CRIMINAL LIABILITY OF ARTIFICIAL INTELLIGENCE (A.I.) MACHINES UNDER INDIAN LAW: A RETHINK ON MENS REA AND CULPABILITY

Kanchan, Assistant Professor (School of Law), MVN University, Palwal, Haryana

ABSTRACT

From a futuristic concept, artificial intelligence (AI) has developed into a vital component of daily life, impacting industries including healthcare, banking, and law enforcement. As these technologies become increasingly independent, they will be able to make choices and take acts that could have real-world repercussions even in the absence of human supervision. The foundations of criminal law, which have always been based on human behavior and purpose, are seriously threatened by this evolution. The ancient concepts of mens rea (the guilty mentality) and actus reus (the guilty act) were developed with humans in mind rather than the autonomous operation of clever computers. This begs the crucial question of who is criminally liable when an AI system does something that would be illegal if done by a human.

This paper examines how crimes involving autonomous AI systems are covered by Indian legislation, particularly the Bharatiya Nyaya Sanhita, 2023 (which superseded the Indian Penal Code, 1860) and the Information Technology Act, 2000. Using a doctrinal and comparative perspective, it explores liability models, looks at international efforts to define AI accountability, and analyzes how courts interpret intent. According to the study, India's current legal system is ill-equipped to handle crimes perpetrated by or using intelligent computers. In order to distinguish between situations in which AI functions only as a tool, as an independent agent, or in a hybrid role, it is necessary to reconsider current ideas. In the end, it supports an adaptable and progressive legal system that encourages responsibility, equity, and moral leadership in the age of swift digital change.

Keywords: Artificial Intelligence, Criminal law, Actus Reus, Mens Rea, autonomous AI systems.

Introduction

The 21st century has seen artificial intelligence (AI) gain prominence. Once confined to science fiction writers' imaginations, artificial intelligence has now impacted every facet of human existence. It has an impact on how people move through the world, communicate, and make decisions. In a number of industries, including healthcare, finance, education, government, and criminal justice, artificial intelligence (AI) is transforming both institutional practices and human experiences.

In today's digital world, artificial intelligence (AI) systems perform tasks that were previously limited to humans. They predict criminal recidivism more precisely than some traditional policing models, diagnose illnesses faster than doctors, and drive cars without a human driver. AI assesses behavioral risks and recommends interventions in law enforcement through predictive policing tools, predicts investment trends in financial markets, and personalizes retail shopping experiences. This integration of machine intelligence has resulted in unprecedented ease and efficiency, but it has also created new legal, philosophical, and ethical issues.

Who is responsible when an AI system causes harm is a key question at the heart of these dilemmas. If an algorithm unintentionally discriminates against someone based on their gender or color, or if an autonomous vehicle collides with a pedestrian, can the algorithm be held accountable? Or should the end user, manufacturer, or programmer bear the responsibility[1]?.

These issues are not merely theoretical; they pose practical difficulties for academics, judges, and politicians. As AI grows more capable of making decisions on its own, its actions may have repercussions similar to those of human wrongdoing. However, AI systems lack mind, emotions, and intentions in contrast to humans. This leads to a substantial discrepancy between the legal structure intended to control human behavior and the technological reality of machine autonomy.

In India, this issue becomes much more complex. The foundation of the country's criminal justice system is still the Indian Penal Code of 1860, a colonial-era statute that was most

[1] Vaish, S. (2025, August 23). *CAN MACHINES COMMIT CRIMES? Rethinking mens rea in the age of AI.* Record Of Law. Retrieved from https://recordoflaw.in/can-machines-commit-crimes-rethinking-mens-rea-in-the-age-of-ai/

recently revised as the Bharatiya Nyaya Sanhita (BNS), 2023. The new code is still firmly grounded in values developed for human behavior, even though it makes several changes to the rules. This legal system's founders could not have imagined a future where robots could make decisions on their own or where "intention" could be ascertained by algorithms. As a result, when an AI-driven act causes harm, the law has a hard time figuring out who or what is responsible.

Consider the following hypothetical scenario, for instance: A pedestrian is struck and killed by an Indian automaker's self-driving car because its AI system failed to respond to an unexpected situation. With minimal human assistance, the vehicle was operating autonomously. Who is responsible—the AI system, the person who turned it on, the car's manufacturer, or the programmer who created the code? Because non-human agency is not taken into consideration by current legal procedures, such culpability is unclear under current Indian criminal law.

This ambiguity emphasizes how important it is to reevaluate criminal law principles in light of artificial intelligence. AI has challenged the basic assumption of criminal guilt in addition to making it more difficult to discern between human and machine behavior.

Legal Foundations: Understanding Actus Reus and Mens Rea

Actus reus and mens rea are the two fundamental pillars of criminal liability in India and most other legal systems worldwide[2].

- The bodily act or omission that qualifies as a crime is referred to as the actus reus. The exterior element of an offense that causes harm is the "guilty act."

- Mens rea, on the other hand, refers to the "guilty mind" that accompanies the conduct. It represents the knowledge, carelessness, recklessness, or intention underlying an action.

Both of these requirements must be present for someone to be deemed criminally accountable. Both an act of wrongdoing (actus reus) and a guilty mental state (mens rea) are required. This twofold condition guarantees that only morally culpable individuals are punished by criminal law, rather than those whose actions are merely unintentional or beyond of their control.

[2] Ankit Kumar Padhy & Amit Kumar Padhy, Criminal Liability of Artificial Intelligence Entities, 8 Nirma U. L.J. 8 (2019).

But this traditional equation is upset by the rise of artificial intelligence. The physical act (actus reus) is evident when a machine performs an illegal act, such as when an algorithm illegally accesses personal data or when an AI trading bot manipulates stock prices. However, is it possible for a machine that only uses code and computational logic to have mens rea?

AI is incapable of moral thought, emotion, or comprehension. It does not experience fear, guilt, or intent. Rather, it makes probabilistic predictions, learns from data, and executes preprogrammed instructions. Its "decisions" are not the result of moral reasoning, but rather of algorithms. This implies that, unlike humans, an AI cannot be said to have intended to cause harm, even though it may do so.

Some scholars claim that advanced AI systems, particularly those that make use of deep learning and neural networks, display behaviors that are comparable to those of human cognition. These systems can "learn" from experience, "predict" outcomes, and even "choose" between options. Artificial intelligence (AI) imitates the external expression of intelligence but not its underlying consciousness. Its "decisions" are mathematical, not moral.

This presents a significant dilemma for criminal law. Conventional frameworks assume that criminal acts are the result of human choice, which involves awareness and intention. Conversely, AI often operates autonomously and in ways that even its creators cannot fully predict. Therefore, attributing mens rea to AI is wrong from a philosophical and legal standpoint.

This conundrum forces us to examine whether the morality and psychology of our current legal system are sufficient to handle crimes involving intelligent machines. As AI advances, our understanding of responsibility, moral agency, and accountability must also evolve.

AI and Criminal Liability under Indian Law

The criminal justice system in India faces a pressing problem due to the quick integration of AI into many facets of the country's society. India's legal system is still primarily based on human agency, intent, and accountability, despite significant advancements in digital governance and technology usage. When AI systems injure people or carry out actions that would obviously be illegal if done by humans, this disparity becomes apparent.

1. The Colonial Legacy and the Bharatiya Nyaya Sanhita (BNS), 2023

India's criminal code was first established by the Indian Penal Code (IPC), 1860, which was a product of colonial rule. It was structured in accordance with the philosophical and moral understanding of human behavior from the 19th century. The Bharatiya Nyaya Sanhita (BNS), 2023, is a modernization effort that reinterprets several offenses and processes. However, it is still essentially predicated on the same conceptual framework, in which the terms "person," "intent," and "act" all suggest human involvement.

This creates a conceptual void in the context of autonomous systems. Because AI entities lack moral consciousness, are unable to form intentions, and cannot be significantly punished, they are not regarded as legal persons under Indian law.

For instance, Section 100 of the BNS (2023), which addresses "culpable homicide," requires evidence of knowledge or intent to cause death. Since AI lacks consciousness and purpose, the mens rea requirements cannot be met if a self-driving car driven by AI causes a fatal collision. This begs the question of who bears criminal responsibility: the programmer who created the algorithm, the user who activated it, or the manufacturer who implemented it[3].

2. The Doctrine of Vicarious and Corporate Liability

The idea that a company, albeit being a legal entity, acts through human agents and can therefore be held accountable for criminal activities has long been acknowledged in Indian jurisprudence. The vicarious liability doctrine permits institutions or employers to be held accountable for the conduct of their agents. Applying this concept to AI, however, becomes challenging because AI systems frequently behave autonomously, without the direct control or vision of their human designers.

For example, a predictive policing algorithm may identify certain towns as "high-risk" zones based on biased data inputs. The discriminatory effect may result in wrongful arrests or rights violations even though no human actor may have purposefully caused this harm. Conventional vicarious liability systems do not cover this kind of algorithmic, indirect misconduct.

[3] Teena Arora , Dr. Shailja Thakur, Criminal Liability of Artificial Intelligence: A Comprehensive Analysis of Legal Issues and Emerging Challenges , International Journal of Research Publication and Reviews, Vol 5, no 11, pp 1886-1891 November 2024

3. The Role of the Information Technology Act, 2000

The Information Technology Act, 2000 (IT Act), India's primary cyber law, provides procedures for punishing online crimes such as identity theft, hacking, and data exploitation. However, the Act primarily assumes that human agency is responsible for these offenses. It ignores circumstances where an AI system initiates or executes the harmful behavior on its own.

When the harm is brought about by autonomous algorithmic behavior, even the "intermediary liability" safeguards provided by Section 79 of the IT Act—which exempt platforms from liability for user-generated content—do not apply. Therefore, despite providing some regulatory control, the IT Act falls short in addressing AI autonomy and criminal liability.

4. The Role of the Consumer Protection Act, 2019

The concept of product or manufacturer liability has begun to take shape within the legal framework when it comes to AI-driven devices that operate with some assistance from humans. The need to place blame when an AI system harms people has led to this development. Since there is no established legal theory that explicitly addresses liability for autonomous technical defaults, the burden of compensating victims must fall on someone, typically the manufacturer, owner, or operator involved in the machine's operation.

This reasoning is reflected in Sections 84, 85, and 86 of the Consumer Protection Act, 2019, which hold manufacturers and owners liable for negligence when an AI-enabled product malfunctions or causes harm, especially when the system requires human supervision or input. These laws ensure that victims have recourse in the event that technology malfunctions while being operated with human assistance.

However, depending on the particular facts and circumstances of each case, responsibility may be divided or assigned between the maker and the operator in situations where human involvement is minimal or complementary. This strategy aims to achieve a balance between promoting technological innovation and guaranteeing accountability.

The discussion that follows will address a more complex issue: criminal liability for fully autonomous AI systems that function without human intervention. Important questions arise in

these circumstances: Who should be responsible when a machine makes a mistake? And how should punishment be administered when there is no direct human actor at fault?

Global Framework

The issue of AI accountability has been addressed by a number of authorities, but no conclusive solution has been achieved[4].

European Union

The European Union[5] has introduced the **AI Liability Directive** and proposed regulations under the **EU Artificial Intelligence Act (2024)[6]**, focusing on transparency and accountability. Even while these are primarily civil in nature, their requirement for traceability and human monitoring opens the door for criminal responsibility in the future.

United States

AI is not acknowledged as a legal person in the United States. Courts rely on well-established theories like product liability and corporation liability. Nonetheless, theories of "electronic personhood" and "algorithmic agency" have been put forth by American academics, implying that highly autonomous computers may eventually have limited legal responsibility.

United Kingdom

The concept of corporate criminal culpability, which links purpose to a corporation's "directing mind," is applied in the United Kingdom. This idea might be expanded to include AI systems that are managed by organizations.

Japan and South Korea

Frameworks for giving sophisticated AI systems limited legal personality have been discussed

[4] MS. SETIKA PRIYAM & DR. KUNVAR DUSHYANT SINGH, ARTIFICIAL INTELLIGENCE, LEGAL PERSONHOOD, AND DETERMINATION OF CRIMINAL LIABILITY, INDIAN JOURNAL OF LEGAL REVIEW (IJLR), 5 (7) OF 2025, PG. 212-222, APIS – 3920 – 0001 & ISSN - 2583-2344

[5] MS. SETIKA PRIYAM & DR. KUNVAR DUSHYANT SINGH, ARTIFICIAL INTELLIGENCE, LEGAL PERSONHOOD, AND DETERMINATION OF CRIMINAL LIABILITY, INDIAN JOURNAL OF LEGAL REVIEW (IJLR), 5 (7) OF 2025, PG. 212-222, APIS – 3920 – 0001 & ISSN - 2583-2344

[6] REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024, Official Journal of the European Union EN L series 12.7.2024

in both nations, especially with regard to economic and civil culpability.

The discourse around the globe demonstrates a balance between promoting technology innovation and guaranteeing legal accountability. Although complete criminal personhood for AI is still debatable, many countries concur that conventional legal structures need to change to address the difficulties posed by autonomous systems.

How Can A.I. Be Punished for Crimes Committed?

One crucial concern that emerges as artificial intelligence (AI) grows more sophisticated and autonomous is what would happen if an AI system committed a crime. Is it possible for a machine to face criminal penalties? With AI being used more and more in industries like banking, healthcare, transportation, and even law enforcement, this question is no longer science fiction.

Theoretical Models of AI Liability

Legal theorists have put forth a number of methods for determining who is responsible for crimes involving artificial intelligence[7]:

1. **The Tool Model[8]:**

   AI is viewed as a tool that humans use. The human controller is always in charge, just like with a weapon or vehicle.

   Limitation: This fails when AI exhibits unpredictable and uncontrollable behaviour.

2. **The Agent Model**:

   AI is viewed as an independent being capable of making decisions for itself. It is possible to hold the AI system directly accountable.

   Limitation: Since AI lacks consciousness, punishment and deterrence are pointless.

[7] https://www.legalserviceindia.com/legal/article-18098-artificial-intelligence-and-crime-charting-the-future-of-criminal-accountability.html

[8] Vaish, S. (2025, August 23). *CAN MACHINES COMMIT CRIMES? Rethinking mens rea in the age of AI.* Record Of Law. Retrieved from https://recordoflaw.in/can-machines-commit-crimes-rethinking-mens-rea-in-the-age-of-ai/

3. **The Hybrid Model:**

   This method combines both points of view, dividing shared accountability between humans and AI based on the degree of control and autonomy.

   Benefit: It acknowledges AI's increasing independence while ensuring human accountability.

   Most academicians believe that the hybrid model is the most balanced approach because it avoids both complete human exoneration and extreme automation of guilt.

When AI only serves as a tool under human control, the person using or programming it bears responsibility. For example, if an AI is designed to alter hospital data or hack a bank, the human operator is clearly at fault. AI is frequently just acting in accordance with instructions rather than on its own initiative, and such offenses are easily handled by the laws in place.

The situation gets considerably more complicated when an AI system acts on its own initiative and causes harm that neither the user nor the developer had planned or anticipated.. Legal scholar **Gabriel Hallevy** [9] calls this as the ***"natural and probable consequence"*** [10] model. Two potential outcomes are suggested by this model:

1. Because the human user or developer was careless or irresponsible, the AI caused harm.

2. The AI committed a different or additional crime than the human had originally planned or expected.

The person in charge of the AI could be held accountable for negligence in the first case and culpability similar to that of an accomplice or abettor in the second, depending on how much control they had over the AI.

The real challenge is figuring out whether AI can be held legally responsible. Since AI lacks

[9] Gabriel Hallevy, (2010) The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control,4 AKRON I.P.J 171 (2016)
https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2

[10] MS. ANANYA MISHRA, "Can Artificial Intelligence Be Punished for Committing Offences? A Critical Analysis of The Applicability of Criminal Law Principles on Artificial Intelligence", Vol.3 & Issue 3, Law Audience Journal (e-ISSN: 2581-6705), Pages 184 to 200 (25th January 2022),
available at https://www.lawaudience.com/can-artificial-intelligence-be-punished-for-committing-offencesa-critical-analysis-of-the-applicability-of-criminal-law-principles-on-artificial-intelligence/.

emotions, consciousness, and moral knowledge, it is unable to produce criminal intent or mens rea in the same sense as humans. However, if the AI is designed to understand that certain actions are prohibited and it intentionally engages in them, there might be grounds for shared accountability among the owner, developer, and the system itself.

A more difficult situation arises when AI commits what are referred to as "irreducible offences"—crimes that cannot be directly connected to human intent or behavior. These occur when AI behaves erratically or learns new behaviors. Criminal laws do not currently address such autonomous behavior because they were developed with human responsibility in mind.

To address these emerging concerns, experts have suggested that special regulations be created for crimes involving AI. These rules may outline the circumstances in which AI, its developer, or its user should be held accountable.

While fully autonomous "hard AI crimes" have not yet occurred, this is the perfect time for politicians to prepare. Instead of punishing machines like humans, the goal is to ensure that accountability and responsibility are present when AI systems cause harm.

By changing the law and implementing rules specific to artificial intelligence, India can continue to foster innovation and technological progress while guaranteeing victims' justice. Ultimately, the legal system must strike a balance between promoting justice and accountability and halting progress in the age of sentient robots.

The Requirement of a Legal Framework

The fundamental principles of criminal law must be re-examined in light of the development of AI. The traditional duality of actus reus and mens rea is no longer sufficient in a world where complex tasks are performed by machines without direct human supervision. Therefore, developing "AI-specific accountability standards" that can assess accountability in intricate technological ecosystems must be the aim of legal reform.

Possible solutions might include:

- Giving human developers, programmers, and implementers of AI systems

responsibility based on their level of control and harm predictability[11].

- AI systems are sometimes recognized as "electronic persons," which grants them procedural standing but not full personhood.

- In the event of harm, algorithmic transparency is necessary for auditing and tracking down decision-making procedures.

- Establishing a system of no-fault compensation for people harmed by AI, such as those involved in car accidents or medical malpractice.

Ultimately, rather than rejecting AI's autonomy, the law needs to evolve by understanding it within a framework of shared responsibility. This means finding a balance between innovation and accountability, ensuring that India upholds justice, equity, and human dignity while simultaneously promoting AI-driven progress.

Conclusion

Artificial intelligence challenges one of the fundamental principles of criminal law, which is the inextricable link between guilty behavior and a conscious mind. As AI systems become more capable of acting independently and rendering decisions without direct human input, pressure is mounting on the traditional frameworks of mens rea and actus reus.

India's legal system, which still relies on assumptions about human agency from the colonial era, needs to adapt to the ethical and practical challenges of the digital age. Rather than viewing machines as moral beings, acknowledging the unique characteristics of AI means ensuring that human accountability is applied to their creation, use, and oversight.

The goal should be to create a flexible, transparent, and ethically sound legal system that encourages innovation while maintaining justice and equity. The future of criminal law will depend on how we redefine responsibility as well as how we rethink humankind's role in a world increasingly populated by intelligent robots.

[11] Vaish, S. (2025, August 23). *CAN MACHINES COMMIT CRIMES? Rethinking mens rea in the age of AI.* Record Of Law. Retrieved from https://recordoflaw.in/can-machines-commit-crimes-rethinking-mens-rea-in-the-age-of-ai/

References

- Vaish, S. (2025, August 23). CAN MACHINES COMMIT CRIMES? Rethinking mens rea in the age of AI. Record Of Law. Retrieved from https://recordoflaw.in/can-machines-commit-crimes-rethinking-mens-rea-in-the-age-of-ai/

- Ankit Kumar Padhy & Amit Kumar Padhy, Criminal Liability of Artificial Intelligence Entities, 8 Nirma U. L.J. 8 (2019).

- Teena Arora, Dr. Shailja Thakur, Criminal Liability of Artificial Intelligence: A Comprehensive Analysis of Legal Issues and Emerging Challenges, International Journal of Research Publication and Reviews, Vol 5, no 11, pp 1886-1891 November 2024.

- Gyandeep Chaudhary, ARTIFICIAL INTELLIGENCE: THE LIABILITY PARADOX, ILI Law Review Summer Issue 2020.

- MS. SETIKA PRIYAM & DR. KUNVAR DUSHYANT SINGH, ARTIFICIAL INTELLIGENCE, LEGAL PERSONHOOD, AND DETERMINATION OF CRIMINAL LIABILITY, INDIAN JOURNAL OF LEGAL REVIEW (IJLR), 5 (7) OF 2025, PG. 212-222, APIS – 3920 – 0001 & ISSN - 2583-2344.

- REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024, Official Journal of the European Union EN L series 12.7.2024, https://www.legalserviceindia.com/legal/article-18098-artificial-intelligence-and-crime-charting-the-future-of-criminal-accountability.html

- Gabriel Hallevy, (2010) The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control,4 AKRON I.P.J 171 (2016) https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2

- MS. ANANYA MISHRA, “Can Artificial Intelligence Be Punished for Committing Offences? A Critical Analysis of The Applicability of Criminal Law Principles on Artificial Intelligence”, Vol. 3 & Issue 3, Law Audience Journal (e-ISSN: 2581-6705), Pages 184 to 200 (25th January 2022), Available at https://www.lawaudience.com/can-artificial-intelligence-be-punished-for-committing-offencesa-critical-analysis-of-the-applicability-of-criminal-law-principles-on-artificial-intelligence/.