvantages minority offenders, and that it was designed for correctional rather than sentencing decisions.

The judgment serves as an unintentional demonstration of why disclaimers do not work. The court acknowledged that the tool may be racially skewed and that it was normed on a different population. The sentencing court was still asked to consider it, while forbidden from relying on it decisively. The problem is that there is no cognitive procedure by which a judge can consider the unverified data without being influenced by it. The margin of error cannot be reviewed and practically a defendant can almost never establish that the score was biased.

The Wisconsin Court's reasoning is complicated. Dressel and Farid¹⁵ found COMPAS's predictive accuracy comparable to that of untrained laypeople given a handful of variables and suggested that the gains purchased at such constitutional cost may be modest. Further, the Houston teachers case shows that where a court squarely confronts trade secrecy against the right to contest, the analysis is not difficult.

4.2 The Netherlands: SyRI and the Aftermath

In The Netherlands, the SyRI (*Systeem Risico Indicatie*) was deployed. The data linked to SyRI spanned across employment, tax, housing, and benefits registries. The objective was to generate risk assessment reports and identify individuals who were likely to have committed fraud. It was deployed selectively, in

¹² *Comm'r of Police, Bombay v. Gordhandas Bhanji*, A.I.R. 1952 S.C. 16 (India); *Purtabpore Co. Ltd. v. Cane Comm'r of Bihar*, (1969) 1 S.C.C. 308 (India).

¹³ *Gullapalli Nageswara Rao v. Andhra Pradesh State Rd. Transp. Corp.*, A.I.R. 1959 S.C. 308 (India).

¹⁴ *Loomis*, 881 N.W.2d at 753–72.

¹⁵ Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 Sci. Advances eao5580 (2018).

practice concentrating in low-income neighborhoods. The model or the indicators were not published and the individuals were not informed that they had been assessed unless a report was generated.

In February 2020 the District Court of The Hague held that the enabling legislation violated Article 8 of the European Convention on Human Rights.¹⁶ The court accepted that combating fraud was a legitimate claim but held that the legislation failed the fair balance test. The absence of transparency about the model and indicators made it impossible to verify how conclusions were reached or to defend against them, and the system carried an unaddressed risk of discriminatory effect and stigmatization. The UN Special Rapporteur described the case as an early skirmish over the digital welfare state.

The SyRI judgment tests the legislation on the grounds of proportionality under an existing rights framework. The same structure is available under Article 21 in India after *Puttaswamy*.¹⁷ It treats verifiability as a component of legality. In other words, a system that cannot be checked cannot be justified. It also recognises that the harm falls on everyone processed, irrespective of whether they are found liable or not.

Within a year of the SyRI judgment, the Dutch childcare benefits case came up. The Dutch tax authorities had used risk classification to identify possible fraud in childcare allowance claims, treating indicators including dual nationality as markers of elevated risk. As a result, tens of thousands of families were wrongly accused and ordered to repay large sums. The data protection authority found the processing unlawful and discriminatory, and the government resigned in January 2021.¹⁸ The impact of the scandal was huge because of its scale, the presumption of guilt and the impossibility of contesting a classification whose basis was undisclosed. The Netherlands had a strong administrative judiciary, an active regulator, and a fresh precedent. Yet, striking down such a monitoring system could not prevent the harm.

4.3 Australia: Robodebt and the Reversal of Onus

The Australian government operated an automated compliance programme called Robodebt between 2015 and 2019. The program compared annual taxation income data against fortnightly welfare declarations. In cases where the annual figure was averaged across fortnights, irregular earnings, such as casual, seasonal, and gig work, generated apparent discrepancies and hence apparent debts. This led to an automatic issue of notices and the burden of disproving the debt fell on recipients. In many cases there were no corresponding payslips.

The Federal Court held that a debt calculated by income averaging was not properly made. The issue ended with a class action settled for approximately AUD 1.8 billion in debts being written off and compensation paid. A Royal Commission reported in 2023 that the scheme had been unlawful from the outset.¹⁹

In this case rather than a machine learning model, there lay instead the substitution of a statistical assumption for individual assessment, a reversal of the burden of proof, and the removal of the officers previously required to investigate a discrepancy before asserting a debt.

¹⁶ *NJCM v. The Netherlands* (SyRI), Rb. Den Haag, Feb. 5, 2020, ECLI:NL:RBDHA:2020:1878 (Neth.).

¹⁷ *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 S.C.C. 1 (India).

¹⁸ Autoriteit Persoonsgegevens, *Decision on the Processing of Nationality by the Tax Administration* (2021); Parliamentary Interrogation Committee, *Ongekend Onrecht* (Neth. 2020).

¹⁹ *Amato v. Commonwealth*, No. VID611/2019 (Fed. Ct. Austl. Nov. 27, 2019) (consent judgment); *Prygodicz v. Commonwealth* (No. 2) [2021] FCA 634 (Austl.); Royal Commission into the Robodebt Scheme, *Report* (Austl. July 2023).

4.4 India: Rapid Adoption, Emerging Governance

India's position is distinctive and, in a sense, more hopeful. Much of the framework is still being written and its jurisprudence on privacy, arbitrariness, and procedural fairness is unusually developed. The manner of adoption within the judiciary has so far been assistive rather than adjudicative. The adoption has been confined, as of now, to translation of judgments into regional languages, case-file analysis, digitization of the district judiciary under e-Courts Phase III, and live transcription piloted at the Supreme Court.

The governance response has been swift. The Ministry of Electronics and Information Technology released the India AI Governance Guidelines in November 2025. The guideline adopted a principles-based, light-touch model anchored in existing law rather than a standalone statute.²⁰ In June 2026 the Supreme Court's AI Committee published Draft Regulations for Use of Artificial Intelligence in Courts, 2026.²¹ The draft creates a classification and states that no AI system may perform the function of adjudication or sentencing. It mandates that deployed systems would only be used in assistive capacity. It requires advocates and litigants to disclose AI use in preparing pleadings, submissions, or evidence. This is in response to the emerging international problem of fabricated citation. It proposes to permanently establish an Apex Body at the Supreme Court. Such a body would set standards, approve tools, coordinate with High Court committees, handle grievances, and publish annual reports. Its scope extends beyond the constitutional courts to tribunals and statutory bodies exercising adjudicatory powers, where high-volume deployment is most likely.

While this serves as a stronger starting position in comparison to most other jurisdictions it still has gaps. The assistive versus adjudicative line is drawn at the wrong place. A system that produces a recommended outcome, a risk tier, or a ranked disposition is assistive in form and determinative in effect. The preferred demarcation should lie where it separates systems that expand the decision-maker's engagement with the individual record from those offering a conclusion derived from a reference class. The latter ought to be prohibited rather than focusing only on formal delegation. This would serve the purposes better.

The judicial framework does not reach the administrative state, where the volume of automated determination is far greater. Revenue, subsidy, and identity-linked benefit systems affect a greater number of people than court-annexed tools. Further, their outputs are typically challengeable only through writ jurisdiction most affected persons cannot access. The Digital Personal Data Protection Act, 2023 contains no equivalent to Article 22 of the General Data Protection Regulation. Thus, there exists no right against decisions based solely on automated processing. In other words, no right to human intervention and no right to contest.²² With broad powers to exempt State instrumentalities, this leaves the individual materially worse protected than in the European Union. The single most consequential reform available is to fill that gap either by rule, amendment, or judicial development under Articles 14 and 21.

On the practical side, procurement is unregulated as a rights question. The crucial decisions such as whether the model is interpretable, whether the vendor must disclose validation data, whether disaggregated performance is published, who owns the audit trail are made when a contract is signed. The terms fixed in a contract begin impacting rights long before a petitioner enters the courtroom. The certification

²⁰ Ministry of Electronics & Information Technology, Gov't of India, *India AI Governance Guidelines* (Nov. 2025).

²¹ Supreme Court of India, Artificial Intelligence Committee, *Draft Regulations for Use of Artificial Intelligence (AI) in Courts, 2026* (June 3, 2026).

²² Digital Personal Data Protection Act, No. 22 of 2023 (India); cf. Regulation 2016/679, art. 22, 2016 O.J. (L 119) 1 (EU).

pathway the draft regulations envisage is the right vehicle for imposing those terms, provided the criteria are published and certification decisions are reviewable.

4.5 The European Union and the Council of Europe

The European Union's AI Act classifies as high-risk both AI intended to assist a judicial authority in researching and interpreting facts and law and in applying the law, and systems used by public authorities to assess eligibility for essential public benefits or to detect fraud in that context.²³ The classification triggers obligations on risk management, data governance, documentation, logging, accuracy, and human oversight. As such public-body deployers must conduct a fundamental rights impact assessment. Post adjudication the affected persons must have a right to an explanation of the system's role in decisions producing legal effects.

The implementation record is itself cautionary. Obligations for standalone high-risk systems were to apply from 2 August 2026. Following the Commission's Digital Omnibus proposal of November 2025, an amending Regulation entered into force on 27 July 2026. The amended regulations deferred the Annex III high-risk obligations to 2 December 2027 and those for artificial intelligence embedded in regulated products to 2 August 2028, with exemptions for systems already on the market.²⁴ It was justified on the ground that harmonised standards, conformity assessment infrastructure, and national competent authorities were not ready. The binding force of an artificial intelligence statute depends on institutional capacity that takes years to build. The compliance deadlines are political variables rather than fixed points.

The Council of Europe's Framework Convention takes a different route. It requires parties to ensure that activities across the AI lifecycle are consistent with human rights obligations and to provide accessible remedies and procedural safeguards.²⁵ It can be easily transplanted across jurisdictions. Its rights-first framing serves as the more useful model for a jurisdiction that is unlikely to enact a comprehensive AI statute soon.

The Court of Justice held in *SCHUFA* that automated generation of a probability value can itself constitute an automated individual decision where a third party draws strongly on that value in deciding the individual's rights. This removes the formal distinction between recommendation and decision on the same lines as Part 3 suggests is necessary.²⁶ In *Bridges* the Court of Appeal held facial recognition deployment unlawful partly because the police force had not taken reasonable steps to satisfy itself that the software did not have a racial or gender bias. It treated the failure to investigate discriminatory performance as itself a breach of the public sector equality duty.²⁷ Such a decision solves the difficulty of proving algorithmic discrimination by shifting the onus from the claimant and transforming it into an obligation on the public body.

²³ Regulation 2024/1689, arts. 6, 14, 27, 86 & annex III, 2024 O.J. (L 1689) 1 (EU).

²⁴ Regulation 2026/1744 (Digital Omnibus on AI), 2026 O.J. (L 1744) (EU) (in force July 27, 2026).

²⁵ Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, C.E.T.S. No. 225 (opened for signature Sept. 5, 2024); European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment* (Dec. 2018).

²⁶ Case C-634/21, *SCHUFA Holding AG*, ECLI:EU:C:2023:957 (Dec. 7, 2023).

²⁷ *R (Bridges) v. Chief Constable of South Wales Police* [2020] EWCA Civ 1058 (Eng.).

5 LOCATING THE HARMS IN EXISTING DOCTRINE

A recurring assumption that algorithmic decision-making requires wholly new law is often false. Instead, it requires the application of existing doctrine to a new object supplemented with evidentiary rules that make that application possible.

Reasons. An order reciting conclusions while the operative ground was an undisclosed score is, on the *Mohinder Singh Gill* principle, an order supported by reasons other than those stated, and should be challengeable on that basis alone.

Hearing. The right to be heard is meaningless if the case to be met is undisclosed. Where an adverse inference is drawn from a model output, *audi alteram partem* should require that the person be told a system was used, what it produced, on what inputs, and how it was validated.²⁸

Non-delegation. Where a statute confers discretion on a named authority, that authority must exercise it. The systematic adoption of model outputs, particularly where deviation carries institutional cost, is discretion exercised under dictation.

Arbitrariness and equality. A learned proxy producing systematically worse outcomes for a protected group satisfies neither limb of the Article 14 test. A state that has not tested for such an effect has not discharged its obligation to act non-arbitrarily.

Proportionality and privacy. After *Puttaswamy*, state action interfering with informational privacy must satisfy legality, legitimate aim, necessity, and proportionality, with procedural safeguards. Data linkage and profiling across administrative databases such as the SyRI pattern falls squarely within this test.

A petitioner cannot plead disparate impact without disaggregated error rates, nor irrelevant consideration without the feature list, nor dictation without deviation statistics. A disclosure obligation is triggered by its reliance, so that documentation, feature sets, validation results, and disaggregated performance data are producible in proceedings, with trade secrecy a ground for protective orders rather than withholding; mandatory retention of decision logs recording the model output, the human decision, and any deviation; and an adverse inference where a body cannot produce this material.

6 A WORKABLE REGULATORY RESPONSE

Increasing transparency through interpretability. Where a system informs a decision affecting liberty, family integrity, immigration status, or subsistence, the model should be inherently interpretable. In practice a scoring rule, decision list, or sparse regression whose operation can be stated in a paragraph and reproduced by hand. Post-hoc explanation should not satisfy the requirement. For triage, post-hoc explanation plus published aggregate performance may suffice whereas for translation or retrieval, verification against source is the appropriate control. The accuracy cost of interpretable models in these domains is, on the evidence, small while the cost of opacity is a right rendered untestable.

Having an independent, published audit. There must be an audit conducted by an entity independent of both the vendor and deploying body. Such an audit should be repeated after any material retraining. Findings should be admissible, and the absence of audit should count against the public body.

Application of human oversight. Given the automation bias evidence, a requirement that a human “review” the output is close to worthless. Effective oversight requires presenting the case record before the

²⁸ *State of Orissa v. Dr. (Miss) Binapani Dei*, A.I.R. 1967 S.C. 1269 (India).

model output, so the human forms an independent view first. There must be recorded reasons for concurrence as well as deviation which would remove the asymmetric cost of disagreement. The concurrence rates should be audited since a rate approaching unity indicates rubber-stamping rather than agreement. The workload standards should make genuine review possible. There should be blind testing of reviewers against known-erroneous outputs. These belong in the certification criteria for any approved tool.

Inclusion of opportunity to contest. The affected person should be notified that an automated system was used, told the output and principal inputs, given a route to correct input data, given a right to fresh determination by a human who has not seen the output, and given reasons addressing her individual circumstances. That last point matters because an explanation stating that people with similar characteristics reoffend at a certain rate is not a reason for deciding individual cases.

Prohibition in certain areas. Some deployments should be off the table irrespective of accuracy. These may include (i) fully automated determination of guilt, sentence, or deprivation of liberty; (ii) automated suspension of subsistence benefits without prior human determination and an opportunity to be heard; and (iii) predictive systems resting on characteristics the constitution treats as suspect, or on proxies not separable from them. The Indian draft regulations' prohibition on AI adjudication should be extended in substance to administrative determinations affecting subsistence.

Procurement as the primary instrument. The state's leverage is greatest at the point of purchase. Standard contractual terms should require interpretability where the tier demands it. It should include disclosure of training data to auditors and to courts under protective order, state ownership of logs, disaggregated performance reporting as a condition of continued payment, and express waiver of trade secrecy objections in proceedings challenging a decision informed by the system.

Increasing institutional capacity and allowing legal aid. None of this functions without institutions able to exercise it. The Apex Body contemplated by India's draft judicial regulations is a promising model provided it is resourced with technical staff rather than constituted solely of judges and officials. It must publish its certification criteria and decisions that are subject to review. The European experience of deferring obligations for want of standards and competent authorities is a warning to build capacity before the deadline. Equally, those most exposed to automated decision-making are least able to litigate. Without funded representation, relaxed standing for public-interest challenges, and a route to aggregate remedies, the rights above will be exercised principally by those who have access to expert counsel. The SyRI litigation happened in the Netherlands because civil society brought it, and the Dutch benefits scandal continued because its victims could not.

7 OBJECTIONS

Uniformity reduces bias. The strongest objection is that discretion is already inconsistent and prejudiced, and that a model is at least uniform and testable. This objection is persuasive but uniformity remains a virtue only where the underlying rule is just. A litigant facing a biased judge may draw a different judge on appeal, while one facing a biased model faces the same model everywhere. Human error is also diverse and partially self-cancelling, whereas model error is correlated across every decision the model touches. A model error would end up converting individual mistakes into systemic ones. The unaudited proprietary model should be replaced with an audited interpretable model and structural guidance.

Scope of interpretation diminishes accuracy. This *claim* may be true in some cases but not always in the tabular, low-dimensional settings typical of justice and welfare administration, where simple models frequently match complex ones. Where a real trade-off exists, it is between marginal predictive gain and a person's ability to contest a decision about her own liberty.

The issue of backlog would be unchanged by regulation. The backlog argument is serious. However, it argues for aggressive investment in exactly the assistive tools this paper does not propose to restrict. What is restricted is the substitution of statistical inference for individual determination, which does not address backlog at all. It rather processes cases faster by not deciding them.

Disclosure would lead to circumvention. Publishing exact feature weights may allow evasion where a system detects deliberate wrongdoing. That justifies tiered disclosure that mandates full access for auditors and courts under protective order, aggregate publication otherwise. This does not lead to secrecy from the affected individual, who learns nothing exploitable from being told why she in particular was flagged.

8 CONCLUSION

Artificial intelligence in its assistive forms is doing valuable work in systems where delay is itself an injustice. The question is which decisions may be informed by statistical inference about a reference class, on what terms of visibility, and with what capacity for a person to answer back.

The present research puts forth some observations. The harm caused cannot be mitigated by technical sophistication. As seen earlier, Robodebt was arithmetic and SyRI a matching system. In practice both were catastrophic because each substituted an assumption for an assessment at scale. It places the burden of correction on the person least able to bear it. It is very common to see formal safeguards failing. Cautionary instructions did not constrain Wisconsin's sentencing courts. Similarly a judgment striking down SyRI did not prevent the Dutch benefits scandal. The comprehensive European statute has had its central obligations deferred by sixteen months for want of institutional readiness. Safeguards depending on a busy official noticing that a machine is wrong will not hold. Existing doctrine relating to the duty to give reasons, the right to be heard, the rule against dictation, the prohibition on arbitrariness, the proportionality standard for privacy intrusions is substantially adequate, and fails not for want of principle but for want of facts. Petitioners lose because they cannot examine the basis of the output.

The reform agenda requires interpretable models where liberty and subsistence are at stake. The price of placing reliance ought to be disclosed. Trade secrecy should be treated as a reason for protective orders rather than silence. There must be independent audits, and the results must be disaggregated and published. Human oversight must be engineered so that it survives contact with a heavy docket, and its performance be measured. The results of adjudication must be notified and explained independently along with providing a route to a second human determination. All these terms should be included in the contract. The agencies must also focus on capacity building initiatives to be able to enforce the obligations.

India is unusually well placed to get this right. Adoption is still mostly assistive, the jurisprudence on arbitrariness and proportionality is strong, and the governance framework is being drafted now rather than retrofitted later. The 2026 draft regulations correctly identify adjudication as non-delegable and

propose a sensible central approval pathway. The need is to extend the same to give the framework evidentiary teeth, without which every right in it is decorative.

* * *