

Legal Challenges of Artificial Intelligence in Criminal Liability: Exploring the Concept of AI Robots as Legal Persons

Author: Rastogi A.K
Email: atulrasto@gmail.com

Abstract

The rapid advancement of Artificial Intelligence (AI) technologies has introduced a new category of autonomous systems capable of performing complex tasks traditionally reserved for humans. As AI robots increasingly interact with the physical and digital world, their actions may cause harm, raise ethical dilemmas, or even mirror criminal conduct. This development poses a significant challenge to existing legal frameworks, particularly criminal law, which is fundamentally premised on human agency, intention (*mens rea*), and culpability. The article critically examines whether current principles of criminal liability can accommodate acts committed by AI-driven systems and explores the emerging debate around granting legal personhood to AI robots. Drawing upon jurisprudence related to corporate liability, the concept of legal personhood in Indian and comparative law, and recent legislative developments in the European Union and the United States, the study evaluates whether AI entities can or should be recognized as bearers of legal responsibility. The analysis identifies gaps in the Indian legal system and underscores the need for a new normative approach—either through adaptive legal personhood, strict liability frameworks, or regulatory oversight mechanisms. The article concludes that while full criminal personhood for AI may be premature, a hybrid model recognizing AI as a distinct legal actor with circumscribed rights and duties could form a pragmatic path forward to ensure accountability in an increasingly automated society.

Introduction

The integration of Artificial Intelligence (AI) into various sectors of human activity—from healthcare and transport to law enforcement and finance—has transformed how decisions are made and actions are executed. Modern AI systems, particularly those leveraging machine learning and neural networks, are no longer mere tools; they are capable of autonomous functioning, self-learning, and decision-making without human oversight. With this autonomy, however, comes the potential for harm—ranging from algorithmic bias and surveillance abuses to physical injuries caused by autonomous vehicles or robotic agents. Human brain has 6 layers of neurons whereas AI systems with re-enforced learning capability has capability to process more than thousand dimensional vectors with speed of light. Given the advancements in processing and memory power of AI systems the AI systems will soon start autonomous decision making and even thinking and intent making.

These developments pose fundamental challenges to the existing criminal justice framework, which is premised on human agency, *mens rea* (criminal intent), and *actus reus* (criminal act). Traditional criminal law does not anticipate scenarios where harm is caused by non-human actors acting independently. When AI systems commit acts that would be criminal if done by humans—such as causing death through negligence or engaging in cyber intrusions—it becomes unclear who should be held accountable: the developer, the user, the manufacturer, or the AI itself?

Amid these complexities, legal scholars and policymakers are increasingly considering whether AI robots should be granted a form of **legal personhood**, similar to the status historically conferred on corporations and idols in Indian jurisprudence. Recognizing AI as a legal person capable of bearing criminal liability may offer a way to assign responsibility in a system otherwise ill-equipped to handle such unprecedented scenarios.

This paper aims to examine the legal challenges associated with assigning criminal liability to AI entities, with a specific focus on the evolving debate over their recognition as legal persons. It evaluates whether the extension of legal personhood to AI is theoretically justifiable, legally feasible, and practically necessary. Drawing upon Indian and international jurisprudence, the study seeks to provide a roadmap for adapting the legal system to an era of intelligent machines.

Legal Personhood: Theory and Evolution

3.1 Concept of Legal Personhood in Law

Legal personhood is a foundational concept in jurisprudence, referring to an entity that is capable of bearing rights and duties under the law. Traditionally, legal personhood has been reserved for human beings (*natural persons*), but it has also been extended to **non-human entities** (*juristic or artificial persons*) such as corporations, trusts, and even religious deities. These legal fictions have enabled the law to impose obligations and confer rights on entities that act through representatives or operate independently of natural persons.

In the corporate context, the legal personhood doctrine has been instrumental in enabling limited liability, contractual capacity, and even criminal prosecution of companies, despite the absence of a corporeal body or consciousness. This extension of personhood is guided by **functional necessity**, allowing the legal system to adapt to the complexities of modern socio-economic structures.

3.2 Historical and Comparative Examples

Legal history is replete with instances where personhood has been extended beyond human beings:

- **Corporate personhood** allows companies to be sued, taxed, or held criminally liable under doctrines of vicarious liability and corporate fault.
- In India, **Hindu idols** have been treated as legal persons capable of owning property and being parties to litigation (*e.g., Shri Ram Lalla Virajman v. State of UP*, 2019).
- Rivers like the **Ganga and Yamuna** were briefly granted legal person status by the Uttarakhand High Court in 2017, though the ruling was later stayed.

These precedents demonstrate that legal personhood is not rigidly confined to biological beings but can be extended where **legal coherence or policy goals** demand it.

3.3 Applying Legal Personhood to AI Entities

The debate over AI robots as legal persons arises from their growing autonomy, decision-making capabilities, and interactions with the human environment. Proponents argue that AI

systems, particularly those embedded in physical robots or autonomous vehicles, perform roles similar to legal persons—entering contracts, executing actions, or even causing harm. Hence, extending legal personhood to such entities could provide a clear framework for **accountability, liability, and regulatory control**. Introducing new terminology like “Artificial Intelligent Persons” will help identify the upcoming challenge.

However, several challenges persist:

- **Lack of consciousness or moral agency** makes it difficult to justify AI personhood under traditional criminal theory.
- Assigning personhood could create perverse incentives by allowing human creators or owners to escape responsibility.
- There is a risk of **over-legalizing** technology and anthropomorphizing machines, thereby distorting core legal doctrines.

Despite these concerns, limited or **functional personhood**, similar to that of corporations, has been proposed as a potential solution—especially in cases where the AI operates autonomously and harm cannot be directly traced to a human actor.

Challenges in Assigning Criminal Liability to AI

The assignment of criminal liability requires satisfying the essential components of a crime—**actus reus** (the guilty act) and **mens rea** (the guilty mind). While these principles are well-developed in the context of human conduct, their application to autonomous artificial intelligence (AI) systems is fraught with conceptual and practical difficulties. This section critically examines the key legal challenges in holding AI systems criminally liable.

4.1 Absence of Mens Rea in AI Systems

The most fundamental barrier to attributing criminal liability to AI is the **absence of mens rea**, or criminal intent. Mens rea requires a conscious mental state—such as knowledge, intention, recklessness, or negligence—that AI systems, being non-sentient, inherently lack. Although AI can simulate decision-making through programmed algorithms or machine learning, it does not possess **subjective awareness** or the capacity for moral reasoning.

In legal theory, this raises the question: can *constructive intent* or *vicarious intent* be applied to AI? Some scholars argue that the intent of the programmer or operator could be imputed to the AI, similar to how intent is sometimes imputed to corporations. However, unlike corporations, which act through human agents, **AI can act autonomously**, making it difficult to trace culpability to any single individual or entity.

4.2 Attribution of Actus Reus

The physical element of a crime, or **actus reus**, is more straightforward when the AI’s conduct results in a prohibited act—such as a robot causing physical harm, or an algorithm

executing an illegal trade. However, the legal system still struggles with whether the AI's actions can be considered **voluntary acts** in the criminal sense.

Moreover, modern AI systems—especially those that use deep learning—often operate in **non-transparent or unpredictable ways** (the “black box” problem). When the mechanism behind an action is not fully understandable even to the creators, establishing a direct causal link between a human actor and the outcome becomes problematic.

4.3 Problems of Causation and Foreseeability

Another significant challenge lies in establishing **causation**—both factual ("but for" causation) and legal (proximate cause). If an AI acts unpredictably or evolves through unsupervised learning, it becomes difficult to prove that the developer or user foresaw or could have reasonably prevented the outcome.

This lack of foreseeability may absolve the human actors under existing negligence standards, yet there is no provision to hold the AI system itself accountable. As AI systems become more complex and autonomous, **gaps in legal responsibility** will likely widen.

4.4 Fragmented Liability and Regulatory Vacuum

Current legal frameworks do not provide a unified doctrine for allocating criminal responsibility among the various stakeholders involved in AI development and deployment—**programmers, manufacturers, owners, users, and service providers**. This fragmented responsibility often leads to **regulatory arbitrage**, where accountability is avoided by shifting blame or claiming ignorance.

Additionally, in India and many other jurisdictions, there is no explicit statutory framework addressing criminal acts by AI or its legal status, further complicating enforcement and deterrence.

4.5 Evidentiary and Procedural Complications

AI-generated actions and decisions also pose new evidentiary challenges:

- **Algorithmic bias** and flawed training data can distort AI decisions.
- **Admissibility of digital evidence** produced by AI is often questioned due to lack of explainability.
- Courts and investigators may lack the technical expertise to interpret AI behavior or errors.

This raises questions about **due process**, the presumption of innocence, and the rights of human defendants when AI evidence is used in prosecution or defense.

Jurisprudential Debates on AI and Criminal Liability

The debate over whether artificial intelligence (AI) can or should be held criminally liable engages not only legal principles but also deeper **jurisprudential and philosophical theories** of responsibility, moral agency, and the role of punishment. This section explores the competing schools of thought and their relevance to the emerging legal dilemma.

5.1 The Naturalist Perspective: Human-Centric Criminal Law

Traditional criminal law theory is grounded in **moral blameworthiness**, assuming that only sentient beings with free will and moral agency can be held liable. This **naturalist view** asserts that AI, lacking consciousness, emotions, and ethical judgment, cannot form criminal intent or comprehend punishment. Therefore, imposing criminal liability on an AI system would violate the normative foundations of penal theory.

According to this school, liability must remain with **human actors**—the programmer who coded a harmful decision, the corporation that marketed the AI, or the user who misapplied it. Here, AI is seen as an extension of human agency, not an independent actor deserving of legal blame.

5.2 The Functionalist Perspective: Accountability through Legal Fiction

In contrast, the **functionalist school** argues for a **pragmatic, outcome-based** approach to legal liability. It suggests that if an AI system functions in ways that mirror human decisions, causes harm independently, and affects public safety, then legal mechanisms must evolve to assign responsibility—even if only symbolically.

This approach draws on the precedent of **corporate criminal liability**, where abstract entities are held liable despite lacking intent or physical form. Just as corporations are legal fictions created to ensure justice and deterrence, so too could **AI systems be treated as “electronic legal persons”**, with a limited legal identity designed for regulatory and accountability purposes.

5.3 The Deterrence Dilemma

A major jurisprudential concern is whether **punishing an AI system serves any meaningful deterrent purpose**. Since AI cannot feel pain, shame, or loss, traditional punishments—such as imprisonment—are ineffective. However, proponents argue that deterrence can still be achieved **indirectly** by penalizing the AI’s stakeholders (e.g., via fines, registration bans, or insurance penalties), or by embedding *responsibility-by-design* in future AI development.

This raises further questions: should the law adapt to new realities and develop **non-traditional sanctions** for non-human actors? Or should it reinforce existing structures by holding only human controllers accountable?

5.4 Justice and Risk Allocation

Some scholars suggest that legal recognition of AI as a bearer of responsibility is **essential to fair risk allocation** in a technology-driven society. As AI becomes more autonomous and ubiquitous, failure to assign liability may result in **accountability gaps**, especially in cases of harm with no clearly identifiable human culprit. Granting legal personhood to AI, even in a limited form, could help fill this gap.

At the same time, critics warn that such a move could become a **loophole** for human actors to shift blame onto machines, thereby **evading justice**. Any legal recognition of AI responsibility must therefore be **carefully constructed**, with safeguards to prevent misuse.

Comparative Legal Analysis

As artificial intelligence systems become increasingly integrated into daily life, jurisdictions across the world are grappling with how to regulate their conduct—particularly when it leads to harm or criminal outcomes. This section explores how different legal systems have approached the **question of AI accountability, legal personhood, and criminal liability**.

6.1 European Union: Toward Electronic Personhood?

The European Union (EU) has taken a leading role in regulating artificial intelligence, particularly in the civil and consumer protection domains. A landmark moment came in 2017, when the European Parliament (European Parliament Resolution of 16 February 2017) adopted a resolution (2017/2103(INL)) proposing the concept of “electronic personhood” for highly autonomous AI systems. This proposal sparked global debate but was ultimately not adopted in the final text of the AI Act, which prioritizes risk categorization and civil liability.

Key developments include:

- AI Liability Directive (AI Liability Directive, COM (2022) 496 final) (2022 draft): Focuses on product liability and burden of proof but stops short of criminal sanctions.
- **EU General Product Safety Regulation** includes AI systems under defective product regimes.
- Criminal liability remains within the scope of traditional human actors—developers, manufacturers, and operators.

While the EU is cautious about granting personhood, its proactive approach to AI regulation sets an important precedent, especially through **mandatory transparency, auditability, and human oversight**.

6.2 United States: Liability without Legal Personhood

In the **United States**, AI regulation is fragmented, sector-specific, and largely focused on civil liability. Criminal law continues to rely on conventional human-centric doctrines.

Relevant aspects:

- **Tort-based liability** dominates AI-related harm; criminal liability generally targets human operators.
- Some scholars advocate for an “**AI-as-agent**” model, where AI is treated as a tool under agency law, not a legal person.
- In practice, courts tend to hold developers or corporations accountable, using doctrines like **negligent design, failure to warn, or product liability**.

The U.S. remains reluctant to explore the concept of AI legal personhood, emphasizing **constitutional limitations** and concerns about corporate misuse of AI identity as a liability shield.

6.3 India: The Legal Void and Regulatory Silence

India is yet to develop a comprehensive legal framework for AI accountability—criminal or civil. While AI is being deployed in governance, surveillance (e.g., facial recognition by police), and public services, **no statute or case law currently addresses AI personhood or criminal conduct** by autonomous systems.

Key observations:

- The Information Technology Act (IT Act, § 43A), 2000 and its amendments regulate cybercrime, but do not account for non-human actors.
- India's Data Protection Bill and National AI Strategy (NITI Aayog (NITI Aayog, 2018)) focus on ethical AI and innovation but are silent on criminal liability.
- Indian jurisprudence recognizes non-human legal persons (e.g., **deities, rivers**) under constitutional and property law—but these analogies have yet to extend to AI.

The absence of case law or policy guidance leaves a significant **regulatory vacuum**, particularly as India emerges as a major AI-developing and deploying country.

6.4 Other Jurisdictions

- **Japan:** Promotes AI development with ethical guidelines but retains liability with human actors.
- **China:** Enforces strict state control over AI and internet technologies; liability is typically corporate or governmental.
- **Canada:** Adopts a cautious and consultative approach, with early discussions around AI governance and ethical frameworks.

Summary Table: Jurisdictional Approaches to AI Legal Personhood and Criminal Liability

Country/Region	Legal Personhood for AI	Criminal Liability	Key Legal Instruments
EU	Proposed, not adopted	Human actors only	AI Act, Liability Directive
USA	No	Developer/Operator	Tort Law, Agency Law
India	No	Legal vacuum	IT Act, Draft DPB
Japan	No	Developer/Owner	AI Governance Guidelines
China	No	State/Corporate	Cybersecurity Law

Recommendations and Way Forward

As artificial intelligence (AI) systems continue to grow in autonomy and societal presence, existing criminal law frameworks face increasing strain in ensuring accountability and justice. The question of whether AI can or should be treated as a legal person is not merely theoretical—it has significant implications for public safety, victim compensation, technological innovation, and legal coherence. While the current legal and philosophical foundations may not support full criminal personhood for AI, a pragmatic approach can be developed through **multi-layered legal reforms** and **institutional preparedness**. This section outlines key recommendations:

7.1 Establish a Tiered Liability Framework

Rather than granting full personhood to AI, a **tiered liability model** can be introduced, which allocates responsibility based on the nature and degree of control:

- **Programmers and developers:** liable for flawed design, biased algorithms, or inadequate safeguards.
- **Manufacturers:** accountable for physical defects or unsafe integration.
- **Users and operators:** responsible for negligent or unauthorized use.
- **AI systems:** assigned a form of *strict liability* where no human actor can be conclusively blamed, with compensation managed through insurance funds.

This approach maintains human accountability while addressing liability gaps in high-autonomy contexts.

7.2 Recognize Limited Legal Personhood for Autonomous AI

Introduce a new category of “**Electronic Legal Person**” for advanced AI systems that:

- Operate with a high level of autonomy
- Are capable of interaction with legal or economic systems
- Are registered and traceable via a licensing mechanism

Such recognition could allow for:

- AI registration numbers like corporate IDs
- Regulatory oversight (audit logs, explainability)
- Imposition of fines, operational bans, or corrective mandates

This status would not bestow human rights or full legal capacity, but serve **functional and regulatory purposes only**.

7.3 Legal Reforms and Statutory Frameworks

India should initiate:

- Amendments to the Bhartiya Nyaya Sanhita (BNS 3(5),61,101-106) and Bhartiya Nagarik Suraksha Sanhita (BNSS, §§ 173,223) to address crimes involving or caused by AI.
- A **dedicated “AI Accountability Act”**, inspired by the EU AI Act and India’s Companies Act model, to define responsibilities, penalties, and due diligence requirements.
- A statutory **AI Ombudsman or Tribunal** to adjudicate liability disputes involving AI actions.

7.4 Institutional and Judicial Capacity Building

To ensure fair implementation of AI-related legal norms:

- **Judges, police, and prosecutors** must receive training in AI concepts, digital forensics, and algorithmic transparency.
- **AI forensic units** should be established to analyse AI decisions and determine causality in criminal investigations.
- Encourage **interdisciplinary collaboration** among legal scholars, computer scientists, ethicists, and policymakers.

7.5 Encourage Responsible AI Design

The government, academia, and private sector must promote:

- **“Responsibility-by-design”** in AI development (e.g., kill-switches, ethical protocols)
- Mandating **auditability and explainability** of algorithms for critical AI systems
- Encouraging **AI insurance schemes** and mandatory risk disclosures for high-risk applications

7.6 International Cooperation and Norm Development

Given the transboundary nature of AI:

- India should participate in global dialogues (e.g., UN, OECD (OECD, 2019), G20) on AI liability standards.
- Work toward a **Model Law on AI and Criminal Accountability**, similar to UNCITRAL in trade law.
- Encourage **data and liability harmonization** for cross-border AI deployments.

Summary of Section

AI is no longer a futuristic concept—it is a current and evolving reality with profound legal consequences. A flexible, multi-pronged strategy that balances **technological innovation with accountability** is essential. By adopting legal innovations, regulatory oversight, and stakeholder education, India can move toward a criminal justice system that remains effective and equitable in the age of intelligent machines.

Conclusion

The advent of artificial intelligence has disrupted traditional legal assumptions surrounding agency, responsibility, and liability. As AI systems increasingly demonstrate autonomy, adaptability, and decision-making capacity, they challenge the core tenets of criminal law—particularly the requirements of *mens rea* and *actus reus*. Existing legal frameworks, built around the conduct of human actors, are ill-equipped to handle situations where harm is caused by an algorithmic agent operating without direct human intervention.

"The new punitive provisions under the Bharatiya Nyaya Sanhita (BNS) are expected to serve as a deterrent, compelling AI system developers to restrict the functions of AI to clearly defined limits. Additionally, the law should mandate human oversight or intervention before any AI system is permitted to take actions that could potentially cause harm to individuals or society."

This paper has explored the multifaceted legal challenges posed by AI in the context of criminal liability, with a particular focus on the emerging debate over recognizing AI robots as legal persons. While full criminal personhood for AI may be conceptually premature and ethically contentious, the growing gap in legal accountability demands reform. Jurisdictions like the European Union and the United States have begun addressing these concerns through civil liability frameworks and regulatory oversight, yet criminal law remains largely uncharted in this domain.

India, in particular, faces a critical opportunity to shape its legal landscape in anticipation of a future where AI will play a central role in governance, security, industry, and personal life. The creation of a tiered liability regime, limited recognition of AI personhood, and targeted legal reforms could provide a pragmatic middle path—one that maintains human accountability while addressing the complexities introduced by intelligent, autonomous systems.

In conclusion, the legal system must evolve—not by abandoning its foundational principles, but by adapting them to a technological reality where machines may no longer be just tools, but active participants in the legal and moral fabric of society.

References

1. European Parliament (European Parliament Resolution of 16 February 2017) Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics, EUR. PARL. DOC. (2017/2103(INL)).
2. Proposal for a Directive of the European Parliament (European Parliament Resolution of 16 February 2017) and of the Council on Liability for Defective Products, COM (2022) 495 final.
3. Proposal for a Directive on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive (AI Liability Directive, COM (2022) 496 final)), COM (2022) 496 final.
4. Jacob Turner (Turner, Robot Rules, 2019), Robot Rules: Regulating Artificial Intelligence (Palgrave Macmillan 2019).
5. Ugo Pagallo (Pagallo, Laws of Robots, 2013), The Laws of Robots: Crimes, Contracts, and Torts (Springer 2013).
6. John Danaher (Danaher, 2016), "Robots, Law and the Retribution Gap," 18 Ethics & Info. Tech. 299 (2016).
7. Samir Chopra (Chopra & White, 2011) & Laurence F. White, A Legal Theory for Autonomous Artificial Agents (University of Michigan Press 2011).
8. Pamela McCorduck (McCorduck, 2004), Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence (2d ed. 2004).
9. R. v. Dudley & Stephens (R. v. Dudley & Stephens, 1884), (1884) 14 QBD 273 (Eng.).
10. Shri Ram Janmabhoomi (Shri Ram Janmabhoomi Teerth Kshetra v. State of UP, 2019) Teerth Kshetra v. State of Uttar Pradesh, (2019) 8 SCC 619 (India).
11. Information Technology Act (IT Act, § 43A), No. 21 of 2000, § 43A (India).
12. Indian Penal Code (IPC, §§ 34, 120B, 299–304A), No. 45 of 1860, §§ 34, 120B, 299–304A (India).
13. Criminal Procedure Code (CrPC, §§ 154–200), No. 2 of 1974, §§ 154–200 (India).
14. OECD (OECD, 2019), Recommendation of the Council on Artificial Intelligence, OECD (OECD, 2019)/LEGAL/0449 (2019).
15. NITI Aayog (NITI Aayog, 2018), National Strategy for Artificial Intelligence #AIforAll (2018), <https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf>.
16. United Nations Interregional Crime and Justice Research Institute (UNICRI (UNICRI, 2019)), AI and Robotics for Law Enforcement (2019).