INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]

VOLUME 5 AND ISSUE 5 OF 2025

APIS – 3920 - 0001 *(and)* ISSN - 2583-2344

Published by

Institute of Legal Education

https://iledu.in

THE BLACK BOX OF AI: WHO'S TO BLAME?

AUTHOR - PREKSHA JAYASWAL* & DR. SHEFALI RAIZADA**

* STUDENT AT AMITY UNIVERSITY NOIDA

** DIRECTOR AND JOINT HEAD OF AMITY LAW SCHOOL, NOIDA

BEST CITATION - PREKSHA JAYASWAL**,** THE BLACK BOX OF AI: WHO'S TO BLAME?, INDIAN JOURNAL OF LEGAL REVIEW (IJLR), 5 (5) OF 2025, PG. 901-905, APIS – 3920 – 0001 & ISSN - 2583-2344

Abstract

Artificial Intelligence (AI), quickly embraced, or incorporated, into decision-making processes that have changed industries, also creates significant legal challenges. Most significantly, trust in AI is complicated by the "black box" nature of AI, which leads to a consternation about whether a decision is being made and how decisions are being made when everyone involved may be in the dark. This paper examines the disadvantages of non-transparency or lack of transparence of AI, and the issue of trying to assign fault, when AI systems cause injury (or breach of contract). We will look at traditional and any new theories of liability, and compare the laws concerning AI in the European Union, United States and India. The paper concludes with recommendations on the legal and policy front with the aim of bolstering accountability in the use of AI systems, and reliability of AI systems, recognizing that we are attempting to fix, or fill, critical gaps in existing legal and quasi-legal frameworks., with unique complications, showing that fault cannot easily be assigned febrile, ever-evolving machine learning models. Therefore, we must emphasis on transparency, explainability and ethical safeguards, as we argue for forward looking, legally young infrastructure and policy that reasonably encourages innovation while fostering public trust and responsibility.[1751]

Keywords: Artificial Intelligence (AI), Black Box Problem, Legal Liability, AI Accountability, AI Transparency, Machine Learning, Autonomous Systems, Fault Theories, Negligence and AI, Strict Liability, Vicarious Liability.

[1751] Binns, R. (2018). *On the Importance of Transparency in AI Systems*. Journal of Artificial Intelligence, 1(2), 14-29.

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]

VOLUME 5 AND ISSUE 5 OF 2025

APIS – 3920 - 0001 *(and)* ISSN - 2583-2344

Published by

Institute of Legal Education

https://iledu.in

Introduction

Artificial Intelligence (AI) has infiltrated human activity in countless ways, across a range of subjects and disciplines, from healthcare and finance to legal services and contracts. One area where AI brings a unique legal question is in autonomous decision-making, where AI develops its own decisions without humans fully knowing or explaining those decisions. This raises the "black box" problem which presents an additional challenge to legal liability. If an AI system makes a mistake that results in harm, who is to blame? The developer? The deployer? The user? Or even the AI? This paper examines this issue in relation to what the law will accept.[1752]

As we depend on AI to make complicated tasks without human oversight, there are concerns related to accountability and responsibility across the private and public sectors. Distinct from a tool or software, AI systems—primarily machine learning and neural networks—are informed by data, allowing AI to learn and make decisions unpredictably and from an untraceable source.[1753] The untraceable nature compromises long-standing legal notions, i.e. those that consider recognition of some human fault or intent.[1754]

The understanding of liability involving AI is a developing area on a global scale, compounded by different approaches to liability across jurisdictions. The European Union has embraced a regulatory precaution (with various checks and balances of human oversight) while the United States has rushed out of the gate with (uselessly) lax oversight as a house of cards focus on itself and allowed innovators to operate unchecked and at their own risk. India is working on developing a regulatory framework, thus establishing a fertile breeding ground of recognized regulatory opportunities and risks (that could evolve). This paper seeks to explain how this changing landscape is taking shape internationally. The paper seeks to identify areas of impropriety as we explore doctrinal fault theories and liability frameworks that have emerged as specific to artificially intelligent systems. Thus, the paper will address the need for new definitions, standards of care that permit flexible adaptations in how we conform to standards of care, and implications for a future regulatory framework to address the implications of AI's total autonomy. More fundamentally, the paper asks: how do we balance promoting rate of technological progress of the capabilities of artificially at a similar time as we promote justice in accountability in the emerging age of artificially intelligent systems

Explaining the Black Box Problem

The term "black box" refers to the lack of transparency or interpretability with which some AI algorithms - most notably deep learning, and neural networks - arrive at their decisions. Black box models analyze enormous amounts of information and then identify relationships or patterns among that information that is so abstract, complex, and/or subtle, that human beings are unable to comprehend it.[1755] As a result, providing any explanation for the internal parameters of the model becomes impossible, even for the data scientist who developed the model. Deep learning models work by tuning millions of parameters across multiple layers of interconnected nodes, resulting in a non-linear, non-obvious, and unexplainable decision. While deep learning models are capable of achieving unprecedented levels of accuracy, they do so at the expense of interpretability or transparency.[1756] Unlike rule-based systems where there are rules that lead to specific decisions that can be traced back to a defined logic or rule, the 'reasoning' of a black box model is buried in math abstractions.

[1752] Calo, R. (2015). *Robotics and the Lessons of Cyberlaw*. California Law Review, 103(3), 513-563.

[1753] European Commission. (2021). *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. Brussels: European Commission.

[1754] Guszcza, J., & Dastin, J. (2019). *AI and the Black Box Problem: Challenges in Accountability and Liability*. Harvard Law Review, 132(4), 944-978.

[1755] European Commission. (2021). *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. Brussels: European Commission.

[1756] O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishing Group.

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]

VOLUME 5 AND ISSUE 5 OF 2025

APIS – 3920 - 0001 *(and)* ISSN - 2583-2344

Published by

Institute of Legal Education

https://iledu.in

This lack of transparency raises major legal and ethical concerns, especially when applied in contexts where there are high stakes, sensitive outcomes including medical diagnostics, credit screenings, loan application decisions, predictive policing, contract implementation, and so on. Consider: if an AI denies a loan application or misdiagnoses a patient, how will the loan applicant understand why their application was denied? Or how would a patient understand why they were misdiagnosed.[1757] The individuals involved may not have any recourse or clear rationale for the decision, and it is unclear how one would assign responsibility and appeal the result. Due process, accountability, and fairness lose their meaning in a situation like this.[1758] Furthermore, if an AI algorithm is black box, the legal standards of negligence or strict liability become problematic without similar visibility. Negligent or strict liability standards for human performance, typically involve a requirement where the conduct or violation must be clear, there must be raw causation connecting a person's action (or inaction) to a particular fault. As AI systems assume a larger applied role in an autonomous way, the disconnect between their performance and ability to explain their performance widens significantly. There will certainly be a critical need for a legal and regulatory response to ensure that we do not allow such technology-driven complexity to insulate these actions from accountability.

Legal Theories of Fault and Their Limitations

Traditional legal systems have developed several theories of fault and liability to hold parties responsible when harm occurs. These theories, including negligence, breach of duty, and intentional torts (among others), are built on the standard assumption that human actors are capable of making judgments, can foresee the consequences of their actions, and are culpable for their choice. When applied to AI systems, however, the traditional concepts (principles and theories) of legal fault and liability have serious limitations because AI does not have human attributes, such as consciousness, intent, and the foresight required to evaluate consequences and can be seen as exhibiting human fault or culpability.[1759]

1. Negligence and Breach of Duty

Under traditional tort law, a party is negligent after failing to use the standard of care that a reasonable person would use in a similar situation, and in doing so causes harm. The idea of "breach of duty" is critical to this standard of negligence. The failure of the defendant to act, according to the obligation, whether legally or socially, creates damages to another party.[1760]

This theory is inherently problem some in relation to AI. First, AI systems operate independently and do not have a human-like recognition of what "reasonable care" may look like.[1761] For example, the AI algorithm may make decisions by drawing relationships from large amounts of data, but the AI algorithm does not "know" or "understand" the harm that its decisions may cause. It is human programmers or users that were responsible for the system's initial design or implementation. The AI itself cannot assume the responsibility of foreseeing or controlling its own decision-making process. This disconnect raises the question of which if any liability may apply as it may be that AI behaves in a way that does not conform to human assessments of reasonable behavior, but also reorganizes our idea of who if anyone has breached their duty: the developer, the user, or the AI itself.[1762]

2. Intent and Fault

Intentional torts, in law, always include the human motive to perform an action that will

[1757] Peters, J., et al. (2020). *Artificial Intelligence and Ethics: Challenges for Legal Accountability*. Ethics and Information Technology, 22(1), 81-100.

[1758] Sullivan, R. (2019). *AI Liability and the New Frontier of Law*. Journal of Law, Technology & Policy, 2019(2), 31-59.

[1759] Guszcza, J., & Dastin, J. (2019). *AI and the Black Box Problem: Challenges in Accountability and Liability*. Harvard Law Review, 132(4), 944-978.

[1760] W. Page Keeton et al., *Prosser and Keeton on the Law of Torts* S. 30, at 164–65 (5th ed. 1984)

[1761] Goodman, B., & Flaxman, S. (2017). *EU Regulations on Artificial Intelligence and Data Privacy*. Journal of International Law & Politics, 49(2), 435-471.

[1762] Matthew U. Scherer, Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, 29 *Harv. J.L. & Tech.* 353, 370–74 (2016)

inevitably lead to damages, where the defendant's state of mind (mens rea) is a salient factor. In criminal law, intent delineates departures in levels of culpability even, differentiating manslaughter and murder along with the associated consequences.[1763]

The lack of intent or knowledge in AI systems presents a challenge for the traditional theories of liability. Generally, when an AI system harms someone, it is difficult to identify fault based on an intentional action. This is particularly troubling in such areas as high-risk areas (e.g., self-driving cars), where AI system actions may cause an accident or a death, even if the AI "did not intend" to harm someone.[1764]

3. Strict Liability

Under strict liability, a party may be held liable for certain activities or harms irrespective of fault or intent. In strict liability claims, the category of conduct is the focal point, as opposed to whether the defendant acted negligently or intentionally. For example, in product liability law, the manufacturer may incur strict liability for harm stemming from its products, notwithstanding the manufacturer's lack of negligence in the design or production of its product.

Some academics have suggested that strict liability might apply to AI systems and further suggested that strict liability might specifically apply to AI systems when used in high risk (high risks include autonomous autos, and medical AI) situations. However, there are challenges to applying strict liability to AI systems. The challenges include defining the scope of "high-risk" AI activities and determining what constitutes any "inherently dangerous" activity.[1765]

4. Vicarious Liability

Vicarious liability is a legal theory under which a party will be found responsible for the acts (and omissions) of another party, generally speaking, in employment relationships. In other words, an employer is responsible for the acts of employees that are committed in the course of their employment. [1766]

In the context of AI, vicarious liability has been invoked in relation to situations where AI is acting through a corporate (or an individual) entity. For example, if an AI entity fails to look after a financial portfolio or drives a self-drive vehicle and causes some harm, then the employer or developer may be found to be vicariously liable. However, vicarious liability encounters challenges in the AI context; most notably in so far as autonomous AI's unverified, unmonitored actions might disengage that relationship of responsibility or connection between the action and an employer even, if there seems to an intuitive comprehension that this is not customary to exist outside those structures of governance or supervision. Discerning whether AI's actions can even be said to be within any responsiveness of an entity's responsibility is substantially unclear as an AI interface engages in autonomous navigation.

5. The Legal Vacuum

While some legal scholars have contended that variants of strict liability and vicarious liability could be applied to an AI scenario, in their current straight versions these remains inadequate to fully comprehend the complexities related to AI driven errors. Most importantly, there still remains a division as a direct split or halted chain of action because AI lacks human features, they rank outside this organized system.

Multiple Stakeholders: There are a variety of actors who play a role in AI systems including: developers, companies and end users. This has

[1763] Dan B. Dobbs et al., *The Law of Torts* S. 31, at 84–86 (2d ed. 2011
[1764] Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353, 366–68 (2016)
[1765] Sullivan, R. (2019). *AI Liability and the New Frontier of Law*. Journal of Law, Technology & Policy, 2019(2), 31-59.
[1766] Mulligan, D., et al. (2018). *Artificial Intelligence, Transparency, and Accountability in the Digital Age*. Stanford Technology Law Review, 21(3), 243-275.

resulted in uncertainty about who is accountable when harm occurs.[1767]

Problems with Assigning Liability

There is an intrinsic problem to assigning liability in the context of AI, because of the multiple actors, unpredictable nature of AI systems and opaque decision-making. These essentially blend into one problem, since AI systems make autonomous decisions in unpredictable ways, which makes it difficult for developers, operators and users to anticipate their intended behaviour or even always understand what decisions were made.

1. Multiple Stakeholders: There are multiple actors in ethical AI systems. Developers, companies and users can all play a part in each AI system, creating a situation whereby it is difficult to assess responsibility when things go wrong.[1768]
2. AI is Dynamic: An AI may learn and adapt over time, creating subsequent decisions no one stakeholder foresaw, thereby adding unpredictability.[1769]
3. Black Box Issue: The lack of interpretability and transparency in accountability of AI also means that decision making may, likewise, lack interpretation, complicating an identification of causes of errors.[1770]
4. Causality: due to the complex, interactive nature of AI systems, accompanied by loss of control as these systems learn, guarantee pinpoint progress of harm, or subsequent causes of claims for accountability, is complicated.
5. Legal and Regulatory Void: While most legal systems have regulations governing liability and processes to oversee considering liability, the traditional context often cannot comprehend AI systems, and as such were left with undefined processes, and rules of liability, leaving something akin to liability vacuums.

Conclusion

The black box problem in AI constitutes a fundamental friction with traditional legal systems. As AI continues to expand and develop in significant areas of life, the existing legal framework must change. This paper proposes the following actions:

1. Mandated Explainability: Legislation should require AI developers to design explainability into their systems.[1771]
2. Shared Liability Frameworks: There is a need to create models that share liability among the groups: developers; users; deployers.[1772]
3. Regulatory Authority: There should be independent regulatory bodies to oversee AI applications and ensure compliance.[1773]
4. International Frameworks: A need exists to promote international standards and open, collaborative frameworks to address cross-border liabilities of AI.[1774]

Ultimately, although AI's black box problem complicates the work of attributing fault, the legal system must adapt and implement measures that require accountability; protect rights; and maintain trust in the face of rapidly developing technologies. [1775]

[1767] Zeng, Y., et al. (2020). *A Comparative Analysis of AI Regulation: Europe, the United States, and China*. Journal of International Business and Law, 24(1), 23-48.
[1768] Brent Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, 3 Big Data & Soc'y 1, 6 (2016)
[1769] Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 Geo. L.J. 1147, 1175–76 (2017)
[1770] Jenna Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, 3 Big Data & Soc'y 1, 3–5 (2016)
[1771] European Commission. (2021). *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*. Brussels: European Commission.
[1772] United Nations. (2021). *Draft Guidelines on AI and Liability*. United Nations Office of Legal Affairs.
[1773] Peters, J., et al. (2020). *Artificial Intelligence and Ethics: Challenges for Legal Accountability*. Ethics and Information Technology, 22(1), 81-100.
[1774] Calo, R., & Citron, D. K. (2020). *The Automated Administrative State: A Crisis of Legitimacy*. University of Pennsylvania Law Review, 168(5), 1497-1554.
[1775]Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 Ga. St. U. L. Rev. 1305, 1328–29 (2019).