

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