IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

From Black-Box to Glass-Box: Redefining Transparency Standards for Legal AI Systems

Dr. Mrityunjai Pandey
Professor, Innovative Institute of Law, Greater Noida
Mrityunjai_pandey@yahoo.com

Abstract: *The growing integration of Artificial Intelligence (AI) into legal and quasi-legal decision-making has intensified concerns about opacity, bias, and accountability. Many contemporary AI systems function as "black boxes," producing outcomes that are difficult to interpret or challenge—an approach fundamentally at odds with legal principles of reasoned decision-making, procedural fairness, and transparency. This paper argues for a shift from black-box models to a "glass-box" framework that redefines transparency standards for legal AI systems. It conceptualizes explainability not merely as a technical feature but as a normative legal requirement grounded in due process, accountability, and data protection law. Drawing on regulatory developments such as the EU General Data Protection Regulation and the EU Artificial Intelligence Act, the paper proposes a context-sensitive transparency model that aligns explanations with legal stakes, affected stakeholders, and regulatory objectives. It further examines the tension between explainability, privacy, and proprietary interests, advocating a tiered disclosure approach. The study concludes that glass-box transparency is essential to maintaining legitimacy, trust, and fairness in AI-assisted legal decision-making and calls for embedding explainability as a core principle of legal AI governance.*

Keywords: Legal AI; Explainable AI; Algorithmic Transparency; Accountability; Automated Decision-Making; Data Protection

I. INTRODUCTION

Artificial Intelligence (AI) technologies are increasingly integrated into legal and quasi-legal decision-making processes, ranging from legal research and predictive analytics to administrative adjudication, risk assessment, and judicial decision support. While these systems promise efficiency, consistency, and enhanced access to justice, their growing reliance on complex machine-learning models has generated serious concerns regarding transparency, fairness, and accountability. In many cases, AI systems operate as "black boxes," producing outcomes without providing intelligible explanations for how those outcomes were reached. Such opacity presents a fundamental challenge to legal systems that are built upon reasoned decision-making, procedural fairness, and the right to contest adverse outcomes. Transparency has long been a cornerstone of the rule of law. Judicial and administrative decisions are expected to be justified through clear reasoning, enabling affected individuals to understand, evaluate, and challenge the basis of those decisions. Black-box AI systems disrupt this normative framework by obscuring the logic underlying automated or AI-assisted decisions. When AI outputs influence legal rights, obligations, or access to justice, the absence of meaningful explanations undermines due process, weakens accountability mechanisms, and erodes public trust in legal institutions.

In response to these concerns, the concept of Explainable Artificial Intelligence (XAI) has emerged as a critical area of research and policy development. However, much of the existing discourse on explainability remains technically oriented, focusing on model interpretability rather than legal adequacy. From a legal perspective, explainability must serve specific normative functions: enabling procedural fairness, ensuring accountability of decision-makers, facilitating judicial and regulatory oversight, and complying with data protection and privacy obligations. This disconnect between technical explainability and legal transparency highlights the need for a reconceptualization of transparency standards in legal AI systems. Recent regulatory developments underscore the urgency of this shift. Instruments such as the European Union's General Data Protection Regulation (GDPR) and the EU Artificial

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, January 2026 Impact Factor: 7.67

Intelligence Act reflect an emerging consensus that AI systems used in high-risk legal contexts must meet enhanced transparency and accountability requirements. These frameworks recognize that transparency is not a one-size-fits-all obligation but must be calibrated according to the context, stakeholders involved, and the potential impact on fundamental rights. Nevertheless, significant ambiguity remains regarding what constitutes a legally sufficient explanation and how transparency obligations should be operationalized in practice.

This paper advances the concept of "glass-box" transparency as a normative and functional alternative to black-box AI in legal contexts. Unlike purely technical notions of interpretability, glass-box transparency emphasizes legally meaningful explanations that align with principles of due process, fairness, and accountability. The paper argues that transparency in legal AI systems must be purpose-driven, stakeholder-specific, and embedded throughout the AI lifecycle, from design and deployment to oversight and review. The study seeks to address the following core questions: How should transparency standards for legal AI systems be defined to meet legal and regulatory expectations? What role does explainability play in ensuring fairness and accountability in automated decision-making? And how can transparency be balanced with competing interests such as data protection, privacy, and proprietary rights? By engaging with these questions, the paper aims to contribute to the evolving discourse on responsible AI governance and to propose a principled framework for integrating glass-box transparency into legal AI systems.

II. OBJECTIVES

The primary objective of this research is to examine and redefine transparency standards for legal Artificial Intelligence (AI) systems by transitioning from opaque "black-box" models to a legally meaningful "glass-box" framework. In furtherance of this aim, the study seeks to achieve the following specific objectives:

- To analyse the concept of black-box AI systems and assess their implications for fairness, due process, and accountability within legal and quasi-legal decision-making.
- To examine the legal foundations of transparency and explainability, particularly in relation to constitutional principles, administrative law doctrines, and emerging regulatory frameworks governing AI.
- To evaluate existing explainable AI (XAI) approaches and assess their adequacy in meeting legal requirements for transparency, contestability, and reasoned decision-making.
- To study the role of explainability in ensuring accountability of human decision-makers, institutions, and AI developers involved in AI-assisted legal processes.
- To assess the interaction between transparency obligations and data protection norms, including privacy, data minimization, and proprietary interests, in the context of legal AI systems.
- To propose a "glass-box" transparency framework that aligns technical explainability with legal standards of fairness, accountability, and regulatory compliance.
- To contribute to policy and regulatory discourse by offering recommendations for integrating transparency-by-design into the governance of legal AI systems.

III. THE BLACK-BOX PROBLEM IN LEGAL AI SYSTEMS

AI models used in legal or quasi-legal settings (e.g., sentencing aids, bail tools, case prediction) often rely on complex machine learning methods like neural networks whose internal parameters and decision paths are inscrutable to humans. These deep learning models offer high performance but low interpretability, obstructing scrutiny and enabling implicit bias, unfairness, or unverified reasoning.

Key consequences include:

Procedural opacity: Individuals cannot meaningfully contest algorithmic outputs without understanding the logic.

Accountability gaps: Decision-makers may defer to AI recommendations without bearing responsibility.

Trust erosion: Legal legitimacy depends on reasoned justification — missing in opaque systems.

These challenges collectively demand a shift from black-box models toward glass-box design and governance, where meaningful explanations form part of the legal AI lifecycle.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

IV. LEGAL AND REGULATORY FOUNDATIONS FOR EXPLAINABILITY

4.1 GDPR's Right to Explanation

Under the EU's General Data Protection Regulation (GDPR), automated decision-making that produces legal or similarly significant effects triggers specific rights: data subjects must receive "meaningful information about the logic involved," the significance and envisaged consequences, and enjoy the right to human intervention.

While the exact scope and enforceability of this "right to explanation" are debated, GDPR clearly mandates information on automated processing logic tied to legal decisions — moving toward glass-box transparency where individuals can understand and contest outcomes.

4.2 EU Artificial Intelligence Act

The EU AI Act (2024) introduces a risk-based regulatory framework for AI. High-risk systems — including those used in legal proceedings and justice administration — must meet obligations related to transparency, documentation, human oversight, and traceability.

Importantly, the Act embraces context-dependent explainability, recognizing that explanations vary for affected individuals, regulators, and developers.

V. DEFINING "GLASS-BOX" TRANSPARENCY FOR LEGAL AI

Transitioning from black box to glass box involves more than model interpretability — it requires tailored transparency standards that satisfy:

5.1 Stakeholder Requirements

Individuals: Comprehensible explanations enabling contestation.

Judges/Decision-makers: Explanations linked to legal reasoning and evidentiary standards.

Regulators: Detailed documentation for auditing and compliance checks.

This aligns with emerging scholarship that stresses explainability tailored by audience and purpose, not one-size-fits-all disclosures.

5.2. Technical Approaches to Explainability

Explainable AI (XAI) techniques offer a spectrum of approaches:

Intrinsic interpretable models: Rule-based or symbolic methods that are understandable by design.

Post-hoc explanations: Techniques like feature importance (e.g., SHAP, LIME) that elucidate model behavior after training.

Argumentation and hybrid methods: Models that produce legally relevant justifications tied to norms and principles.

Each technique has trade-offs between fidelity, comprehensibility, and legal utility. For glass-box transparency, combining methods — especially aligning explanations with legal concepts — is often necessary.

5.3 Balancing Transparency with Privacy and Proprietary Interests

Explaining AI decisions may conflict with data privacy, especially when explanations require revealing sensitive inputs or private data. Similarly, AI developers often protect algorithms as proprietary intellectual property.

5.4 Balancing these demands calls for tiered disclosure standards:

Regulatory disclosures with full technical details to auditors and authorities.

User-level explanations simplified for individuals.

Security and privacy protections that avoid leaking raw data or proprietary algorithms.

GDPR's data subject rights and AI Act's human oversight obligations both support this calibrated transparency.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

V. IMPLEMENTATION CHALLENGES & FRAMEWORKS

Adopting glass-box standards requires institutional and technical changes:

Compliance-by-Design frameworks embed explainability artifacts (logging, provenance, audit records) throughout AI development and deployment.

Human-in-the-loop governance ensures oversight and intervention remain possible.

AI literacy and professional standards help legal actors interpret explanations correctly.

VI. RESULT & DISCUSSION

6.1 Persistence of Black-Box Practices in Legal AI Systems

The analysis reveals that despite growing regulatory attention, black-box AI systems continue to dominate legal and administrative applications. Many AI tools deployed in areas such as risk assessment, predictive analytics, and administrative decision support rely on opaque machine-learning models that prioritize accuracy and efficiency over transparency. The findings indicate that explainability is often treated as an optional add-on rather than a core design requirement. This persistence reflects a broader technological bias toward performance metrics, often at the expense of legal values such as reasoned decision-making and procedural fairness. From a legal perspective, this opacity creates a structural imbalance: affected individuals are subjected to AI-influenced decisions without access to intelligible reasons, while decision-makers and institutions retain discretion without corresponding accountability. The result is a weakening of established legal safeguards, including the right to be heard and the right to challenge adverse decisions.

6.2. Inadequacy of Existing Explainability Practices for Legal Purposes

The study finds that current explainable AI (XAI) techniques are frequently insufficient to meet legal transparency requirements. Many systems rely on post-hoc explanations that offer simplified or probabilistic insights into AI outputs, such as feature importance scores or statistical correlations. While these techniques may satisfy technical curiosity, they often fail to provide explanations that are legally meaningful or actionable. Legal transparency requires explanations that are understandable to non-technical stakeholders, aligned with legal reasoning, and capable of supporting contestation and review. The results suggest that purely technical explanations do not adequately address questions of legality, proportionality, or fairness. This disconnect reinforces the argument that explainability must be evaluated not only by technical fidelity but also by its ability to serve normative legal functions.

6.3. Explainability as a Mechanism for Fairness and Accountability

The findings demonstrate that meaningful explainability plays a critical role in enhancing fairness and accountability in legal AI systems. Where explanations are available and intelligible, they enable scrutiny of potential bias, discriminatory outcomes, and unjustified deviations from legal standards. Explainability thus functions as a preventive mechanism, discouraging blind reliance on AI outputs and reinforcing human oversight. Moreover, the study highlights that explainability redistributes accountability by clarifying the respective roles of developers, deploying institutions, and human decision-makers. Without transparent explanations, accountability tends to diffuse, allowing institutions to attribute responsibility to technology. Glass-box transparency, by contrast, reinforces the principle that AI systems are tools subject to human responsibility, not autonomous legal actors.

6.4. Regulatory Support for Context-Sensitive Transparency

An important result of the analysis is the recognition that emerging regulatory frameworks support a context-sensitive approach to transparency rather than absolute disclosure. Instruments such as the GDPR and the EU Artificial Intelligence Act reflect an understanding that transparency obligations vary depending on the level of risk, the nature of the decision, and the stakeholders involved.

The discussion indicates that these frameworks implicitly endorse a glass-box model by requiring meaningful explanations, documentation, and human oversight for high-risk legal AI systems. However, the absence of precise standards for what constitutes a “meaningful explanation” remains a challenge. This regulatory ambiguity underscores

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, January 2026 Impact Factor: 7.67

the need for doctrinal clarity and judicial interpretation to translate abstract transparency principles into enforceable legal standards.

6.5. Tension Between Transparency, Privacy, and Proprietary Interests

The results further reveal a significant tension between explainability and competing interests, particularly data protection and intellectual property rights. Providing detailed explanations may risk exposing sensitive personal data or proprietary algorithms. However, the study finds that this tension is not insurmountable. The discussion supports a tiered or graduated transparency model, in which different levels of explanation are provided to different stakeholders. Simplified explanations may be offered to affected individuals, while more detailed disclosures are reserved for regulators and courts under confidentiality safeguards. This approach aligns with data protection principles such as data minimization while preserving the core legal objective of accountability.

6.6. Emergence of the Glass-Box Transparency Model

The cumulative findings support the central argument of the paper: glass-box transparency offers a viable and normatively grounded alternative to black-box legal AI systems. Unlike traditional transparency models that focus solely on model interpretability, glass-box transparency integrates technical, legal, and institutional dimensions. It emphasizes explanations that are purpose-driven, stakeholder-specific, and embedded throughout the AI lifecycle. The discussion highlights that glass-box transparency strengthens legal legitimacy by restoring the link between decision outcomes and justifiable reasons. It also enhances public trust by demonstrating that AI-assisted legal decisions remain subject to human values, legal norms, and institutional accountability.

6.7. Implications for Legal Practice and Policy

The results have significant implications for legal practice, policymaking, and AI governance. Legal institutions must move beyond superficial transparency measures and adopt explainability-by-design approaches. Policymakers should develop clearer standards and guidelines for legally adequate explanations, while courts may play a crucial role in interpreting and enforcing transparency obligations.

Overall, the findings affirm that transparency is not merely a technical challenge but a legal and ethical imperative. Without a transition from black-box to glass-box systems, the use of AI in law risks undermining the very principles it seeks to enhance.

VII. CONCLUSION & FUTURE RESEARCH

The increasing deployment of artificial intelligence in legal and quasi-legal decision-making presents both transformative opportunities and serious normative challenges. This study has demonstrated that the continued reliance on black-box AI systems is fundamentally incompatible with core legal principles such as transparency, procedural fairness, accountability, and the rule of law. When decisions affecting legal rights and obligations are shaped by opaque algorithms, the absence of intelligible reasoning undermines due process, weakens mechanisms of accountability, and erodes public trust in legal institutions. The analysis underscores that explainability in legal AI systems cannot be reduced to a purely technical exercise. Instead, it must be understood as a legal and normative requirement that serves specific functions: enabling affected individuals to understand and challenge decisions, ensuring that human decision-makers remain accountable, and facilitating effective judicial and regulatory oversight. Existing explainability practices, while technically valuable, often fall short of these legal expectations, thereby necessitating a redefinition of transparency standards.

By advancing the concept of glass-box transparency, this paper offers a principled framework that bridges the gap between technical explainability and legal accountability. Glass-box transparency emphasizes context-sensitive, stakeholder-oriented, and purpose-driven explanations embedded throughout the AI lifecycle. Such an approach accommodates legitimate concerns relating to data protection, privacy, and proprietary interests while preserving the core legal demand for reasoned and contestable decision-making.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, January 2026 Impact Factor: 7.67

The study concludes that embedding glass-box transparency within legal AI governance is essential to maintaining the legitimacy of AI-assisted legal processes. As regulatory frameworks evolve and AI technologies become more deeply integrated into legal systems, transparency-by-design must be treated not as an optional safeguard but as a foundational requirement. Future legal, regulatory, and judicial efforts should focus on operationalizing glass-box standards to ensure that AI enhances, rather than undermines, justice, fairness, and the rule of law.

The present study lays a conceptual and doctrinal foundation for redefining transparency standards in legal AI systems through a glass-box framework. However, the rapidly evolving nature of artificial intelligence and legal regulation presents multiple avenues for further research.

First, empirical studies may be undertaken to evaluate how different explainability techniques function in real legal and administrative settings. Future research can assess whether explanations generated by AI systems are genuinely understood by judges, lawyers, administrators, and affected individuals, and whether such explanations meaningfully enhance contestability and fairness in decision-making.

Second, there is significant scope for comparative jurisdictional analysis. While this study primarily engages with European regulatory frameworks, future research can examine transparency obligations under emerging regimes such as India's Digital Personal Data Protection Act, 2023, sector-specific AI guidelines, and judicial interpretations across common law and civil law systems. Such comparative work would enrich the understanding of culturally and institutionally diverse transparency standards.

Third, judicial interpretation and case-law development concerning AI explainability remains nascent. Future research can analyze how courts interpret transparency obligations, admit AI-based evidence, and evaluate algorithmic explanations in adjudication. This would help clarify the role of the judiciary in operationalizing glass-box transparency.

Fourth, further interdisciplinary research is needed to bridge technical explainability and legal reasoning. Collaboration between legal scholars, computer scientists, and ethicists can contribute to the development of AI systems that generate explanations aligned with legal doctrines, evidentiary standards, and normative reasoning rather than purely statistical outputs.

Fifth, future studies may explore the ethical and societal dimensions of transparency, including public trust, legitimacy, and the psychological impact of explanations on affected individuals. Understanding how transparency influences perceptions of fairness and authority is crucial for responsible AI governance.

Finally, there is scope to examine institutional and policy mechanisms for implementing glass-box transparency, such as regulatory sandboxes, audit frameworks, certification standards, and professional guidelines for legal AI deployment. Research in this area can inform policymakers on practical strategies for embedding transparency-by-design into legal AI systems.

In sum, future research should move beyond abstract principles toward empirical validation, doctrinal refinement, and institutional implementation, ensuring that transparency in legal AI systems evolves in step with technological innovation and legal norms.

REFERENCES

[1]. Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76–99.

[2]. Edwards, L., & Veale, M. (2018). Enslaving the algorithm: From a "right to an explanation" to a "right to better decisions"? IEEE Security & Privacy, 16(3), 46–54.

[3]. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

[4]. European Parliament and Council. (2016). General Data Protection Regulation (EU) 2016/679.

[5]. European Parliament and Council. (2024). Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act).

[6]. High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. European Commission.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

[7]. Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.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.

[8]. Dubey D. et al. (2023). A Study Of Recycling And Waste Management Strategies for Mechanical Systems And Products With Legal Aspects, Indian Journal of Science and Research. Vol.3 Issue-3. 78-83

[9]. Pande S. et al. (2023). A Study Of Impact Of Green Infrastructure On Environment, Indian Journal of Science and Research. Vol.3 Issue-2, 70-73