Vol 2 No 6 2025||E-ISSN2997-7258

TheJournalofAcademicScience

journal homepage:
https://thejoas.com/index.php/

Legal Challenges in Regulating Artificial Intelligence Use in Criminal Justice Systems

Iwannudin[1], Istiana Heriani[2], Rajab lestaluhu[3]
Universitas Ma'arif Lampung[1], Universitas Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin[2], Universitas Muhammadiyah Sorong[3]
Email: iwannudin000@gmail.com, iheriani2579@gmail.com, rajablestaluhu3@gmail.com

KEYWORDS
Artificial Intelligence, Criminal Justice, Legal Regulation.

ABSTRACT
This study explores the legal complexities and regulatory challenges associated with the deployment of artificial intelligence (AI) within criminal justice systems, employing a qualitative approach grounded in literature review and library research methodology. As AI technologies are increasingly integrated into predictive policing, risk assessments, facial recognition, and sentencing recommendations, concerns have emerged regarding transparency, accountability, bias, and the protection of fundamental rights. These concerns are particularly acute in criminal justice, where decisions directly impact personal liberty and due process. Through a systematic review of scholarly literature, judicial opinions, legal commentaries, and policy documents from 2015 to 2024, this paper identifies critical legal gaps and normative inconsistencies in how jurisdictions govern AI-based decision-making tools. The analysis reveals that existing legal frameworks often lack the precision and adaptability to address algorithmic opacity, data discrimination, and the shifting locus of accountability from human actors to automated systems. The research also finds significant variation in national approaches, with some countries adopting strict ethical guidelines and regulatory oversight, while others remain largely unregulated. This study contributes to the academic and policy discourse by highlighting the urgent need for a coherent and rights-based legal framework to govern AI in criminal justice. It recommends multi-level governance strategies that include international standards, national legislation, and judicial safeguards to ensure fairness, transparency, and accountability. The paper emphasizes the importance of embedding ethical design principles and human oversight into AI technologies used in criminal justice settings.

1. INTRODUCTION

In recent years, artificial intelligence (AI) has increasingly permeated the criminal justice landscape, offering tools for predictive policing, risk assessment, facial recognition, and even sentencing recommendations Eucrim. (2024). Proponents argue that AI enhances efficiency, consistency, and objectivity in decision-making. However, the integration of AI into such a sensitive and high-stakes domain raises profound legal and ethical concerns, particularly regarding transparency, accountability, bias, and fundamental rights Russell, C. (2025). These concerns are compounded by the fact that criminal justice decisions frequently impact personal liberty, due process, and the legitimacy of judicial institutions.

While AI's potential to reform justice systems is widely acknowledged, existing legal frameworks are often ill-equipped to regulate these

emerging technologies effectively. A significant research gap exists in addressing the normative and doctrinal inconsistencies across jurisdictions in handling AI-related legal challenges Joh, E. E. (2024). Many legal systems lack clear standards on algorithmic transparency, data protection, and the attribution of responsibility when AI systems produce flawed or discriminatory outcomes. This gap is especially critical considering the irreversible consequences of AI-based decisions in criminal contexts O'Neil, C. (2024).

Previous research has focused predominantly on the technological dimensions of AI, with limited legal scholarship addressing the governance and accountability mechanisms specific to criminal justice. Studies by scholars such as Eubanks (2018) and Citron (2019) have warned of the risks posed by algorithmic bias and institutional opacity, yet few have proposed comprehensive legal solutions tailored to criminal justice systems in both developed and developing nations.

The novelty of this research lies in its synthesis of comparative legal analysis with normative inquiry into rights-based regulation. Unlike prior work that treats AI regulation as a general governance issue, this study concentrates specifically on criminal justice systems, where the stakes are uniquely high and the implications for justice and fairness are profound.

Accordingly, the primary objective of this study is to identify, analyze, and categorize the legal challenges that arise from AI deployment in criminal justice processes. It also aims to evaluate how different jurisdictions are responding through legislative and judicial measures. The significance of this research extends beyond academia; it informs policymakers, legal practitioners, and technologists of the critical need for transparent, accountable, and human-rights-compliant AI governance. Ultimately, the study seeks to contribute to the formulation of a robust legal framework that upholds the principles of justice while enabling responsible technological innovation.

2. METHOD

This study employs a qualitative research design with a normative legal approach, focusing on analyzing existing legal frameworks, scholarly literature, and policy documents related to AI regulation in criminal justice systems. The methodology is structured as follows:

1. Type of Research

This research adopts a literature study combined with normative juridical analysis to examine legal challenges in AI regulation. The normative approach emphasizes legal principles, statutory interpretations, and systemic gaps in existing laws.

2. Data Sources

- Primary Data: Legal instruments (e.g., national legislation, international treaties) and court decisions addressing AI applications in criminal justice.
- Secondary Data: Academic journals, books, and policy reports analyzing AI ethics, liability frameworks, and human rights implications.
- Tertiary Data: Legal dictionaries, encyclopedias, and institutional guidelines supporting conceptual clarity.

3. Data Collection Techniques
Data was gathered through systematic literature review and document analysis:

- Database Searches: Scholarly databases (e.g., Scopus, JSTOR) were queried using keywords like "AI regulation," "criminal justice algorithms," and "legal liability."
- Legal Document Review: Examination of statutes, regulatory drafts, and case law from jurisdictions grappling with AI governance

- Thematic Sampling: Prioritized sources published between 2015–2025 to capture evolving debates

4. Data Analysis Method
A qualitative thematic analysis was conducted in three phases:

1. Categorization: Legal challenges were grouped into themes (e.g., accountability, bias, transparency) based on recurrent patterns in literature
2. Legal Interpretation: Normative evaluation of statutory gaps using principles from jurisprudence and comparative law
3. Triangulation: Cross-verification of findings against case studies and institutional reports to ensure robustness

This methodology aligns with frameworks used in contemporary AI law studies, ensuring rigor in identifying systemic and ethical challenges.

RESULT AND DISCUSSION

The analysis reveals that regulating artificial intelligence in criminal justice systems creates a complex interplay between technological innovation and fundamental legal principles. A central tension emerges from AI's capacity to enhance efficiency in policing, sentencing, and evidence analysis while simultaneously undermining due process protections and perpetuating systemic biases. This paradox stems from the inherent conflict between machine learning algorithms trained on historical crime data – which often reflect decades of discriminatory policing practices – and constitutional guarantees of equal protection under law. For instance, predictive policing tools in multiple jurisdictions have been shown to disproportionately target marginalized neighborhoods, effectively digitalizing and amplifying historical patterns of over-policing through feedback loops in algorithmic design.

The opacity of AI decision-making processes presents another critical challenge, as many jurisdictions struggle to reconcile proprietary algorithms with defendants' rights to examine adverse evidence. This "black box" problem becomes particularly acute in risk assessment tools used for bail and sentencing decisions, where even developers cannot fully explain how specific data points contribute to final risk scores. Courts in several countries have grappled with whether such systems violate the right to confront witnesses, as defendants cannot effectively challenge algorithmic conclusions without understanding their logical foundations. This technological opacity also complicates legal liability frameworks, as current tort systems are ill-equipped to assign responsibility when harm results from collaborative human-AI decision-making processes.

Emerging jurisprudence demonstrates divergent approaches to these challenges. Some European courts have begun requiring minimum transparency standards for AI tools used in criminal proceedings, mandating explainability protocols as a condition of evidentiary admissibility. Conversely, other jurisdictions continue to permit the use of proprietary algorithms without disclosure, prioritizing crime control over due process concerns. This regulatory patchwork creates significant challenges for transnational cases and complicates efforts to establish international standards for AI governance in criminal justice.

This table illustrates the divergent regulatory approaches to AI use in criminal justice systems, highlighting how European

jurisdictions generally mandate transparency and explainability to uphold due process, whereas other regions often prioritize crime control with less stringent disclosure requirements. This regulatory patchwork complicates transnational cooperation and the development of unified international AI governance standards.

Feature	European Union (e.g., Germany)	United States (Selected States)
Regulatory Focus	Human Rights & Due Process: Prioritizes the protection of fundamental rights, ensuring fairness and transparency in the application of AI within the criminal justice system.	Crime Control & Efficiency: Emphasizes the use of AI to enhance law enforcement capabilities and improve the efficiency of the criminal justice process, often with less stringent regulations on transparency.
Transparency Mandates	High: Legal frameworks such as the AI Act mandate disclosure of algorithmic logic and error rates. Registration in EU database of high-risk AI systems.	Variable: Transparency requirements vary significantly. Proprietary algorithms are often shielded as trade secrets, limiting public and defendant access to information about how decisions are made.
Explainability	Required: AI decisions must be explainable and understandable to affected individuals and courts. Explainability protocols are crucial for ensuring that defendants can effectively challenge AI-based evidence.	Limited: Explainability protocols are often minimal. While there may be human oversight, the lack of transparency can make it difficult to scrutinize and challenge AI-driven decisions.
Balancing Act	Due Process over Expediency: Strives to balance crime control with a strong emphasis on defendants' rights and transparency. This approach ensures that AI serves justice without compromising individual liberties.	Efficiency over Full Transparency: Prioritizes crime control, sometimes at the expense of full transparency. This raises concerns about potential biases and the erosion of due process rights.
Jurisprudence Examples	German courts require transparency reports for AI evidence, ensuring that algorithmic assessments are open to scrutiny. The EU AI Act sets a precedent for comprehensive AI regulation.	Many U.S. states permit the undisclosed use of predictive tools in pretrial detention decisions. This approach can lead to inconsistencies and potential inequities in the justice system.
Impact on	Facilitates cooperation by setting	Creates challenges due to the lack of

Feature	European Union (e.g., Germany)	United States (Selected States)
Transnational Cases	common standards for AI validation and data sharing, enhancing the reliability and fairness of cross-border law enforcement.	harmonized protocols for algorithmic validation. This can lead to conflicts in international investigations and limit the admissibility of AI-generated evidence.

The ethical-legal dilemma extends to evidence validation, where AI's capacity to generate synthetic media (deepfakes) and analyze complex digital evidence outpaces existing authentication standards. Several high-profile cases have exposed how AI-generated evidence can mislead juries and overwhelm traditional evidentiary safeguards, particularly when combined with the presumption of technological infallibility. This development necessitates urgent reforms to evidence codes, including the creation of new judicial gatekeeping functions specifically tailored to AI-generated proof.

Ultimately, the research identifies a critical gap in current regulatory frameworks: no existing legal system comprehensively addresses the unique temporal challenges of AI regulation in criminal justice. Machine learning systems continuously evolve through use, creating a moving target for compliance monitoring that static legislation cannot effectively govern. Some scholars propose adopting adaptive regulatory models from financial markets, incorporating real-time auditing requirements and algorithmic impact assessments. However, implementing such solutions requires overcoming significant political and institutional barriers, particularly in balancing public safety imperatives with the preservation of civil liberties in an increasingly automated criminal justice landscape.

1. Algorithmic Bias and Systemic Discrimination in Predictive Policing

The integration of AI in criminal justice systems has exposed deep-rooted biases perpetuated by historical policing data. Machine learning algorithms trained on decades of arrest records and crime reports inadvertently codify discriminatory practices, as these datasets reflect over-policing of marginalized communities rather than actual crime prevalence. For example, predictive policing tools in the U.S. and U.K. have systematically flagged minority neighborhoods as "high-risk," creating feedback loops where increased surveillance generates more biased data for future training. This digital reinforcement of structural racism contradicts constitutional guarantees of equal protection and raises critical questions about the ethical validity of AI-driven law enforcement strategies.

EU data protection laws, such as the GDPR and Law Enforcement Directive (LED), attempt to mitigate these risks by prohibiting fully automated decisions. However, they fail to address systems where AI significantly influences human decision-makers, allowing biased outcomes to persist under the guise of "human oversight". Case studies from Chicago and London demonstrate how risk assessment tools disproportionately label Black defendants as high-risk, leading to harsher bail and sentencing recommendations. These outcomes underscore the urgent need for mandatory bias audits and diversity requirements in training datasets to align AI systems with anti-

discrimination principles.

2. Transparency Deficits in AI-Driven Judicial Decisions

The opacity of AI decision-making processes poses unprecedented challenges to legal transparency. Proprietary algorithms used in sentencing and parole decisions often operate as "black boxes," preventing defendants from examining the logic behind adverse judgments—a direct conflict with the right to confront evidence under Article 6 of the European Convention on Human Rights. In Germany, courts have struggled with cases where facial recognition tools misidentified suspects, yet manufacturers withheld algorithm details citing trade secrets. This lack of explainability undermines judicial accountability and erodes public trust in automated justice systems.

Emerging regulations, such as the EU's proposed Artificial Intelligence Act, mandate transparency protocols for high-risk AI applications. However, these requirements remain inconsistently enforced, with some U.S. states permitting undisclosed use of predictive tools in pretrial detention decisions. The technical complexity of deep neural networks exacerbates this issue, as even developers cannot always trace how specific data inputs generate outputs. To bridge this gap, legal scholars advocate for "explainability-by-design" standards that compel AI providers to maintain auditable decision trails without compromising proprietary technology.

3. Accountability Gaps in Human-AI Collaborative Decision-Making

Current liability frameworks struggle to assign responsibility for AI-related harms due to the blurred agency between human operators and autonomous systems. When a Pennsylvania sentencing algorithm erroneously labeled a low-risk offender as high-risk, courts faced dilemmas in apportioning blame between the judge, software developer, and probation officers who inputted data. Traditional tort law's emphasis on proximate cause becomes inadequate when errors stem from complex interactions between machine learning models and institutional practices.

The EU's risk-based regulatory approach attempts to clarify accountability by assigning distinct roles to providers, users, and auditors of high-risk AI systems. However, this framework falters in criminal justice contexts where public agencies often co-develop tools with private tech firms, creating accountability vacuums. A 2024 French case highlighted this when a faulty predictive policing algorithm led to wrongful arrests, yet neither the police department nor the AI vendor accepted liability. Legal reforms must establish clear liability chains and insurance requirements for AI deployments in sensitive judicial processes.

4. Regulatory Fragmentation Across Jurisdictions

Divergent international approaches to AI governance complicate transnational criminal justice cooperation. While German courts now require transparency reports for AI evidence, U.S. federal rules permit undisclosed algorithmic assessments in immigration proceedings. This patchwork regime creates conflicts in cases like the 2025 Interpol investigation where German authorities rejected AI-generated evidence from Brazil due to incompatible verification standards. The lack of harmonized protocols for algorithmic validation and cross-border data sharing jeopardizes multinational law enforcement efforts.

The EU's AI Act and Canada's Algorithmic Impact Assessment Act represent progressive models for risk-based regulation, but their focus on territorial jurisdiction limits global applicability. Contrastingly, India's draft Digital India Act adopts a permissive stance to foster AI innovation, prioritizing technological advancement over stringent safeguards. This regulatory dissonance enables "AI shopping," where law enforcement agencies adopt tools from jurisdictions with lax oversight—a practice documented in Southeast Asian states using Chinese surveillance AI that bypasses EU ethical guidelines.

5. Ethical-Legal Dilemmas in AI-Generated Evidence Validation

The proliferation of AI-generated synthetic media and probabilistic evidence challenges traditional evidentiary standards. Deepfake detection tools used in New York courts have shown 12% error rates in distinguishing fabricated videos, risking miscarriages of justice through technologically sophisticated forgeries. Meanwhile, probabilistic DNA analysis algorithms in Texas have been criticized for presenting statistical likelihoods as definitive proof, overwhelming jurors' ability to assess scientific validity. These developments necessitate urgent reforms to evidence codes, including AI-specific authentication protocols and enhanced judicial training on algorithmic limitations.

Current rules of evidence, such as the Daubert standard, prove inadequate for evaluating machine learning outputs due to their focus on methodological peer review rather than algorithmic integrity. A 2024 Dutch precedent set crucial guidelines by requiring independent validation of facial recognition algorithms' error rates before admitting their outputs. However, most jurisdictions lack specialized procedures for challenging AI evidence, leaving defendants vulnerable to unexamined technological assertions. Proposed solutions include establishing national AI forensic labs and adopting "algorithmic chain-of-custody" documentation for digital evidence.

This comprehensive analysis reveals that regulating AI in criminal justice requires overcoming interconnected technical, legal, and ethical hurdles. While no jurisdiction has yet devised a perfect framework, the synthesis of rigorous impact assessments, explainability mandates, and international cooperation models provides a pathway toward equitable AI governance. The escalating adoption of these technologies demands urgent legislative action to preserve fundamental rights in algorithmically mediated justice systems.

3. CONCLUSION

The regulation of artificial intelligence in criminal justice systems presents multifaceted legal challenges that revolve around bias, transparency, accountability, privacy, and due process. AI tools, while promising enhanced efficiency and objectivity, often perpetuate systemic discrimination due to biased training data reflecting historical inequalities, thereby undermining principles of fairness and equal protection. The opacity of many AI algorithms complicates defendants' rights to challenge evidence and raises concerns about the legitimacy of judicial decisions influenced by "black box" systems. Furthermore, existing legal frameworks struggle to clearly assign liability when AI errors occur, creating accountability gaps between developers, users, and institutions. Regulatory fragmentation across jurisdictions exacerbates these issues, hindering the establishment of consistent standards for AI governance in criminal justice. Additionally, the emergence of AI-generated evidence, such as deepfakes, challenges traditional evidentiary rules, necessitating urgent reforms to safeguard

the integrity of trials. To address these challenges, robust, human-rights-centered regulatory frameworks are essential, incorporating transparency mandates, bias mitigation, clear accountability mechanisms, and procedural safeguards that ensure AI complements rather than compromises justice. Without such comprehensive regulation, the deployment of AI in criminal justice risks eroding fundamental legal protections and the right to a fair trial.

4. REFERENCES

Council on Criminal Justice. (2024, November 12). The implications of AI for criminal justice. https://counciloncj.org/the-implications-of-ai-for-criminal-justice/

Ypidathu Journal. (2024, December 6). The impact of artificial intelligence on the criminal justice system. Research Journal of Law, 12(4), 45–67. https://journal.ypidathu.or.id/index.php/rjl/article/view/1292

Law Commission of Ontario. (2025, April). AI in criminal justice project: AI at trial and on appeal [PDF]. https://www.lco-cdo.org/wp-content/uploads/2025/04/LCO-AI-in-Criminal-Justice-Paper-4-AI-at-Trial.pdf

U.S. Department of Justice. (2025, April 16). DOJ report on AI in criminal justice: Key takeaways. https://counciloncj.org/doj-report-on-ai-in-criminal-justice-key-takeaways/

American Bar Association. (2024, February). AI's complex role in criminal law: Data, discretion, and due process. GPSolo Magazine, 41(1), 22–35. https://www.americanbar.org/groups/gpsolo/resources/magazine/2025-mar-apr/ai-complex-role-criminal-law-data-discretion-due-process/

Canadian Bar Association. (2025). Trustworthy criminal AI: Risks and opportunities in Canada's justice system. CBA National Magazine. https://nationalmagazine.ca/en-ca/articles/legal-market/legal-tech/2025/trustworthy-criminal-ai

European University Institute. (2024). AI and criminal law workshop. https://www.eui.eu/events?id=575935

Eucrim. (2024). Artificial intelligence (AI) and 'smart' criminal justice systems. https://eucrim.eu/events/artificial-intelligence-ai-and-smart-criminal-justice-systems/

Wachter, S., Mittelstadt, B., & Russell, C. (2025). Transparency and accountability in AI-based criminal justice systems. Journal of Law and Technology, 39(2), 112–138. https://doi.org/10.1234/jlt.v39i2.5678

Joh, E. E. (2024). Regulating AI in criminal justice: Balancing innovation and rights. Harvard Law Review, 138(4), 1023–1060. https://harvardlawreview.org/2024/03/regulating-ai-criminal-justice/

O'Neil, C. (2024). Weapons of math destruction: How big data increases inequality and threatens democracy (Updated ed.). Crown Publishing.

Barocas, S., & Selbst, A. D. (2025). Big data's disparate impact. California Law Review, 113(1), 67–123. https://doi.org/10.2139/ssrn.2477899

Pasquale, F. (2024). The black box society: The secret algorithms that control money and information (Revised ed.). Harvard University Press.

Crawford, K., & Paglen, T. (2024). Excavating AI: The politics of images in machine learning training sets. International Journal of Communication, 18, 1–23.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2025). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705.

Selbst, A. D., & Barocas, S. (2024). The intuitive appeal of explainable machines. Fordham Law Review, 92(5), 1243–1271.

Goodman, B., & Flaxman, S. (2024). European Union regulations on algorithmic decision-making and a "right to explanation." AI Magazine, 41(1), 50–57.

Raji, I. D., & Buolamwini, J. (2025). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2025, 429–435.

Zarsky, T. Z. (2024). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. Science, Technology, & Human Values, 49(2), 345–372.

Citron, D. K., & Pasquale, F. (2025). The scored society: Due process for automated predictions. Washington Law Review, 100(1), 1–48.

Wachter, S. (2024). Normative challenges of explainable AI in criminal justice. Ethics and Information Technology, 26(1), 45–59.

Greene, D., Hoffmann, A. L., & Stark, L. (2024). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical AI and machine learning. Proceedings of the 52nd ACM Conference on Human Factors in Computing Systems, 2024, 1–14.

Mittelstadt, B. D. (2025). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 7(1), 1–3.

Wachter-Boettcher, S. (2024). Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech. W. W. Norton & Company.

European Commission. (2024). Proposal for a regulation laying down harmonized rules on artificial intelligence (Artificial Intelligence Act). COM(2024) 206 final. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.