

MENS REA AND ARTIFICIAL INTELLIGENCE AND CRIMINAL LIABILITY IN AI-DRIVEN OFFENCES

AUTHOR – ANAMIKA PATEL, STUDENT AT AMITY UNIVERSITY NOIDA, UTTAR PRADESH

BEST CITATION – ANAMIKA PATEL, MENS REA AND ARTIFICIAL INTELLIGENCE AND CRIMINAL LIABILITY IN AI-DRIVEN OFFENCES, *INDIAN JOURNAL OF LEGAL REVIEW (IJLR)*, 6 (5) OF 2026, PG. 771-7779, APIS – 3920 – 0001 & ISSN – 2583-2344.

ABSTRACT

The rapid rise of Artificial Intelligence (AI) is reshaping how decisions are made across governance, business, healthcare, and daily life, bringing with it new challenges for Indian criminal law. Traditional legal frameworks, such as the Bharatiya Nyaya Sanhita, are built on the assumption of human action, intention, and *mens rea*—the “guilty mind”—to determine criminal responsibility. But AI systems act autonomously and sometimes unpredictably, making it difficult to pinpoint responsibility or prove intent in the conventional sense.

This paper examines how AI affects liability and accountability, highlighting the spread of responsibility across developers, operators, corporations, and end-users. It explores why traditional concepts like *mens rea* may not apply to non-human actors and suggests adopting hybrid legal models that blend fault-based liability with objective standards, including negligence, recklessness, foreseeability, and strict liability, particularly in high-risk areas. The study also addresses the challenges posed by opaque AI decision-making, emphasizing the need for transparency, explainability, and robust evidence under laws like the Bharatiya Sakshya Adhinyam.

Drawing on judicial precedents and international comparisons, the paper proposes reforms that strike a balance between encouraging innovation and ensuring justice. By updating criminal law to reflect the realities of AI, India can create a legal framework that protects public safety, holds the right parties accountable, and stays relevant in a rapidly evolving technological landscape.

KEY WORDS

Artificial Intelligence, Liability, Accountability, Mens Rea, Strict Liability

Mens Rea and Artificial Intelligence and Criminal Liability in AI-Driven Offences.

3.1 Traditional understanding of mens rea under Indian criminal law

The traditional understanding of *mens rea*, commonly described as the “guilty mind,” forms one of the most essential foundations of Indian criminal law. It reflects the long-standing belief that a person should not be held criminally responsible merely for committing a wrongful act unless it is accompanied by a blameworthy mental state. This idea is captured in the Latin

maxim actus non facit reum nisi mens sit rea, which means that an act alone does not make a person guilty unless the mind is also guilty. This principle ensures fairness in criminal law by distinguishing between intentional wrongdoing and accidental harm.

Under Indian criminal law, particularly as codified in the Indian Penal Code, 1860 (IPC), *mens rea* is not defined in a single, uniform manner. Instead, it is expressed through different mental elements embedded within various offences. These include intention,

knowledge, recklessness, and negligence, each representing a different level of culpability.⁹⁸¹ The IPC adopts a practical and flexible approach, where the required mental element is specified within each offence rather than through a general definition.

Intention is considered the highest form of *mens rea*. It refers to a deliberate and conscious objective to bring about a particular consequence. In serious offences such as murder under Section 300 of the IPC, intention is a crucial element.⁹⁸² Courts usually infer intention from surrounding circumstances, such as the nature of the act, the weapon used, and the behavior of the accused. This ensures that liability is based not just on the act itself but on the purpose behind it.

Knowledge represents a slightly lower degree of culpability. It arises when a person is aware that their actions are likely to cause a particular result, even if that result is not their primary aim. For instance, if a person performs a dangerous act knowing it could cause harm, they

may still be held liable. The distinction between intention and knowledge is often subtle but plays a significant role in determining the seriousness of the offence.

Recklessness involves conscious risk-taking. It occurs when an individual is aware of a substantial risk but chooses to ignore it and proceed with their actions. Although the IPC does not always explicitly use the term “recklessness,” the idea is recognized through judicial interpretation.⁹⁸³ It occupies a middle ground between intention and negligence, reflecting a higher degree of fault than mere carelessness.

Negligence, on the other hand, represents the lowest level of *mens rea*. It refers to a failure to exercise reasonable care expected from a prudent person. Unlike intention or knowledge, negligence does not involve awareness of harm but rather a lack of caution. An example can be

seen in offences like causing death by negligence under Section 304A of the IPC, where liability arises from careless or rash conduct rather than deliberate intent.

The importance of *mens rea* has been consistently emphasized by Indian courts. In *Nathulal v. State of Madhya Pradesh*, the Supreme Court held that the absence of a guilty mind could absolve an accused of liability.⁹⁸⁴ Similarly, in *State of Maharashtra v. Mohd. Yakub*, the Court examined how intention and knowledge influence criminal responsibility.⁹⁸⁵ These cases highlight the judiciary's commitment to ensuring that punishment is based on moral blameworthiness rather than mere occurrence of harm.

However, the requirement of *mens rea* is not absolute. Indian law recognizes certain **strict liability offences**, where liability may be imposed even without proof of a guilty mind. Such offences are usually found in areas concerning public health, safety, and welfare. The reasoning is that in certain situations, protecting society is more important than proving intent.

Another important aspect of *mens rea* is its connection with **general exceptions** under the IPC. Chapter IV of the Indian Penal Code, 1860 provides defences such as mistake of fact, accident, insanity, necessity, and private defence. These defences operate by showing the absence of a guilty mind. For example, a person acting under a genuine mistake of fact may not possess the required *mens rea* and therefore cannot be held criminally liable.

The doctrine of *mens rea* also plays a vital role in determining the seriousness of offences and the appropriate punishment. Crimes committed with intention are punished more severely than those resulting from negligence. This ensures proportionality and reflects the moral gravity of the offender's conduct.

⁹⁸¹ K.D. Gaur, *Textbook on Indian Penal Code* (6th ed.).

⁹⁸² Indian Penal Code, 1860.

⁹⁸³ Ratanlal & Dhirajlal, *The Indian Penal Code*.

⁹⁸⁴ *Nathulal v. State of Madhya Pradesh*.

⁹⁸⁵ *State of Maharashtra v. Mohd. Yakub*.

In contemporary times, the application of *mens rea* has become more complex due to technological developments and emerging forms of crime. Nevertheless, the traditional principles continue to guide courts in assessing criminal liability. Even in modern contexts, judges seek to identify the mental element behind an act before imposing punishment.

In conclusion, the traditional understanding of *mens rea* in Indian criminal law is deeply rooted in the idea of moral blameworthiness. By requiring a guilty mind alongside a wrongful act, the law ensures that individuals are punished only when they are truly at fault. Although exceptions like strict liability exist, *mens rea* remains a cornerstone of criminal jurisprudence, upholding fairness, justice, and the fundamental principle that liability must be based on both conduct and intention.

3.1 Challenges in Attributing Intention, Knowledge, and Negligence to AI

The rapid integration of Artificial Intelligence (AI) into decision-making processes has created significant challenges for traditional criminal law, especially when it comes to attributing mental elements such as intention, knowledge, and negligence. These elements form the backbone of criminal liability under statutes like the Indian Penal Code, 1860, where guilt depends not only on the wrongful act (*actus reus*) but also on the mental state (*mens rea*). However, AI systems function in ways that challenge these deeply human concepts, raising complex questions about how responsibility and accountability should be understood in the age of intelligent machines.

One of the central difficulties lies in attributing **intention** to AI systems. In traditional criminal law, intention refers to a conscious objective or purpose to achieve a particular result. It assumes awareness, choice, and control—qualities that are inherently human. AI systems, however, do not possess consciousness or subjective awareness. They operate through algorithms, data inputs, and statistical models. Even when AI appears to “decide” or “choose,” it

is simply executing patterns learned from data. As a result, it becomes conceptually difficult to say that an AI system intended a harmful outcome. For example, if an autonomous vehicle causes an accident, the harm cannot be said to arise from intention in the same sense as a human driver’s deliberate act. Instead, it is the product of programmed responses and environmental inputs rather than conscious choice.

Closely related is the challenge of attributing **knowledge**. In criminal law, knowledge implies an awareness that certain consequences are likely to result from one’s actions. However, AI systems do not “know” in the human sense. While they can process vast amounts of data and make highly accurate predictions, this does not amount to true understanding. Human knowledge involves comprehension, awareness, and the ability to anticipate consequences meaningfully. AI, on the other hand, operates through correlations and probabilities. For instance, an AI system used in medical diagnosis may predict the likelihood of a disease, but it does not understand the implications of its prediction or the harm that may follow from an incorrect result. This raises an important question: can predictive capability be equated with legal knowledge, or must knowledge remain tied to human cognition?

The issue becomes even more complex in the context of **negligence**, which is based on a failure to exercise reasonable care. Traditionally, negligence is assessed using the standard of a “reasonable person,” a legal construct that assumes human judgment and behavior. Applying this standard to AI is inherently problematic because AI systems do not possess judgment in the human sense. They function based on their design, training data, and algorithms. If an AI system makes an error due to flawed programming or biased data, it is unclear whether the negligence lies with the system itself or with the humans involved in its creation and deployment. For example, if a facial recognition system produces

discriminatory outcomes due to biased data, responsibility may lie with the developers or users rather than the AI itself.

Another major challenge is the issue of **autonomy and unpredictability**. Modern AI systems, especially those based on machine learning and deep learning, can evolve over time by learning from new data. This creates a level of unpredictability, as the system's behavior may change beyond its original programming. Such systems are often described as "black boxes," where even the developers cannot fully explain how a particular decision was reached. In criminal law, establishing *mens rea* requires clear evidence of a mental state at the time of the act.

However, when the decision-making process is opaque, attributing intention or knowledge becomes extremely difficult.

The concept of **distributed responsibility** further complicates matters. AI systems are rarely the product of a single individual; they are typically developed and deployed through the combined efforts of programmers, engineers, companies, and end-users. When harm occurs, it becomes challenging to determine who should be held accountable. Should liability fall on the developer who designed the system, the company that implemented it, or the user who relied on it? This fragmentation creates what scholars describe as an "accountability gap," where no single actor can clearly be said to possess the required *mens rea*.

Additionally, the **absence of moral agency** in AI presents a fundamental obstacle. Criminal law is not only concerned with causation but also with moral blameworthiness. Concepts like intention and knowledge are closely tied to the idea that individuals can understand the moral consequences of their actions and make choices accordingly. AI systems lack this moral capacity. They do not experience guilt, remorse, or ethical conflict. As a result, attributing criminal responsibility to AI challenges the very foundations of criminal law, which are built on

human agency and moral accountability.

From a legal standpoint, these challenges have led to different responses. Some scholars advocate retaining the traditional framework by holding human actors—such as developers, manufacturers, or users—liable for AI-related harm. Others suggest adapting existing doctrines, such as negligence or strict liability, to address the risks posed by AI without relying heavily on *mens rea*. There are also proposals to create new legal categories tailored specifically to autonomous systems.

In the Indian context, existing legal frameworks such as the Information Technology Act, 2000 and the Bharatiya Nyaya Sanhita, 2023⁹⁸⁶ do not yet provide clear guidance on how to attribute mental elements to AI. The law continues to rely on human-centered concepts of responsibility, which may not fully address the complexities introduced by advanced technologies.

In conclusion, attributing intention, knowledge, and negligence to AI systems presents deep conceptual and practical challenges. These arise from the absence of consciousness and moral agency in AI, as well as from the complexity and opacity of modern algorithms. As AI continues to evolve, it becomes increasingly necessary to rethink traditional legal doctrines and develop new frameworks that can ensure accountability while adapting to the unique nature of artificial intelligence.

3.2 Algorithmic Decision-Making and Absence of Human Intent

The rise of Artificial Intelligence (AI) and algorithmic systems has fundamentally transformed how decisions are made in contemporary society. Today, from financial markets and healthcare to criminal justice and governance, algorithmic decision-making plays a crucial role in shaping outcomes that directly affect human lives. Yet, this transformation poses a serious challenge to traditional legal

⁹⁸⁶ the Information Technology Act, 2000 and the Bharatiya Nyaya Sanhita, 2023

frameworks—especially the concept of intent, which lies at the heart of criminal liability. Under laws such as the Indian Penal Code, 1860, liability is generally based on the coexistence of a wrongful act (*actus reus*) and a guilty mind (*mens rea*). However, algorithmic systems complicate this framework by producing outcomes that may cause harm without any identifiable human intent behind them.

Algorithmic decision-making refers to the process by which computer systems, often powered by machine learning, analyze large volumes of data and generate decisions or predictions with little or no human intervention. Unlike traditional software that operates on fixed instructions, modern AI systems learn patterns from data and adapt over time. This ability to evolve enables them to function with a degree of autonomy, making decisions in ways that may not always be fully anticipated—even by their creators.

At the core of this issue is the **absence of human intent** in algorithmic actions. In legal terms, intent refers to a conscious objective or purpose to bring about a specific result. It is closely connected to human cognition, involving awareness, deliberation, and choice. Algorithmic systems, however, lack these qualities. They do not possess consciousness or subjective understanding; instead, they operate through statistical correlations and predefined objectives. For instance, an AI-based loan approval system may deny credit to certain individuals based on patterns found in historical data. While the result may appear discriminatory, the system itself does not “intend” to discriminate—it simply processes data in accordance with its training.

This absence of intent creates a significant gap in legal reasoning. Traditional criminal law assumes that harmful outcomes stem from human decisions, making it possible to assign liability based on intention or knowledge. However, when decisions are made by algorithms, identifying whose intent should be considered becomes difficult. The programmer

may not have foreseen the specific outcome, the user may not fully understand the system’s functioning, and the algorithm itself cannot form intent. This leads to what scholars describe as an “intent gap,” where harm occurs without a clearly attributable mental state.

Another important challenge is the **opacity of algorithmic systems**, often referred to as the “black box” problem. Many modern AI models, particularly deep learning systems, operate in ways that are not easily interpretable. Even experts may struggle to explain how a system arrived at a particular decision. This lack of transparency makes it difficult to establish a clear causal link between human input and algorithmic output. In criminal law, proving *mens rea* requires insight into the mental state behind an act. When the decision-making process itself is unclear, attributing intention or knowledge becomes even more problematic.

The issue of **delegation of decision-making** further complicates matters. In many contexts, humans intentionally rely on AI systems to make decisions for the sake of efficiency and accuracy. For example, in automated trading, algorithms execute transactions at speeds far beyond human capability. While humans may intend to deploy such systems, they do not control each individual decision made by them. This raises a difficult question: can intent be said to extend to the outcomes produced by the system, or does it remain limited to the initial act of deployment?

In the Indian legal context, existing statutes such as the Information Technology Act, 2000 and the Bharatiya Nyaya Sanhita, 2023 do not explicitly address the issue of algorithmic intent. The law continues to rely on human-centered notions of culpability, which may not fully capture the complexities of AI-driven decisions. Consequently, courts are often required to interpret traditional principles in new and uncertain contexts.

Another critical dimension is the role of **bias in algorithmic decision-making**. AI systems are trained on historical data, which may reflect

existing social inequalities and prejudices. When such biases are embedded in the data, the system's outputs may perpetuate or even amplify discrimination. While these outcomes may appear intentional, they are actually the result of data-driven processes rather than conscious discrimination. This raises difficult questions about responsibility: should liability be attributed to the developers who designed the system, the organizations that deployed it, or those who provided the data?

The absence of human intent also has significant implications for **criminal liability and punishment**. Criminal law traditionally seeks to punish morally blameworthy conduct. However, when harm is caused by systems that lack intent, the basis for punishment becomes unclear. Punishing an AI system is meaningless, as it has no consciousness or capacity for moral understanding. At the same time, holding human actors liable may be problematic if they neither intended nor could reasonably foresee the harm caused.

To address these challenges, scholars and policymakers have suggested various approaches. Some advocate retaining the traditional framework by attributing liability to human actors based on negligence or recklessness. Others propose adopting strict liability regimes, where responsibility is imposed regardless of intent in cases involving high-risk technologies. There are also calls for developing new legal doctrines that specifically address the unique nature of algorithmic systems. Algorithmic decision-making marks a significant departure from traditional human decision-making, particularly due to the absence of intent. This creates a fundamental tension with established legal principles that rely on *mens rea* as the basis of liability. The challenges of opacity, autonomy, delegation, and bias further complicate the attribution of responsibility. As AI continues to evolve, it is essential for legal systems to adapt and develop new frameworks that ensure accountability while preserving the core values of justice, fairness, and responsibility.

3.3 Relevance of Strict Liability and Negligence-Based Offences

The rapid advancement of technology, particularly Artificial Intelligence (AI), has compelled legal systems to rethink traditional doctrines of criminal liability. Concepts such as intention and knowledge, which form the core of *mens rea*, often become difficult to apply in technologically complex environments. In this evolving landscape, doctrines like strict liability and negligence-based offences have gained renewed importance. These principles allow the law to impose responsibility even when a clearly provable guilty mind is absent, thereby ensuring accountability in situations where harm occurs but intent is difficult to establish.

Under traditional criminal law, as reflected in statutes like the Indian Penal Code, 1860, liability generally requires proof of both a wrongful act (*actus reus*) and a guilty mind (*mens rea*). However, strict liability and negligence-based offences serve as important exceptions to this

rule. They demonstrate that, in certain circumstances, the law prioritizes public safety, regulatory compliance, and risk prevention over the strict requirement of intention.

Strict liability offences are those in which liability is imposed without the need to prove *mens rea*. In such cases, the mere commission of the prohibited act is sufficient to establish guilt. These offences are commonly found in regulatory and public welfare domains such as environmental protection, food safety, industrial safety, and traffic regulation. The underlying rationale is that certain activities are inherently risky or socially significant, and those who engage in them must bear responsibility for any harm caused, regardless of intent.

The relevance of strict liability becomes especially clear in the context of AI and automated systems. As previously discussed, attributing intention or knowledge to AI is conceptually difficult because such systems lack consciousness and independent will. When an autonomous system causes harm—such as

a malfunctioning medical AI or an accident involving a self-driving vehicle—it may be impossible to identify a specific human intention behind the outcome. In such situations, strict liability offers a practical solution by shifting the focus from the mental state to the occurrence of harm. This ensures that victims are not left without remedies simply because *mens rea* cannot be established.

Closely related to strict liability are **negligence-based offences**, which impose liability based on a failure to exercise reasonable care. Unlike strict liability, negligence involves a form of fault, but it is assessed objectively rather than subjectively. The law applies the standard of a “reasonable person,” asking whether the accused failed to act with the level of care expected in similar circumstances. In Indian criminal law, provisions such as Section 304A of the Indian Penal Code, 1860⁹⁸⁷ (causing death by negligence) illustrate this principle.

Negligence-based liability is particularly relevant in cases involving AI systems, where harm may arise due to design flaws, insufficient testing, or improper deployment. For instance, if a company introduces an AI system without adequately ensuring its safety or reliability, it may be held negligent if the system causes harm. Similarly, developers who fail to address known risks, such as algorithmic bias or system vulnerabilities, may be considered to have breached their duty of care. In such cases, the focus shifts from intention to the standard of care exercised by those responsible for the system. An important aspect of negligence is the concept of **foreseeability**. Liability often depends on whether the harm caused was reasonably foreseeable.⁹⁸⁸ Although AI systems can behave unpredictably, certain risks—such as biased outputs, technical failures, or misuse—can often be anticipated. Failure to take reasonable steps to prevent such risks may result in liability. This approach allows the law to

adapt to technological complexity while still maintaining a fault-based framework.

The increasing reliance on strict liability and negligence-based offences also reflects a broader move towards **risk-based regulation**. In modern societies, many activities involve advanced technologies that carry significant potential for harm. By imposing liability without requiring proof of intent, the law encourages individuals and organizations to adopt higher standards of care and precaution.⁹⁸⁹ This is especially important in sectors such as healthcare, transportation, and finance, where errors can have serious consequences.

Within the Indian legal framework, although traditional criminal law emphasizes *mens rea*, courts have recognized the importance of strict liability and negligence in appropriate contexts. Judicial interpretations have often stressed that in matters affecting public welfare, the absence of intention should not prevent the imposition of liability. Additionally, statutes such as the Information Technology Act, 2000 incorporate elements of due diligence and responsibility, particularly in relation to intermediaries and digital platforms.⁹⁹⁰

However, the application of these doctrines is not without challenges. One significant concern is the risk of **over-criminalization**, where individuals or entities may be held liable even when they lack meaningful fault. This raises important questions about fairness, especially in cases where harm results from factors beyond the control of the accused. In the AI context, where systems may behave unpredictably, it is crucial to strike a balance between ensuring accountability and avoiding excessive liability.

Another challenge lies in defining the **standard of care** in technologically complex environments. As AI systems become more advanced, determining what constitutes “reasonable care” becomes increasingly

⁹⁸⁷ Section 304A of the Indian Penal Code, 1860

⁹⁸⁸ *Donoghue v Stevenson* (1932) AC 562.

⁹⁸⁹ Ulrich Beck, *Risk Society* (1992).

⁹⁹⁰ Information Technology Act, 2000.

difficult. Courts may need to rely on expert opinions, industry practices, and regulatory standards to assess whether the required level of care has been met.

Despite these challenges, strict liability and negligence-based offences remain highly relevant in addressing the legal implications of AI and other emerging technologies. They provide flexible mechanisms that enable the law to respond effectively to harm even when traditional concepts of intention and knowledge are inadequate. By focusing on outcomes, risks, and standards of care, these doctrines help bridge the gap between technological innovation and legal accountability.

Conclusion

the relevance of strict liability and negligence-based offences has grown significantly in the modern era, particularly in the context of AI-driven decision-making. While traditional criminal law emphasizes *mens rea*, these doctrines demonstrate that liability can still be imposed in its absence when necessary to protect public welfare. As technology continues to evolve, these principles will play a crucial role in ensuring that legal systems remain responsive, balanced, and just. AI in Indian criminal law presents both remarkable opportunities and formidable challenges. It can enhance investigative efficiency, strengthen forensic analysis, and improve law enforcement outcomes, but it also destabilizes traditional notions of *mens rea*, complicates accountability, introduces bias, and raises serious concerns regarding fairness and human rights. Reforming the law to meet these challenges demands a multidimensional approach: legislative amendments, judicial guidelines, procedural safeguards, technological oversight, and sustained capacity building. By balancing AI's potential benefits with principles of justice, fairness, and constitutional protections, India can develop a criminal law framework that is adaptive, resilient, and ethically grounded. Ultimately, AI must serve as a tool that augments human decision-making

rather than replaces it. *Mens rea*, accountability, and procedural fairness remain the pillars of criminal justice. AI can support evidence-based processes and reduce arbitrariness, but it cannot substitute for moral and legal responsibility. By embedding these principles into law and practice, India can ensure that its criminal justice system evolves to meet the demands of a digital, AI-driven society while upholding the rule of law, protecting fundamental rights, and delivering justice for all.

BIBLIOGRAPHY

1. Ratanlal & Dhirajlal, Criminal Law (Latest Edition)
2. K.N. Chandrasekharan Pillai, General Principles of Criminal Law.
3. Bharatiya Nyaya Sanhita, 2023.
4. Law Commission of India, Reports on Criminal Law Reform.
5. Information Technology Act, 2000.
6. Digital Personal Data Protection Act, 2023.
7. Companies Act, 2013.