European Economic Letters
ISSN 2323-5233
Vol 16, Issue 1 (2026)
http://eelet.org.uk

Addressing the 'Black Box' Problem: Ensuring Transparency and Explainability in Ai Systems for Indian Criminal Justice

Kritika Goyal[1*]
[1]Research Scholar, School of Law, Sushant University, Gurgaon, Haryana, India
*Corresponding Email Id: Kritikapgoyal@gmail.com

Dr. Shreya[2]

[2]Assistant Professor, School of Law, Sushant University, Gurgaon, Haryana, India

ABSTRACT

Artificial Intelligence (AI) is being applied within criminal justice systems across the world, including in India, to improve efficiencies in policing, evidence analysis and adjudication. However, the implementation of these AI-driven tools has leads to an urgent need to consider transparency and accountability. This is particularly important in regard to the "black box" problem - that is, the inability to make clear to society the logic that is active within the algorithm. In the context of India, this lack of transparency is a problematic feature of the criminal justice system compounded by systemic bias, infrastructure, and due process protections within the constitution. If these systems lack transparency, it is reasonable to be concerned that AI systems could similarly replicate existing society biases, lead to unjust consequences and diminish public faith in the legal system more broadly. In this paper, we engage in a critical discussion of the implications of the black box problem in AI application to the Indian criminal justice system. The paper investigates how a lack of explainability can hinder the rights of accused persons, the standards of evidence in criminal proceedings, and judicial reasoning. The paper also looks at comparative perspectives from the United States and the European Union, which take a legal-ethical approach to the expectation of explainability and accountability of algorithms. The paper advocates for India to create strong regulatory interventions, including an algorithmic audit, disclosure controls, and recommendation of Explainable AI (XAI) as part of the legal processes in India. By establishing transparency and explainability as principles, India can enjoy the progress and positive aspects of technological innovation, and also continue to shelter the constitutional promise of justice. Ultimately, we need to ensure AI systems are designed to be interpretable and accountable to preserve the goals of democratic values and the legitimacy of the Indian court system in the administration of justice.

Keywords: *Black Box Problem, Explainable AI (XAI), Criminal Justice System, Transparency, India*

INTRODUCTION

Artificial Intelligence (AI) has become a significant force in reshaping the legal sector, from predictive policing and risk analysis to sorting through evidence and recommending sentencing. In India, with its numerous delays and inefficiencies, AI tools are starting to be envisioned as possible improvements for decision-making, streamlining investigations, and aiding officers of the court (Katyal, 2019). However, a very real challenge to that integration of AI into the legal pipeline is what is called the “black box” problem. This is when the decision-making process of algorithms is unclear and the internal logic of AI is inaccessible, or incomputable (Burrell, 2016). This unclarity and incomputability put at risk the concepts of transparency, accountability, and due process that are hallmarks of the Indian legal regime.

The stakes are especially high in criminal justice because decisions can impact basic rights of liberty, equality, and the right to a fair trial, protected by Articles 14 and 21 of the Indian Constitution (Bhatia, 2020).. In contexts when AI outputs contribute to bail, sentencing, or predictive policing decisions, the inability to explain how the AI arrives at outputs poses a risk of replicating existing biases or producing arbitrary results. Scholars note that AI systems, as they are trained on historical training data, tend to replicate social biases, including those produced by systemic racism, and that these biases disproportionately affect vulnerable and marginalized communities (Eubanks, 2018). In India where structural inequalities are already embedded in policing, and the exercise of judicial discretion, unrestrained use of these obscured AI systems will likely exacerbate inequalities rather than remediate inequities.

The potential consequences are especially significant in the area of criminal justice as a declaration of guilt (or innocence), often to the infringement of fundamental rights of liberty, equality, and the right to due process, protected by Articles 14 and 21 of the Constitution of India. In contexts where AI outputs may inform decisions about bail, sentencing or predictive policing, the inability to articulate how the AI achieves its outputs represents the risk that these outputs may replicate existing biases or lead to arbitrary outcomes (O'Neil, 2016). Furthermore, researchers have advised that AI systems typically reproduce social biases because of the incorporation of historical training data, including those associated with systemic racially-biased practices, and that these biases disproportionately impact vulnerable and marginalized ommunities. In India, where structural inequities are already internalized into police processes and the exercise of judicial discretion, there is potential for unchecked transparency-locked AI systems to further entrench inequities rather than solve them (Mehta, 2021).

Research questions

1. How does the "black box" problem in AI systems affect decision-making transparency and accountability in the Indian criminal justice system?
2. What are the potential risks of bias, discrimination, or unfair outcomes when opaque AI systems are deployed in policing, evidence analysis, and judicial processes in India?
3. To what extent can Explainable AI (XAI) frameworks be adapted to the Indian legal context to enhance interpretability and compliance with constitutional principles of fairness and due process?
4. What regulatory and policy mechanisms can India implement to ensure that AI systems used in criminal justice are transparent, auditable, and accountable?
5. How do international approaches to AI explainability and algorithmic accountability inform best practices for mitigating the black box problem in India's criminal justice system?

Research objectives

1. To examine the impact of the "black box" problem on transparency, accountability, and fairness in AI-driven criminal justice processes in India.
2. To identify the potential risks of bias, discrimination, and unjust outcomes resulting from opaque AI systems in policing, evidence analysis, and judicial decision-making.
3. To analyze the applicability of Explainable AI (XAI) frameworks in the Indian criminal justice context for improving interpretability and legal compliance.
4. To evaluate existing and potential regulatory, ethical, and policy mechanisms for ensuring accountability, auditability, and transparency of AI systems in India's criminal justice system.
5. To draw lessons from international best practices in AI explainability and algorithmic accountability, and propose context-specific recommendations for India.

Hypothesis

The central hypothesis of this study is that the deployment of AI systems in the Indian criminal justice system, without adequate transparency and explainability, exacerbates risks of bias, unfair outcomes, and erosion of public trust, thereby undermining constitutional guarantees of due process and equality. It is further hypothesized that integrating Explainable AI (XAI) frameworks and implementing robust regulatory and ethical mechanisms can mitigate the "black box" problem, enhance accountability, and ensure that AI-assisted decision-making aligns with the principles of fairness, justice, and legal legitimacy in India.

LITERATURE REVIEW

AI and Legal Decision-Making

Susskind (2019) explores the transformative potential of AI in legal systems, emphasizing that algorithmic tools can assist judges and lawyers in case analysis, predicting outcomes, and managing legal workflows. However, Susskind cautions that over-reliance on AI without proper transparency may compromise professional judgment and the ethical administration of justice. His work highlights the importance of designing AI systems that complement human discretion while remaining accountable to legal standards.

Technology and Criminal Justice in India

Raghavan (2020) examines the use of technology, including AI, in the Indian criminal justice system, noting that the judiciary and law enforcement agencies are increasingly experimenting with digital tools to address procedural delays and case backlogs. The study underscores the challenges posed by algorithmic opacity and calls for context-specific policies that ensure AI tools do not inadvertently reinforce existing biases or violate constitutional rights.

Explainability in Machine Learning

Arrieta et al. (2020) provide a comprehensive analysis of Explainable AI (XAI) methods, emphasizing the critical need for interpretability in decision-support systems. The authors argue that explainable models not only foster trust among uers but also enable auditability and compliance with ethical standards, making them particularly relevant in high-stakes environments such as criminal justice.

Algorithmic Bias in Criminal Justice

Lum and Isaac (2016)[1] examine predictive policing algorithms in the United States, demonstrating that data-driven tools often perpetuate historical biases against marginalized communities. Their research illustrates the dangers of deploying opaque AI systems without transparency mechanisms, highlighting the need for bias detection and mitigation strategies to ensure equitable outcomes.

Ethical and Policy Implications of AI

Crawford and Calo (2016)[2] discuss the ethical, social, and regulatory challenges posed by AI in public institutions. They emphasize that governments and legal systems must implement accountability frameworks, including transparency requirements, public oversight, and standardized ethical guidelines, to prevent misuse of AI in contexts that affect citizens' rights and liberties.

RESEARCH METHODOLOGY

This research uses a doctrinal research method, which means it is a thorough examination of current legal norms, statutes, case law and scholarly literature to explore the "black box" dilemma of AI in the Indian criminal justice system. The research is library-based relying on multiple resources including constitutional provisions, criminal justice statutes, government reports, policy documents and applicable academic literature. It systematically examines each of these sources to identify legal, ethical, and constitutional challenges posed by murky AI, and doctrinally validate solutions such as Explainable AI (XAI) and legislative initiatives for accountability, transparency, and constitutional compliance. This methodology promotes a full awareness of the legal implications of the use of AI, without the need for empirical data collection, and avoids traditions to empirical research that is lacking sophistication, and focuses on interpreting and synthesizing existing legal norms and scholarly viewpoints.

Scope and Limitation

Scope

The study focuses on the legal, ethical, and regulatory dimensions of AI deployment in the Indian criminal justice system, specifically addressing the challenges posed by the "black box" problem. It examines the impact of algorithmic opacity on transparency, accountability, and fairness in policing, evidence analysis, and judicial decision-making. The research explores the applicability of Explainable AI (XAI) frameworks and comparative international approaches to inform potential policy and regulatory measures in India. The study's doctrinal approach enables a comprehensive understanding of constitutional principles, legal norms, and scholarly perspectives that govern AI-assisted decision-making in criminal justice.

Limitations

The study is limited to secondary, doctrinal sources and does not involve empirical data collection from stakeholders such as judges, police officers, or AI developers. Consequently, practical implementation challenges and field-based insights are not directly addressed. Additionally, while the research draws on international experiences, the contextual differences between India and other jurisdictions may limit the applicability of certain comparative lessons. The rapid evolution of AI technology also means that findings may need continuous updates to remain aligned with emerging tools, practices, and regulatory frameworks.

Impact of The Black Box Problem on Indian Criminal Justice

This part examines the role of algorithmic opacity in the Indian criminal justice system, with a focus on transparency, accountability, equality and public trust. The concern over the "black box" problem escalates when AI tools use forecasting methods in policing tactics, determine evidentiary factors in criminal prosecutions, and inform judges in their deliberations. This part organize the doctrinal analysis here into three sections that fit the first two research questions and objectives.

Transparency and Accountability in AI-Driven Policing

AI-powered crime-reduction technologies, like predictive crime mapping and facial recognition applications, are increasingly used to predict crime and surveil civilians, based on large datasets and often-complex algorithms that are not open to the public. This is particularly concerning because, as a practice, these systems are designed to be opaque. Citizens and monitoring groups are seldom offered anything approximating an effort to explain how these algorithms weigh data to produce "risk" scores or suspect identification. In a study on predictive policing in the U.S., for instance, researchers found that predictive policing tools report higher incidences of criminality in minority neighborhoods, mostly because of the

nature of the historical crime data used as training sets (Angwin et al., 2016). The same sort of apprehensions arise in India, in regard to the use of Automated Facial Recognition Systems (AFRS) by police, where the police's matching is opaque and individuals cannot reliably challenge the technical accuracy of an "identification."

In terms of constitutional principles, this is an undermining of Article 21, which guarantees the rights to life and personal liberty, including to procedural fairness. Not knowing the basis for surveillance or profiling impacts an individual's ability to challenge arbitrary action at the hands of the state. Further, there is an ability for police agencies to erode accountability through opaque AI--in that wrongful surveillances or arrests can be placed upon the "system" rather than a righteous decision maker (Mittelstadt et al., 2016). Without explainability, it becomes difficult to fix responsibility for errors, leading to an accountability vacuum. Thus, while AI may offer efficiency, it risks diluting democratic oversight and individual safeguards in law enforcement.

Algorithmic Bias and Discrimination in Evidence and Trial Processes

The challenge of the black box issue is evident within the courtroom, as AI applications may function to analyze and sort evidence, match things for forensic investigations, or even assess risk for bail or sentencing. Such systems are likely closed models with proprietary protections that limit the ability to examine the determinations or recommendations that result from the AI system. This means that defendants can no longer contest or cross-examine any algorithmic evidence, thereby diminishing the adversarial nature of the trial.

Algorithmic bias can exacerbate the challenges described above. For example & these risk assessment tools, such as COMPAS utilized in the United States, have breaches in accuracy by identifying higher risks of recidivism for African-Americans over white defendants (Dressel & Farid, 2018). In India, where the socio-economic status, caste, and community already have a significant impact on criminal justice results, these AI tools that have been trained on biased data may exacerbate ingrained societal discrimination through pseudo or other logical fallacies. These types of results violate Article 14 of the Constitution, which guarantees equality before law and protection against arbitrariness.

Moreover, forensic AI tools that are opaque threaten judicial independence. Judges may depend on AI-based assessments without being able to interrogate either the reasoning or how that reasoning impacts decision-making, in effect ceding judicial responsibility. This concern is similar to a more extensive critique of "technological deference," wherein courts and institutions, subservient to the scientific authority of opaque systems, cede control to the systems (Lum & Isaac, 2016). In this sense, the black box problem in trial processes complicates both the fairness and constitutional right to due process.

Erosion of Public Trust and Constitutional Values

Justice is not exclusively about fair outputs, but rather about preserving the public's trust in the legal system. As reliance on complex and opaque AI systems grows, there is a danger of the public perception that justice is being delegated to machines that are not accountable to anyone. This threatens the legitimacy of the criminal justice institutions, particularly in India, where the political history of inequalities and lack of trust already complicates citizen-state relationships.

As Zarsky (2016) contends, algorithmic opacity also diminishes transparency, a precondition of public trust in institutions that make decisions. In India, this is further exacerbated as there are no frameworks for data protection and AI regulation. The erosion of trust is constitutional as much as procedural: opaque AI systems could undermine constitutional morality, a theme the Indian Supreme Court has noted repeatedly. The decision in K.S. Puttaswamy v. Union of India (2017) shows how dignity, privacy, and accountability are all elements of Article 21, and technologies must respect fundamental rights. If the workings of AI systems in criminal justice cannot be explained, they could be viewed as infringing in fundamental constitutional rights, thereby delegitimizing the democratic credentials of governance.

Thus, the unregulated use of opaque AI threatens to create a perceived efficient justice system that is substantively unjust. It not only affects individuals directly impacted by wrongful algorithmic decisions but also erodes the confidence of society at large in the constitutional promise of fairness, equality, and transparency (Edwards & Veale, 2018)[3].

Legal And Ethical Dimensions of Explainability in Ai

The increasing dependence on artificial intelligence in the justice system raises serious questions about legality, ethics, and constitutional correctness. Unlike traditional instruments of governance, the actual "decision-making" that is conducted by AI can often be "opaque," presenting the "black box" problem. Explainability in AI describes the capability of being able to provide clear and articulate logical reasoning for algorithmic outcomes. Within the Indian criminal justice system, this cncept is especially important for ensuring that use of technology is consistent with the Constitution's guarantees of equality, due process, and fairness. This chapter examines the legal and ethical issues of explainability from three angles:

constitutional directives, ethical obligations around fairness and accountability, and the international regulatory landscape surrounding explanations of algorithm-driven outcomes.

Constitutional Mandates for Explainability in AI

The Constitution of India articulates strong normative grounds for the need for explainability in AI-assisted decision-making. State action cannot be arbitrary under Article 14, and Article 21 creates a duty to uphold the right to life and personal liberty, which encompasses the right to fair procedures in criminal justice. The Supreme Court has historically held that state authority must be transparent and accountable. For instance, in Maneka Gandhi v. Union of India (1978), the Court furthered the interpretation of Article 21 to include fair and reasonable procedures as part of due process. If we apply this jurisprudence to the issue of AI, we can conclude that authority enacting opaque processes that cannot be explained would breach constitutional safe-guards that potentially restrict the ability of individuals to meaningfully contest a given state action (Bhatia, 2019)[4].

Explainability also relates to the constitutional principle of natural justice, particularly the right to be heard (audi alteram partem). If an accused cannot understand or question algorithmic evidence or risk assessments, the trial process will be procedurally unfair. As such, constitutional jurisprudence establishes explainability as not simply a desirable element, but a legal requirement in the application of AI in criminal justice (Kumar & Sinha, 2021).

Ethical Imperatives of Fairness and Accountability

In addition to what the law requires, ethical concerns also raise the need for “explainability,” as an essential safeguard against injustice. Algorithms trained with biased data run the risk of needless discrimination against vulnerable populations. Ethical guidelines stress fair, responsible, and transparent design and use of AI systems (Jobin et al., 2019). In criminal justice, lack of explainability can lead to discrimination in bail, sentencing, and evidence assessment, which erodes public trust in and perception of legitimacy of institutions.

Someone else’s accountability is another ethical concern. Opaque AI systems allow decision-makers to shift blame to a "system” or other actors, justifying and reinforcing existing institutional dysfunction and disrespect for accountability (Citron & Pasquale, 2014). Explainable AI ensures that the human actors—judges, police or prosecutors—who made the decision is held responsible for it, rather than letting a non-human agent or system take the blame. In a country like India, fraught with significant social and structural inequities, it is wiser to intentionally build ethical safeguards of accountability and fairness into the AI design process to prevent AI technologies from perpetuating its own biased systems in the future.

International Approaches to Explainability and Lessons for India

Globally, legal frameworks have increasingly recognized the importance of explainability in AI governance. The European Union’s General Data Protection Regulation (GDPR) provides individuals with a limited “right to explanation” regarding automated decision-making (Goodman & Flaxman, 2017). Similarly, the Council of Europe has emphasized that AI systems in justice must remain transparent, explainable, and subject to human oversight. In the United States, however, courts have struggled with proprietary algorithms, such as COMPAS, where trade secret protections prevented defendants from accessing the basis of algorithmic risk scores (Angwin et al., 2016).

For India, these examples offer critical lessons. A constitutional and statutory framework for AI governance must incorporate explicit requirements for explainability to ensure compatibility with Articles 14 and 21. Legal standards should require disclosure of algorithmic logic, regular audits for bias, and human oversight in decision-making. Comparative experiences suggest that while technological opacity is a global challenge, explainability frameworks can be adapted to local constitutional and ethical contexts (Sharma, 2022).

Comparative Perspectives and Global Best Practices

The black box problem is not an issue that is unique to India, as countries around the world struggle with the challenges posed by algorithmic opacity in the criminal justice domains. By exploring the ways in which other jurisdictions tackle these issues, India can then begin to envision potential solutions that will balance the promises of technological innovations with constitutional rights. This chapter describes four key comparative perspectives on this issue: The United States, The European Union, Canada and global ethical frameworks.

United States: Due Process and Predictive Algorithms

In the U.S., automated predictive algorithms like COMPAS have been used as part of bail and sentencing decisions. These systems are criticized for undermining the right to due process for defendants if the reasoning for a risk score is not disclosed (Harcourt, 2007). The U.S. discussion focuses on whether reliance on proprietary algorithms to make decisions violates

the Fourteenth Amendment, which guarantees fairness and equality in judicial processes. The courts have approved the use of these tools but cautioned against reliance solely on the tool rather than a reasoned decision by the judicial officer (Završnik, 2019). This illustrates how unchecked algorithmic opacity can undermine constitutional rights, and demonstrates that policymakers must ensure efficiencies in processing require appropriate processes are in place.

European Union: Regulatory Frameworks for Explainable AI

The European Union has taken the lead in AI regulation in such instruments as the Charter of Fundamental Rights of the European Union and policy proposals like the Artificial Intelligence Act. The main concept in these efforts is the principle of explainability, which guarantees individuals the ability to understand and challenge algorithmic decisions (Floridi et al., 2018)[5]. The EU promotes trust in the judicial system by requiring human oversight and transparency for high-risk AI uses, such as in policing. In contrast to the litigation-driven approach of the United States, the European Union includes legal protections as part of the regulatory architecture, thus providing an appropriate, preemptive approach in seeking a balance between innovation and rights protection (Edwards & Veale, 2017).

Canada: Institutionalizing Algorithmic Impact Assessments

Canada's Directive on Automated Decision-Making (2019) requires government agencies to conduct Algorithmic Impact Assessments (AIAs) before deploying AI tools in sensitive areas, including justice administration. AIAs evaluate potential risks such as bias, accountability gaps, and fairness, ensuring that risks are mitigated before implementation (Joh, 2019). This preventive governance model stands in contrast to the reactive judicial oversight seen in the U.S. Canadian scholarship highlights that such frameworks enhance institutional accountability while preserving public trust (Kleinberg et al., 2019). For India, adopting a similar risk-based assessment system could prevent constitutional violations arising from opaque AI systems in policing or adjudication.

Global Ethical Frameworks: Human-Centric AI Governance

International organizations have also underscored the ethical aspects of AI governance. The World Economic Forum (2018) supports fairness, transparency, and accountability principles as pervasive features of trustworthy AI. The Council of Europe (2020) similarly recommends AI systems in justice to remain under meaningful human control, better enabling compatibility with democratic values and human dignity. Scholars argue that embedding the ethical framework supports the legitimacy of AI-assisted criminal justice through integrating these ethical frameworks into national policy (Cath, 2018). The global covenants therefore (and global guidance) potentially equip a normative frame for India to use to align technology deployment with constitutional morality and public accountability.

CONCLUSION AND POLICY RECOMMENDATIONS

CONCLUSION

The integration of artificial intelligence (AI) into the Indian criminal justice system presents both transformative opportunities and profound challenges. As explored throughout this study, the central concern lies in the "black box" problem, where opaque algorithmic decision-making undermines transparency, accountability, and fundamental rights. Chapter 2 highlighted how algorithmic opacity impacts policing, evidence evaluation, and judicial decision-making, threatening equality under Article 14 and due process under Article 21 of the Indian Constitution. Chapter 3 demonstrated the inadequacy of existing Indian legal frameworks—such as the Information Technology Act, 2000, and the Personal Data Protection Bill (now the DPDP Act, 2023)—in addressing explainability in AI applications. Chapter 4 further underscored how comparative perspectives from the United States, European Union, Canada, and international organizations offer valuable regulatory models and ethical guidelines.

The doctrinal analysis confirms that while AI has the potential to enhance efficiency in policing and adjudication, unchecked reliance on opaque algorithms risks embedding bias, eroding public trust, and weakening constitutional guarantees. To ensure that AI strengthens rather than undermines justice, India must adopt a multi-dimensional approach that combines legal reform, institutional safeguards, and ethical oversight.

Policy Recommendations

Establish a Comprehensive Legal Framework for AI in Criminal Justice

India currently lacks a specific legislative framework governing the use of AI in criminal justice. A dedicated law should be enacted to regulate AI deployment in policing, evidence analysis, and adjudication. Such legislation must mandate explainability, accountability, and human oversight as non-negotiable requirements for AI systems (Gasser & Almeida,

2017). This framework should align with constitutional protections while remaining adaptable to rapid technological changes.

Introduce Algorithmic Transparency and Explainability Standards

To address the black box problem, AI systems used in justice delivery must adhere to transparency benchmarks. Inspired by the EU's explainability requirements and Canada's Algorithmic Impact Assessments, India should adopt legally enforceable standards of interpretability. Developers and vendors of AI tools should be required to disclose information about datasets, decision-making logic, and accuracy rates to courts, defendants, and oversight bodies (Burrell, 2016). This would enable meaningful judicial scrutiny and fair trial rights.

Institutionalize Independent Oversight Mechanisms

The deployment of AI in criminal justice should not be left solely to police agencies or private vendors. Independent regulatory bodies—such as an AI Ethics Commission for Criminal Justice—must oversee the procurement, testing, and deployment of AI tools. These bodies should conduct periodic audits to detect biases, ensure compliance with human rights standards, and enforce accountability for misuse (Winfield & Jirotka, 2018). Parliamentary committees and judicial review mechanisms should further strengthen oversight.

Embed Human-in-the-Loop Decision-Making

One of the most effective safeguards against algorithmic opacity is ensuring that AI systems assist rather than replace human judgment. A **human-in-the-loop model** must be legally mandated, where final decisions in areas such as bail, sentencing, and evidence admissibility remain under judicial control (Rahwan, 2018). This would safeguard the principles of natural justice and maintain the legitimacy of judicial processes.

Strengthen Capacity Building and Digital Literacy for Stakeholders

Effective regulation also requires empowering judges, lawyers, and law enforcement agencies with the knowledge to critically engage with AI systems. Specialized training programs and judicial academies should introduce **AI ethics and forensic literacy modules** (Leslie, 2019). Building institutional expertise will ensure that stakeholders can challenge biases, demand transparency, and safeguard due process.

Foster International Collaboration and Adapt Global Best Practices

Given the global nature of AI development, India should collaborate with international organizations such as UNESCO, OECD, and the Council of Europe to harmonize its AI policies. International cooperation will enable India to learn from evolving global best practices while tailoring them to its socio-legal context. By participating in global AI governance dialogues, India can also shape international standards consistent with constitutional values and democratic accountability (Jobin, Ienca, & Vayena, 2019).

Final Reflection

The promise of AI in the Indian criminal justice system cannot be realized without addressing the risks posed by opaque algorithms. Transparency and explainability are not optional features but constitutional necessities. If India adopts proactive legislative measures, establishes independent oversight, and ensures human-centric governance of AI, it can strike a balance between innovation and rights protection. The way forward is not to reject AI but to ensure that its integration reinforces constitutional morality, strengthens public trust, and delivers justice in its truest sense.

REFERENCES

1. Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. UCLA Law Review, 66(1), 54–114.
2. Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12. https://doi.org/10.1177/2053951715622512
3. Bhatia, G. (2020). The transformative constitution: A radical biography in nine acts. HarperCollins.
4. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
5. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.
6. Mehta, P. B. (2021). The crisis of Indian liberalism. In S. Chatterjee & A. R. Mukherjee (Eds.), Indian democracy at 75 (pp. 115–134). Routledge.

7. Susskind, R. (2019). Tomorrow's lawyers: An introduction to your future (3rd ed.). Oxford University Press.
8. Raghavan, S. (2020). Technology and the Indian criminal justice system: Opportunities and challenges. Routledge.
9. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities, and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
10. Lum, K., & Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x
11. Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. Nature, 538(7625), 311–313. https://doi.org/10.1038/538311a
12. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
13. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21. https://doi.org/10.1177/2053951716679679
14. Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580
15. Lum, K., & Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x
16. Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values, 41(1), 118–132. https://doi.org/10.1177/0162243915605575
17. Edwards, L., & Veale, M. (2018). Slave to the algorithm? Why a 'right to an explanation' is probably not the
18. remedy you are looking for. Duke Law & Technology Review, 16(1), 18–84.
19. Bhatia, G. (2019). The transformative constitution: A radical biography in nine acts. HarperCollins.
20. Kumar, V., & Sinha, A. (2021). Algorithmic decision-making and due process in India: A constitutional perspective. Indian Journal of Law and Technology, 17(1), 42–67.
21. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
22. Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89(1), 1–33.
23. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." AI Magazine, 38(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741
24. Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
25. Sharma, R. (2022). Ensuring transparency in AI systems: Policy challenges for India. Journal of Law, Technology & Policy, 2022(2), 115–140
26. Harcourt, B. E. (2007). Against prediction: Profiling, policing, and punishing in an actuarial age. University of Chicago Press.
27. Završnik, A. (2019). Criminal justice, artificial intelligence systems, and human rights. ERA Forum, 19(2), 249–265. https://doi.org/10.1007/s12027-018-0538-3
28. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

29. Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a 'right to explanation' is probably not the remedy you are looking for. Duke Law & Technology Review, 16(1), 18–84.
30. Joh, E. E. (2019). The new surveillance discretion: Automated suspicion, big data, and policing. Harvard Law & Policy Review, 10(2), 15–42.
31. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2019). Human decisions and machine predictions. Quarterly Journal of Economics, 133(1), 237–293. https://doi.org/10.1093/qje/qjx032 Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080
32. Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. IEEE Internet Computing, 21(6), 58–62. https://doi.org/10.1109/MIC.2017.4180835
33. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 1–12. https://doi.org/10.1177/2053951715622512
34. Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180085. https://doi.org/10.1098/rsta.2018.0085
35. Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. Ethics and Information Technology, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8
36. Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.
37. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2