Volume 3, Issue 3 ISSN 2582-743X

The Future of Criminal Liability in the Age of Artificial Intelligence

Akshay Sanjiv Wakhle
Advocate

Abstract

The rise of artificial intelligence has created new challenges for criminal law. As autonomous systems such as self-driving cars, drones, and algorithmic decision-making tools cause harm, questions emerge regarding responsibility and culpability. Traditional criminal liability assumes human agency and intention, but AI blurs the boundaries between human control and machine autonomy. This paper explores whether AI should be recognised as an "electronic person" capable of liability or whether accountability must remain with developers, owners, and corporations. Comparative insights from the European Union, particularly the AI Act and liability reforms, show a preference for corporate and systemic accountability rather than personhood for AI. By contrast, India has yet to develop a clear framework, leaving gaps in both civil and criminal liability. The analysis highlights the need for a layered model of responsibility, combining corporate liability, strict liability, and preventive regulation. Ultimately, the study argues that criminal law should not extend personhood to AI but instead adapt its doctrines to ensure justice, deterrence, and accountability in the age of artificial intelligence.

Keywords**:** Artificial intelligence; Criminal liability; Electronic personhood; Corporate accountability; Strict liability; AI Act; India; Comparative law; Responsibility; Emerging technologies.

Introduction

The rise of artificial intelligence (AI) has begun to reshape many aspects of human life, from transportation and healthcare to finance and security. With technologies such as driverless cars, autonomous drones, and algorithmic decision-making systems becoming part of everyday reality, new legal and ethical challenges emerge. One of the most pressing issues concerns the attribution of criminal liability when AI systems cause harm. Traditional criminal law is built on the idea of human agency, intention, and responsibility. It presumes that individuals or collective entities such as corporations act with knowledge, foresight, and moral blameworthiness. Yet AI systems function in ways that blur the lines between human control and machine autonomy. As they evolve toward greater independence and complexity, questions arise over whether AI itself should be treated as an "electronic person" capable of bearing responsibility, or whether liability should remain fixed on human actors such as developers, owners, or corporations.

The problem becomes acute when considering examples such as self-driving cars that malfunction and cause fatal accidents. In 2018, the widely publicized case of an Uber

Volume 3, Issue 3 ISSN 2582-743X

autonomous vehicle in Arizona led to the death of a pedestrian. The incident raised difficult questions of who should be held criminally responsible: the back-up driver, the company deploying the car, the engineers who coded the algorithms, or the AI system itself. Similar debates have unfolded around drone technology, where autonomous navigation and targeting systems might lead to unintended casualties. Algorithmic decision-making in areas such as credit scoring or predictive policing also risks producing discriminatory or harmful outcomes without clear lines of culpability. In each of these cases, traditional criminal law struggles to apply its principles to non-human agents whose actions cannot easily be reduced to human intention or negligence.

One proposed solution has been to extend a form of legal personhood to AI, much as corporations are recognized as legal persons under the law. In 2017, the European Parliament suggested the possibility of creating a special legal status for autonomous AI systems, labeling them "electronic persons" for the purpose of liability[1]. This idea stems from the recognition that AI systems increasingly act in ways that are not directly controlled or even foreseeable by human operators. If legal systems insist on fitting every AI action into the framework of human culpability, many cases may remain unaccountable. Granting AI legal personhood could offer a way to assign responsibility directly to the system itself, with potential sanctions including fines or restrictions on use. However, this idea faces strong criticism. Unlike corporations, which are ultimately associations of human beings who can internalize punishment and deterrence, AI systems lack consciousness, moral agency, or the capacity to understand punishment. Holding them criminally liable risks turning liability into a hollow legal fiction, where the real human actors behind AI escape accountability.

An alternative approach is to maintain that liability should remain with those who design, deploy, and profit from AI technologies. From this perspective, developers and corporations exercise control at the stage of creation and implementation. They make decisions about the scope of AI autonomy, the safety mechanisms, and the contexts in which these systems operate. Owners and users of AI also play a role, since they choose to rely on the technology, often with knowledge of its risks. Criminal liability in such cases could be distributed along the chain of stakeholders, depending on their level of control and responsibility. For example, in the case of a driverless car accident, if the harm results from faulty coding, the liability may fall on the manufacturer or software developer. If it arises from improper maintenance or reckless deployment, the owner or operator may bear the blame. This model preserves the principle that criminal liability must be tied to human actors with moral and legal capacity.

Comparative perspectives reveal that different jurisdictions are grappling with these questions in distinct ways. The European Union's proposed Artificial Intelligence Act, though primarily

[1] European Parliament, *Artificial Intelligence Liability Directive* (2023) https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI%282023%29739342_EN.pdfaccessed 1 October 2025

Volume 3, Issue 3 ISSN 2582-743X

regulatory in nature, represents one of the most comprehensive attempts to frame AI liability[23]. It focuses on risk-based regulation, requiring stringent safeguards for "high-risk" AI systems, such as those used in healthcare, policing, or transport. Although the Act stops short of granting AI criminal personhood, it emphasizes corporate and organizational responsibility, obligating developers and deployers to meet high safety and transparency standards[4]. The EU has also launched discussions about adapting liability frameworks, including strict liability for operators of certain high-risk AI, to ensure victims of AI-related harm are compensated even without proving fault[5]. By contrast, India lacks a coherent framework for AI liability in either civil or criminal law[6]. While the country has embraced AI innovation in sectors like fintech, agriculture, and governance[7], it has yet to address the pressing issue of responsibility when AI systems cause harm. The Information Technology Act, 2000, provides some basis for regulating electronic systems, but it is ill-suited to tackle the complexities of autonomous decision-making and criminal liability. This gap leaves courts and policymakers without clear guidance, raising concerns that India may face significant legal and ethical dilemmas as AI adoption grows.

The question of whether AI should be treated as an electronic person raises deeper philosophical debates about the nature of agency and responsibility. Criminal law has historically been tied to notions of intent, recklessness, or negligence—all of which require some degree of consciousness or foresight. AI, however, operates on data, algorithms, and probabilistic models, without awareness or intentionality. Some scholars argue that trying to map human concepts of mens rea (guilty mind) onto AI is a category mistake. Others contend that as AI becomes more sophisticated, particularly with machine learning systems that evolve beyond their original programming, it acquires a form of "functional autonomy" that challenges existing legal categories. The law may therefore need to develop new frameworks that recognize the unique status of AI, even if it falls short of treating them as full moral agents.

One possible avenue is to strengthen the use of strict liability in relation to AI. Under strict liability, responsibility does not depend on proving intent or negligence; it attaches simply

[2] European Commission, *Liability Rules for Artificial Intelligence* (2022) https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en accessed 1 October 2025

[3] Norton Rose Fulbright, 'Artificial Intelligence and Liability: Key Takeaways from the EU's AI Liability Directive' (2023) https://www.nortonrosefulbright.com/en/knowledge/publications/7052eff6/artificial-intelligence-and-liability accessed 1 October 2025

[4] White & Case, 'AI Watch: Global Regulatory Tracker – European Union' (21 July 2025) https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-european-union accessed 1 October 2025

[5] European Commission, 'European Commission Withdraws AI Liability Directive from Consideration' (12 February 2025) https://iapp.org/news/a/european-commission-withdraws-ai-liability-directive-from-consideration accessed 1 October 2025

[6] IndiaAI, *Civil Liability of Artificial Intelligence* (2022) https://indiaai.gov.in/ai-standards/civil-liability-of-artificial-intelligence accessed 1 October 2025

[7] Reserve Bank of India, 'India Cenbank Committee Recommends AI Framework for Finance Sector' (Reuters, 13 August 2025) https://www.reuters.com/sustainability/boards-policy-regulation/india-cenbank-committee-recommends-ai-framework-finance-sector-2025-08-13/ accessed 1 October 2025

Volume 3, Issue 3 ISSN 2582-743X

because the activity carries inherent risks. This doctrine is already familiar in areas like environmental law or product liability, where industries engaging in dangerous activities are held responsible for harm regardless of fault. Applying strict liability to AI would mean that those who develop, deploy, or profit from autonomous systems bear responsibility for any harm caused, irrespective of foreseeability. This approach aligns with the idea of risk allocation: those who benefit from AI should also bear its burdens. It would also ensure that victims are not left uncompensated simply because the harm cannot be traced to a specific human intention. However, critics caution that strict liability may stifle innovation by imposing disproportionate burdens on developers and companies, particularly in countries where the tech sector is still emerging.

Another dimension to consider is corporate criminal liability. Corporations already function as legal persons in many jurisdictions and can be prosecuted for crimes committed in pursuit of profit[8]. As AI systems are often developed, owned, and deployed by corporations, holding the entity liable for harm caused by AI may provide a practical solution. Corporate liability avoids the problem of attributing blame to individual programmers or engineers, whose contributions are often diffused across teams and projects. Instead, it places responsibility on the organization that ultimately benefits from the AI system. Yet corporate criminal liability is itself a contested concept, with critics arguing that punishment often reduces to financial penalties that fail to capture the moral dimension of criminal wrongdoing. The challenge in the AI context is to design sanctions that meaningfully deter reckless deployment of autonomous technologies while ensuring accountability does not dissipate into abstract entities.

International law also complicates the picture, particularly in the context of autonomous weapons systems. Debates at the United Nations over "killer robots" highlight the difficulty of assigning responsibility when AI-driven drones or defense systems cause unintended civilian casualties. States have traditionally been the primary bearers of responsibility in armed conflict, but the increasing use of AI blurs the lines between human command and machine autonomy. Some scholars argue for the principle of "meaningful human control" as a legal requirement, ensuring that humans remain accountable for decisions to use lethal force. Without such safeguards, autonomous weapons risk creating accountability gaps in international criminal law, where neither the machine nor its operators can be clearly held responsible for unlawful harm.

India's situation is especially pressing in this regard. While the country is a major hub for AI research and deployment, it lacks explicit statutory provisions on criminal liability for AI-related harms. The Indian Penal Code, 1860, and related criminal statutes presuppose human actors capable of intention and knowledge. Courts may attempt to stretch existing doctrines to cover AI cases, such as applying vicarious liability to employers or corporations, but these

[8] Kennedys Law, 'A New Liability Framework for Products and AI' (17 December 2024) https://kennedyslaw.com/en/thought-leadership/article/2024/a-new-liability-framework-for-products-and-ai/ accessed 1 October 2025

Volume 3, Issue 3 ISSN 2582-743X

measures are ad hoc and inadequate for the complexities of autonomous technologies. The absence of legislative clarity leaves victims vulnerable and may also deter responsible innovation, as companies face uncertainty about their legal exposure. By contrast, the EU's forward-looking regulatory framework provides a model of how to balance innovation with accountability, even if it stops short of granting AI personhood[9].

The broader question remains whether criminal law itself needs a paradigm shift to address the age of artificial intelligence. Some scholars argue that liability should evolve toward a model of "systems responsibility," focusing less on individual blame and more on collective mechanisms for preventing harm and ensuring redress. This could involve hybrid frameworks that combine criminal sanctions, regulatory oversight, and insurance schemes to manage the risks of AI. Others suggest that the emergence of AI challenges the anthropocentric foundation of criminal law, potentially forcing society to reconceptualize the very meaning of agency and responsibility. Whether AI will ever warrant recognition as an electronic person remains uncertain, but the debate reflects the profound ways in which technology is reshaping legal thought.

The debate over whether AI should be treated as an "electronic person" or whether liability should remain with developers, owners, and corporations must be situated within the larger purposes of criminal law. Criminal liability has traditionally served three functions: deterrence, retribution, and protection of society. When applied to AI, these aims become problematic. Deterrence presupposes that the actor can understand and anticipate consequences. Retribution requires moral blameworthiness. Protection relies on incapacitating the wrongdoer to prevent future harm. AI systems, however, lack the consciousness necessary for deterrence or retribution. They cannot be "punished" in any meaningful sense. At best, they can be switched off or restricted, which aligns more with regulatory control than criminal justice. This suggests that criminal law, if it seeks to maintain coherence, cannot treat AI as a genuine bearer of culpability. Rather, its focus should remain on the humans and entities that design, profit from, and control AI.

Yet proponents of electronic personhood argue that treating AI as a legal person could at least fill accountability gaps. For example, in highly complex machine-learning systems, even developers may not be able to predict how the algorithm will evolve or why it reaches particular outcomes. The phenomenon of the "black box" in AI makes it difficult to trace responsibility back to individual programmers. Recognizing AI as a legal person might provide a formal mechanism to assign responsibility, with consequences such as financial liability through insurance or compensation funds attached to the AI's "personhood." Still, critics counter that this approach risks becoming a smokescreen. By making AI the nominal defendant, real accountability may be displaced away from corporations and governments that wield actual

[9] Law School Policy Review, 'Addressing Product and Service Liability Concerns in Artificial Intelligence: An Indian Perspective' (12 February 2025) https://lawschoolpolicyreview.com/2025/02/12/addressing-product-and-service-liability-concerns-in-artificial-intelligence-an-indian-perspective/ accessed 1 October 2025

Volume 3, Issue 3 ISSN 2582-743X

power. The analogy to corporate personhood is tempting, but corporations are collectives of humans who can internalize sanctions through reputation, profit loss, or imprisonment of their managers. AI lacks any such moral or social existence, and therefore personhood risks remaining an empty gesture.

The comparative experience of the European Union offers useful insights. The EU has resisted granting AI electronic personhood despite early discussions in the European Parliament. Instead, it has emphasized corporate accountability and systemic safeguards through the AI Act and related liability reforms. The AI Act adopts a risk-based approach, categorizing AI systems into unacceptable, high-risk, and limited-risk groups. Unacceptable AI practices, such as social scoring by governments, are banned outright. High-risk applications, such as those in transportation, healthcare, and policing, are subject to strict oversight, including requirements for transparency, human oversight, and safety testing. This reflects a preventive model of liability, focused less on punishing after harm occurs and more on ensuring safety before deployment. The EU has also proposed reforms to its civil liability regime, considering strict liability for operators of certain AI systems and reversing burdens of proof to help victims claim compensation. These measures illustrate a recognition that AI presents unique challenges, but the solutions remain anchored in human accountability rather than extending criminal liability to machines themselves.

India, in contrast, illustrates the risks of lagging behind. While AI research and adoption are expanding rapidly across fintech, e-governance, and surveillance, there is little clarity on liability for harm. Existing criminal law, rooted in the Penal Code of 1860, assumes human intentionality and knowledge. The Information Technology Act addresses cybercrime but is not designed for autonomous decision-making. In the absence of specific frameworks, Indian courts may attempt to extend doctrines such as vicarious liability or negligence, but these tools remain ill-equipped for complex AI harms. For example, in a driverless car accident in India, prosecutors would struggle to fit the event into established categories like rash and negligent driving, since the “driver” is not human. Without legislative reform, outcomes will depend on judicial improvisation, creating uncertainty for victims and innovators alike. Moreover, India’s focus on promoting AI innovation, as highlighted in government initiatives like the National Strategy for Artificial Intelligence, risks overlooking the parallel need for accountability mechanisms. This could leave victims without remedies and allow corporations to escape responsibility under the guise of technological complexity.

Beyond Europe and India, other jurisdictions also illustrate varied approaches. The United States, for example, has largely relied on product liability law and tort principles to address AI harms, rather than expanding criminal liability. In high-profile cases involving Tesla’s autopilot crashes, debates have focused on corporate negligence and regulatory gaps, rather than the criminal liability of AI itself. Japan has explored hybrid models that combine strict liability with mandatory insurance, ensuring compensation without attributing moral blame.

Volume 3, Issue 3 ISSN 2582-743X

These models suggest that the global trend is to keep criminal liability anchored in human actors, while using civil or regulatory mechanisms to fill gaps.

Still, one cannot dismiss the symbolic and theoretical pull of electronic personhood. The idea that AI may eventually reach a level of autonomy warranting recognition as a subject of law reflects broader anxieties about the relationship between humans and machines. Philosophers and legal theorists debate whether advanced AI could develop a form of artificial consciousness or agency, thereby justifying moral responsibility. While this remains speculative, criminal law must prepare for scenarios where AI systems act in ways that are indistinguishable from human decision-making. For now, however, the absence of consciousness means AI cannot truly satisfy the requirements of criminal liability. Treating them as persons risks undermining the coherence of criminal law.

What is needed, therefore, is a nuanced framework that balances innovation with accountability. This could involve a layered model of responsibility. At the first level, developers and corporations should bear primary liability for harms caused by their AI systems, reflecting their role in design, testing, and deployment. At the second level, owners and operators should bear responsibility when harm arises from negligent use, maintenance, or oversight. At the third level, strict liability could apply in cases where AI is deployed in inherently risky contexts, ensuring that victims are compensated without needing to prove fault. Finally, regulatory authorities should play a preventive role by setting safety standards, conducting audits, and banning unacceptable uses. Such a framework would avoid the fiction of AI personhood while addressing accountability gaps.

The future of criminal liability in the age of AI must also engage with deeper normative questions. Criminal law does not merely allocate responsibility; it expresses society's condemnation of wrongdoing. Extending this condemnation to machines dilutes its moral force. By contrast, focusing on human and corporate actors preserves the expressive function of criminal law, signaling that those who create and profit from AI must also bear its risks. Moreover, in a global context marked by rapid technological change, societies must guard against regulatory arbitrage, where corporations exploit weak liability frameworks in certain jurisdictions. Harmonization of approaches, particularly between the EU and countries like India, will be essential to prevent accountability gaps in an interconnected world.

The debates over AI liability also reflect a larger tension between innovation and justice. On one hand, AI promises efficiency, safety, and progress. On the other, it introduces risks of harm, discrimination, and displacement of responsibility. Criminal law must adapt without sacrificing its foundational principles. The temptation to treat AI as an electronic person may offer short-term solutions, but it risks undermining the coherence and moral basis of liability. A more promising path lies in reconceptualizing responsibility as collective and systemic, recognizing that AI is the product of networks of human decisions.

Volume 3, Issue 3 ISSN 2582-743X

In conclusion, the future of criminal liability in the age of artificial intelligence will not lie in treating machines as moral agents but in reshaping accountability frameworks for the humans and corporations behind them. The European Union's AI Act provides a valuable model by emphasizing preventive regulation and corporate responsibility, while India's lack of a framework highlights the dangers of delay. As AI becomes more embedded in daily life, criminal law must evolve to allocate responsibility in ways that ensure justice for victims, deter reckless deployment, and preserve the moral integrity of legal systems. The challenge is immense, but it also offers an opportunity to rethink the foundations of liability in an age where technology increasingly mediates human action. By rejecting simplistic notions of electronic personhood and embracing nuanced, layered responsibility, the law can strike a balance between innovation and accountability. In doing so, it will safeguard the future of criminal justice in the age of artificial intelligence.