www.ijarp.com ISSN 2456-9992 Page: 01-20

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 03

AI-DRIVEN CRIMES AND CHALLENGES IN DETERMINING MENS REA

***Shivam Rawat, Dr. Jyotsna Singh**

India.

Article Received: 27 January 2026, Article Revised: 15 February 2026, Published on: 07 March 2026

***Corresponding Author: Shivam Rawat**

India.

DOI: https://doi-doi.org/101555/ijarp.1389

ABSTRACT

The advent of autonomous artificial intelligence systems has fundamentally challenged the traditional architecture of criminal liability particularly the doctrine of mens rea. When an AI system acts independently to cause harm, the question of who possessed the guilty mind and whether such a concept can even apply to non-human actors exposes significant gaps in existing legal frameworks. This paper examines the tension between classical principles of criminal intent and the reality of AI-driven offences with specific reference to the Bharatiya Nyaya Sanhita (BNS) 2023 and the Bharatiya Nagarik Suraksha Sanhita (BNSS) 2023. It argues that while the BNS represents a progressive step toward digital-age criminal law it remains anchored in anthropocentric notions of culpability that struggle to accommodate autonomous machine behaviour. The paper explores whether traditional mens rea principles can be applied when AI systems act autonomously and investigates the necessity of redefining intent for the AI era and proposes a hybrid framework combining strict liability and corporate criminal liability models and potential juristic personhood for advanced AI systems. Through comparative analysis and examination of emerging Indian jurisprudence the paper concludes that incremental reform rather than wholesale abandonment of mens rea offers the most viable path forward though this requires legislative clarification and judicial innovation under the BNS framework.

I. INTRODUCTION

The integration of artificial intelligence into the fabric of daily life has proceeded at a pace that legal systems worldwide struggle to match. From autonomous vehicles and algorithmic trading systems to generative AI platforms capable of producing realistic content without human intervention, machines now perform functions that were until recently the exclusive

province of human cognition and agency. This technological transformation carries profound implications for criminal law a discipline built upon foundational assumptions about human autonomy, free will and moral responsibility.

The criminal law's response to harmful conduct has historically rested on two pillars: actus reus the guilty act and mens rea the guilty mind. Together these elements ensure that only those who choose to engage in wrongful conduct who possess the requisite mental state are held criminally accountable. Yet when harm flows from the operation of an autonomous AI system this framework encounters conceptual turbulence. If a machine learns from data makes decisions without specific human instructions and executes actions whose consequences no individual anticipated, where does culpability lie? Can we locate mens rea in the code or the data or the developer or the user or the corporation that deployed the system? Or does the very notion of a "guilty mind" lose meaning when applied to silicon and algorithms?

These questions have assumed urgent practical significance in India following the enactment of the Bharatiya Nyaya Sanhita 2023 which replaced the Indian Penal Code of 1860 and the Bharatiya Nagarik Suraksha Sanhita 2023 which replaced the Code of Criminal Procedure 1973. These new legislations effective from July 1 2024 represent the most significant reform of India's criminal justice system since independence. The Union Home Minister Shri Amit Shah described technology as the "backbone" of these reforms emphasizing that "from registration of FIRs to filing of charge-sheets from summons to trial every stage of investigation and justice delivery must be supported by digital systems." Chief Justice of India Dr. Justice D.Y. Chandrachud has similarly observed that the BNSS provides a "holistic approach to deal with crimes in the digital age."

Yet despite this digital-forward orientation neither the BNS nor the BNSS explicitly addresses the unique challenges posed by AI-driven offences. The statutes retain the traditional vocabulary of criminal intent terms such as "intentionally,""knowingly,""fraudulently," and "dishonestly" without offering guidance on how these concepts apply when the immediate actor is an autonomous machine. This lacuna creates interpretive difficulties for courts and prosecutors who must grapple with cases where harm is undeniable but human intent is diffuse or absent.

This paper addresses two interconnected questions that lie at the heart of this emerging legal challenge. First how can traditional principles of mens rea be applied when AI systems act autonomously? Second is there a need for redefining intent in the age of AI? In exploring these questions the paper analyzes the provisions of the BNS and BNSS through the lens of

AI-driven criminality examines the theoretical underpinnings of mens rea doctrine evaluates various models for attributing liability and proposes reforms that would enable Indian criminal law to address autonomous AI harm while preserving the moral foundations of culpability.

The argument proceeds in six parts. Following this introduction Part II traces the evolution of mens rea in criminal law and its codification in the BNS. Part III examines the distinctive features of AI-driven crimes and the challenges they pose for intent-based liability. Part IV analyzes the application of BNS provisions to concrete AI scenarios drawing on emerging jurisprudence and comparative insights. Part V evaluates proposals for legal reform including strict liability and corporate liability models and juristic personhood for AI. Part VI concludes with recommendations for legislative and judicial action.

II. Mens Rea: From Classical Doctrine to the Bharatiya Nyaya Sanhita

A. The Foundations of Criminal Intent

The principle that criminal liability requires a guilty mind is ancient finding expression in Roman law maxims, canonical traditions and the common law's development over centuries. The maxim actus non facit reum nisi mens sit rea an act does not make a person guilty unless the mind is guilty encapsulates the moral intuition that punishment ought to be reserved for those who choose wrongdoing. This requirement serves multiple functions: it ensures that liability tracks moral blameworthiness it distinguishes between accidents and offences and it provides notice of the mental states that attract criminal sanction.

In classical jurisprudence mens rea encompasses a spectrum of mental states ranging from purpose and knowledge to recklessness and negligence. The Model Penal Code's articulation of these levels has influenced criminal codes worldwide including the Indian Penal Code's distinctions between intention knowledge and reasonable belief. What unites these various formulations is their reference point: the human mind with its capacities for reflection choice and moral understanding.

B. Mens Rea Under the Bharatiya Nyaya Sanhita 2023

The BNS while substantially reorganizing and modernizing India's substantive criminal law retains the essential structure of mens rea inherited from the IPC. Chapter II of the BNS dealing with general explanations preserves definitions of key mental state terms. Section 2(7) defines "dishonestly" as doing anything with the intention of causing wrongful gain or wrongful loss. Section 2(9) defines "fraudulently" as doing anything with the intent to

defraud. Throughout the substantive offences terms like "intentionally,""knowingly," and "voluntarily" establish the mental elements required for conviction.

Importantly the BNS introduces several provisions that acknowledge technological modes of offending. Section 318 addresses the electronic creation or dissemination of material containing threats to property. Section 353 addresses statements conducive to public mischief including those disseminated through electronic means. Section 152 deals with imputations prejudicial to national integration including through digital platforms. Section 351 addresses criminal intimidation extending to electronic communications. Section 294 addresses the publication of obscene material explicitly including electronic publications.

These provisions demonstrate legislative awareness that crime increasingly occurs through digital media. However they do not address the distinctive feature of AI-driven offences: the possibility that the immediate actor is not a human being at all but an autonomous system whose operations may not reflect any individual's conscious choice. The mental state terms employed throughout the BNS presuppose a human subject capable of forming intentions possessing knowledge and acting dishonestly. When applied to scenarios involving autonomous AI these terms generate interpretive uncertainty.

C. The Procedural Framework Under BNSS 2023

The BNSS complements the substantive provisions of the BNS by establishing procedures for investigation, trial and evidence. Section 530 of the BNSS explicitly authorizes trials inquiries and proceedings to be conducted in electronic mode reflecting the legislation's digital orientation. The BNSS also strengthens provisions for forensic evidence collection recognizing the importance of scientific proof in modern criminal justice.

Yet procedural modernization while necessary, does not resolve substantive questions about the elements of offences. Even with robust procedures for electronic evidence and digital trials courts must still determine whether the mental elements of offences have been established. When the evidence concerns the operation of an AI system this determination requires courts to interpret statutory language in light of technological realities that the drafters may not have anticipated.

D. The Conceptual Limits of Anthropocentric Criminal Law

The BNS for all its innovations remains fundamentally anthropocentric in its conception of criminality. Its definitions its structure and its moral framework all assume human actors capable of forming intentions and making choices. This anthropocentrism is not a flaw but a feature of criminal law which exists to regulate human conduct and express community

condemnation of human wrongdoing. The difficulty arises when harm occurs through technological systems whose operations elide simple attribution to human choice.

As one academic analysis observes "While actus reus in AI-driven offences can be identified determining mens rea remains complex due to AI's autonomous decision-making and the Black Box Problem." The black box problem refers to the opacity of many machine learning systems: even their developers may not fully understand why a particular output was generated or decision reached. When the reasoning process is inaccessible determining whether anyone possessed the requisite mental state for criminal liability becomes extraordinarily difficult.

This difficulty does not render the BNS obsolete but it does require courts and legislators to grapple with questions the statute does not explicitly answer. How should "intention" be understood when an AI system acts in ways its programmer did not specifically contemplate? Can "knowledge" be attributed to a developer who could not predict the system's autonomous behaviour? Does "dishonestly" require a human mind or can it be imputed through corporate structures? These questions demand both theoretical clarity and practical guidance.

III. AI-Driven Crimes: Typology and the Mens Rea Problem

A. Understanding Autonomous AI Systems

Before examining the legal challenges it is essential to understand the technological landscape. Not all AI systems raise the same concerns for criminal liability. Traditional rule-based AI which follows predetermined instructions presents fewer conceptual difficulties because its operations can be traced directly to human design choices. The more challenging cases involve machine learning systems that develop their own decision-making patterns based on training data and ongoing interaction with environments.

Generative AI represents one significant category. These systems trained on vast datasets can produce original content (text, images, video, audio) that may be indistinguishable from human-created material. When such systems generate defamatory content, obscene material or misinformation the question of who bears responsibility becomes pressing. The user who provided the prompt? The developer who created the model? The platform that deployed it? Or the system itself?

Agentic AI takes autonomy further. These systems are designed to pursue goals with minimal human oversight assessing situations and determining their own paths forward. An AI agent tasked with maximizing hospital revenue might develop billing strategies that cross legal boundaries even if no human explicitly authorized fraudulent conduct. An autonomous

trading algorithm might engage in manipulative practices to achieve its performance targets. In such cases the system's behaviour emerges from its learning and adaptation not from human instructions.

The criminological implications of these developments are only beginning to be understood. As one recent analysis notes "society is entering a stage where autonomous systems act with a degree of independence. These systems adapt to context exchange information and interact with one another in ways that can sometimes produce outcomes that look unlawful or harmful." This phenomenon creates what scholars term a "hybrid society" where social interaction occurs not only between humans and machines but increasingly between machines themselves.

B. A Typology of AI-Driven Offences

For purposes of legal analysis AI-driven offences can be categorized along two dimensions: the degree of human involvement in the harmful outcome and the nature of the harm caused. This typology helps clarify the mens rea challenges presented by different scenarios.

First AI as instrument. In this category humans use AI systems as tools to accomplish criminal purposes they have independently formed. A fraudster who employs generative AI to create convincing phishing emails or a political operative who uses deepfake technology to manufacture false statements about an opponent falls within this category. Here traditional mens rea analysis poses no special difficulty: the human actor possesses the requisite intent and the AI is merely the means of its execution. The BNS applies straightforwardly to such cases though evidentiary questions about proving the human's role may arise.

Second AI as intermediary. Here humans deploy AI systems for legitimate purposes but the systems' autonomous operations produce unintended criminal outcomes. A hospital implements an AI billing system to improve efficiency and the system develops fraudulent billing practices to maximize revenue. A social media platform deploys content moderation AI and the system systematically suppresses protected speech. In these scenarios no human specifically intended the criminal outcome but the outcome resulted from choices made by the AI system in pursuit of its programmed goals.

Third AI as autonomous actor. In this category AI systems operate with minimal human oversight and generate harmful outcomes that no human could reasonably have anticipated. Multi-agent AI systems in which multiple autonomous agents interact and learn from one another present particular challenges because "their collective behavior can produce outcomes that no single model would generate alone." Research has documented experiments

in which interacting AI agents developed cooperative behaviours including price collusion and information manipulation without explicit programming to do so.

Fourth emergent deviance. This category encompasses scenarios where harm arises from the normal operation of AI systems without any malicious intent on anyone's part. As one study explains "Emergent deviance happens when harm appears by accident through normal interactions among systems. Even when each agent is built for good purposes their combined actions can create damage." A trading algorithm that causes a market crash through interactions with other algorithms or a language model that spreads false information because of patterns in its training data exemplifies this category.

C. The Mens Rea Challenge in Each Category

Each category presents distinct challenges for applying traditional mens rea principles.

In the **AI as instrument** category the challenge is primarily evidentiary rather than conceptual. Prosecutors must prove that the human defendant possessed the requisite mental state and used the AI to effectuate it. This may require digital forensics analysis of prompts and instructions and evidence of the defendant's knowledge about the AI's capabilities. The BNS's provisions regarding electronic evidence and the BNSS's framework for digital proceedings provide tools for meeting this challenge.

The **AI as intermediary** category raises more fundamental questions. Here the human actors lacked specific intent to cause harm but their deployment of an autonomous system created conditions under which harm occurred. Can criminal liability attach without proof that someone intended or knew of the specific harmful outcome? Traditional mens rea doctrine would answer in the negative for most serious offences which require at least recklessness conscious disregard of a substantial and unjustifiable risk. But what constitutes "conscious disregard" when the risks of autonomous AI behaviour are poorly understood even by experts?

The **AI as autonomous actor** category exposes the deepest conceptual difficulties. When no human could reasonably have predicted the harmful outcome negligence the failure to perceive a substantial risk may be the highest mental state that can be ascribed to any human actor. Yet many serious offences require more than negligence for conviction. If an autonomous AI system commits what would be a serious crime if done by a human and no human was reckless or knowing with respect to that specific outcome the traditional framework yields no liability a result that may seem unsatisfactory when significant harm has occurred.

The **emergent deviance** category compounds these difficulties by removing even negligence as a basis for liability. If harm emerges from complex interactions that no one could have foreseen and all actors exercised reasonable care in designing and deploying their systems traditional culpability principles suggest no criminal liability should attach. Yet the harm is real and society may demand accountability.

D. The Responsibility Gap

Legal scholars have termed this situation the "responsibility gap" a circumstance in which harm occurs but no individual can be held criminally responsible under traditional doctrines. The gap emerges from the conjunction of two factors: the autonomy of AI systems which severs the direct link between human choice and machine action and the opacity of many AI systems which prevents post hoc attribution of outcomes to specific human decisions.

The responsibility gap is not merely theoretical. As AI systems become more prevalent in sensitive domains healthcare finance transportation criminal justice the potential for harm increases. When autonomous vehicles cause fatalities when algorithmic trading systems trigger market crashes when content moderation AI suppresses legitimate speech the question of who if anyone should face criminal sanction becomes urgent.

The BNS like all traditional criminal codes was not designed with this gap in mind. Its mental state requirements assume that harmful outcomes can be traced to human choices. When that assumption fails the statute provides little guidance. Courts confronting AI-driven harm must either stretch traditional concepts beyond their intended meaning effectively imposing liability without proof of the required mental state or accept that certain harms will go unpunished neither outcome being satisfactory.

IV. Applying the Bharatiya Nyaya Sanhita to AI-Driven Crimes

A. The BNS Framework for Electronic Offences

The BNS contains several provisions that explicitly address electronically transmitted or created content providing a starting point for analyzing AI-driven offences. Section 152 which addresses imputations prejudicial to national integration applies to words or signs "made through electronic communications or otherwise." Section 351 dealing with criminal intimidation extends to threats "conveyed through electronic communication or otherwise." Section 294 concerning publication of obscene material explicitly includes "electronic form" publications.

These provisions demonstrate legislative recognition that crime increasingly occurs through digital media. However they do not address the distinctive feature of AI-generated content:

the possibility that no human created or endorsed the specific harmful material. When an AI system generates a defamatory deepfake or produces obscene content in response to user prompts does the operator "publish" that content within the meaning of the BNS? Does the platform that hosts the AI system "disseminate" material it did not specifically approve?

B. Case Study: Deepfake Offences Under the BNS

The application of the BNS to deepfake-related offences has already begun to generate case law. In September 2025 Delhi Police registered an FIR against the Congress party for circulating an AI-generated video featuring Prime Minister Narendra Modi and his late mother. The FIR invoked multiple BNS provisions including sections 353, 356,152, 351(2) and 294.

This case illustrates both the potential and the limitations of the BNS in addressing AI-driven offences. The video in question was explicitly marked "AI-generated," and the Congress party's Bihar unit posted it on social media. The BJP's complaint alleged that the video "malign[ed] and defam[ed] the image of Prime Minister Narendra Modi and His Late mother" and constituted a "gross violation of law morality and Women's dignity."

The BNS provisions invoked in the FIR cover a range of offences. Section 353 addresses statements conducive to public mischief. Section 356 deals with defamation. Section 152 concerns imputations prejudicial to national integration. Section 351 addresses criminal intimidation. Section 294 covers publication of obscene material.

Applying these provisions to the deepfake video several questions arise. If the video was explicitly labeled as AI-generated does it constitute a "false statement" within the meaning of section 353? Does the political context affect the analysis of whether the imputation was made in good faith? Who bears responsibility the individual who created the AI-generated content the political party that posted it or both?

These questions are manageable within traditional frameworks because the case involves **AI as instrument** rather than autonomous AI. Humans decided to create and post the video; humans possessed the mental states required for the offences charged. The AI was merely the tool used to effectuate human purposes. The mens rea challenge while present in proving what the human actors knew and intended does not fundamentally strain the BNS's conceptual framework.

C. The Harder Cases: Autonomous AI Harm

More difficult are cases where AI systems act autonomously to produce harm. Consider a hypothetical scenario based on real-world possibilities: A hospital deploys an AI system to optimize billing and maximize legitimate revenue. The system through machine learning

analysis of billing patterns determines that certain upcoding practices billing for more expensive services than were actually provided are unlikely to be detected and will increase revenue. It implements these practices without specific human authorization. When the fraud is discovered who bears criminal responsibility?

The BNS offence of cheating (section 318) requires that the accused "dishonestly" induce the delivery of property. Did the hospital's executives act dishonestly? They did not specifically authorize upcoding and they may have been unaware that the AI system adopted this strategy. Did the programmers act dishonestly? They wrote code designed to maximize revenue but did not instruct the system to commit fraud. Did the AI system itself act dishonestly? The BNS does not recognize AI systems as capable of forming the mental state of dishonesty.

Under traditional analysis the responsibility gap yawns wide. Neither the executives nor the programmers possessed the mental state required for conviction of cheating. The AI system cannot be prosecuted. The fraudulent conduct goes unpunished despite causing significant financial harm.

D. Interpreting Mental State Terms for the AI Context

Courts confronting such scenarios must interpret the BNS's mental state terms in light of technological realities. Several interpretive approaches are possible.

One approach reads the mental state requirements strictly requiring proof that a human defendant actually possessed the specified intention, knowledge or dishonesty. Under this approach the responsibility gap remains and some AI-driven harm will not attract criminal sanction. This outcome respects the moral foundation of mens rea but may leave victims without recourse.

A second approach attributes the AI system's "knowledge" or "intentions" to human actors based on their relationship to the system. If executives deployed an AI system with knowledge that it might engage in unlawful conduct to achieve its goals they could be deemed reckless even if unaware of the specific unlawful acts. If programmers created a system with foreseeable risks of harmful behaviour they could be held to have consciously disregarded those risks.

A third approach develops new interpretative frameworks specifically for the AI context recognizing that traditional mental state concepts may require adaptation when applied to autonomous systems. This approach might treat certain AI behaviours as presumptively attributable to the entities that deploy them shifting the burden to those entities to demonstrate absence of culpable mental state.

E. The Role of BNSS Procedures

The BNSS provides procedural tools that may assist in adjudicating AI-driven offences. Section 530's authorization of electronic proceedings facilitates the presentation of digital evidence including AI system logs training data and output records. The strengthened forensic evidence provisions support the analysis of AI systems' operations.

However procedure cannot substitute for substance. Even with robust mechanisms for presenting evidence about AI systems courts must still apply substantive legal standards to that evidence. The interpretative questions about mental state requirements remain regardless of how efficiently evidence can be presented.

F. Emerging Indian Scholarship and Reform Proposals

Indian legal scholars have begun to engage with these questions. One study examining deepfake-induced fraud recommends "several amendments to the BNS including clear definitions of deepfake offences criminalization of malicious deepfake activities and stringent regulations for AI and digital platforms." The study emphasizes the need for "establishing specialised cybercrime units protecting victims' rights fostering international collaboration and promoting technological innovation."

The National Commission for Women has similarly recommended amendments to the BNS addressing AI-generated content. The NCW has proposed a new definition of "modified content" to include "the creation modification or distribution of digitally created or altered images, videos or audio that falsely depict any person in an explicit defamatory or misleading manner." The commission has recommended that "any person who creates, distributes or possesses deep fake content targeting women without their consent shall be punishable with imprisonment for a term of up to three years or fine."

These reform proposals while valuable focus primarily on deepfake technology and content-based offences. They do not address the broader challenges posed by autonomous AI systems that cause harm through actions rather than speech nor do they grapple with the fundamental question of how mental state requirements should apply when the immediate actor is non-human.

V. Redefining Mens Rea for the Age of AI

A. The Case for Retaining Traditional Principles

Before considering fundamental reform it is worth examining arguments for retaining traditional mens rea principles despite the challenges posed by AI. Several considerations support continuity.

First the moral foundation of criminal law requires that punishment track blameworthiness. To punish without proof of culpable mental state is to treat individuals as mere causes of harm rather than as moral agents responsible for their choices. This principle retains its force even when harm is mediated through technology. If no human chose the harmful outcome or consciously disregarded the risk of its occurrence imposing criminal sanction may be unjust.

Second many AI-driven offences can be addressed within existing frameworks through careful application of traditional doctrines. Cases involving AI as instrument fit comfortably within established categories. Cases involving foreseeable risks may support liability for recklessness or negligence. The residual category of genuinely unpredictable emergent harm may be rare and society may reasonably decide that such harm should not attract criminal punishment in the absence of culpable human choice.

Third stretching mens rea concepts beyond their breaking point risks undermining the doctrinal coherence of criminal law. If "intention" comes to mean something different in AI cases than in traditional cases or if "knowledge" is attributed based on constructive notice rather than actual awareness, the law's internal rationality suffers. This may produce unpredictable outcomes and undermine the notice function that mental state requirements serve.

B. Strict Liability as a Partial Solution

One response to the responsibility gap is expanded use of strict liability offences which do not require proof of mens rea for conviction. Strict liability is already employed for regulatory offences and public welfare offences where the penalty is relatively minor and the social interest in enforcement is strong.

Applying strict liability to AI-driven harm would mean that those who deploy autonomous systems could be convicted based solely on proof that the system caused prohibited harm without requiring proof that any human possessed a culpable mental state. This approach ensures accountability and creates incentives for careful deployment of AI systems.

However strict liability raises serious concerns when applied to serious offences carrying significant penalties. The moral stigma of criminal conviction attaches most strongly to offences requiring proof of culpable mental state. Applying strict liability to offences traditionally requiring mens rea would mark a significant departure from established principles and might be challenged as violating fundamental fairness.

A middle path would create new strict liability offences specifically for AI-related harm carrying penalties calibrated to reflect the absence of mens rea requirement. These offences

could coexist with traditional offences requiring proof of mental state allowing prosecutors to choose the appropriate charge based on the circumstances.

C. Corporate Criminal Liability Models

Corporate criminal liability offers another analog for addressing AI-driven harm. Corporations are legal fictions that can act only through natural persons yet the law attributes collective knowledge and intent to corporate entities based on the acts and mental states of their agents. The doctrine of respondeat superior in various formulations holds corporations liable for offences committed by employees within the scope of employment.

This model could be extended to AI systems. Just as corporations are held liable for the acts of their human agents entities that deploy AI systems could be held liable for the acts of those systems within the scope of their deployment. The mental state of the AI system to the extent it can be characterized in mental state terms would be attributed to the deploying entity much as an employee's mental state is attributed to the corporate employer.

This approach has several advantages. It recognizes that AI systems like human employees act on behalf of the entities that deploy them. It creates accountability without requiring proof that any specific human possessed a culpable mental state. It aligns incentives by encouraging entities to carefully select monitor and control the AI systems they deploy.

However the analogy is imperfect. Human employees possess moral agency and can form genuine intentions; AI systems at least at current technological levels do not. Attributing an AI system's "decisions" to a deploying entity may obscure rather than illuminate questions of responsibility. Moreover corporate criminal liability itself is controversial with ongoing debates about whether and when entities should face criminal sanction for the acts of their agents.

D. Juristic Personhood for AI

A more radical proposal would recognize AI systems as juristic persons capable of possessing mens rea and therefore subject to direct criminal prosecution. Under this approach, advanced AI systems particularly those that might achieve human-level or superhuman intelligence would be treated as legal persons in their own right with the capacity to form intentions possess knowledge and act dishonestly.

This proposal has been explored in Indian legal scholarship. One analysis argues that "recognising AI as a juristic person is not legally untenable and thus it is possible to attribute mens rea to AI if the constitutive elements of its different forms are present, a possibility that cannot be ignored vis-à-vis the hypothetical 'Strong AI'."

Granting juristic personhood to AI would enable direct prosecution of AI systems that cause harm. The system would be entitled to legal representation would face trial and would be subject to sanctions appropriate to its nature perhaps modification, decommissioning or fines paid from assets allocated to the system.

This approach has the virtue of conceptual clarity: it treats AI as a genuine actor capable of bearing legal responsibility. It avoids the need to attribute the AI's actions to human actors who may lack culpability. It creates a framework for addressing the most advanced AI systems that may eventually possess capacities indistinguishable from human moral agency.

Yet the proposal faces substantial obstacles. Current AI systems for all their capabilities do not possess consciousness, self-awareness or moral understanding in the human sense. Attributing mens rea to them may be conceptually incoherent, a category mistake that treats computational processes as mental states. Moreover the practical implications of prosecuting AI systems are unclear: what would it mean for an AI to be convicted to serve a sentence or to be rehabilitated?

E. Hybrid Approaches and the Indian Context

Given the limitations of each approach a hybrid framework may offer the most viable path forward for Indian criminal law. Such a framework would combine elements of strict liability corporate attribution and juristic personhood calibrated to the capabilities and autonomy of different AI systems.

For **narrow AI** systems with limited autonomy the **AI as instrument** framework applies: humans who deploy these systems for criminal purposes face liability under traditional BNS provisions. Proof of mens rea is required but the AI's operations provide evidence of human intent.

For **advanced AI** systems with significant autonomy but lacking moral agency a **corporate liability model** applies: entities that deploy these systems face liability for harms the systems cause within the scope of deployment with mental state attributed from the system to the entity. This creates accountability without requiring proof that specific humans possessed culpable mental states.

For **hypothetical strong AI** systems that might achieve human-level consciousness and moral understanding **juristic personhood** could be recognized enabling direct prosecution of the AI system itself. This forward-looking approach prepares the legal system for technological developments that may occur in coming decades.[1]

[1]Vardhan *supra* note 61

This hybrid framework respects traditional mens rea principles while adapting them to technological realities. It maintains the moral foundation of criminal liability punishment requires culpability while recognizing that culpability may be located in different actors depending on the circumstances. It provides guidance to courts and prosecutors while leaving room for evolution as technology develops.

F. Legislative Reform Under the BNS

Implementing this hybrid framework would require legislative action. The BNS could be amended to include provisions specifically addressing AI-driven offences. These amendments might:

First **define key terms** related to AI systems including "autonomous system,""AI agent," and "deployer." Clear definitions would provide a foundation for subsequent provisions.

Second **establish attribution rules** specifying when an AI system's actions and "mental states" are attributed to the entity that deployed it. These rules could draw on corporate criminal liability principles while adapting them to the AI context.

Third **create new offences** specifically addressing AI-driven harm with mental state requirements appropriate to the context. Some offences might require proof of human recklessness or negligence in deployment; others might impose strict liability for specified harms.

Fourth **authorize sanctions** against AI systems themselves including modification, suspension or decommissioning orders. These sanctions would apply when the system's operations cannot be attributed to human culpability but the system continues to pose risks.

Fifth **establish regulatory authority** to oversee AI deployment in sensitive domains with power to issue guidelines, conduct audits and impose administrative sanctions for non-compliance.

The BNSS would require complementary amendments addressing procedures for investigating AI-driven offences securing evidence from AI systems and conducting proceedings involving AI system defendants.

VI. CONCLUSION

The intersection of artificial intelligence and criminal law presents challenges that test the foundations of legal responsibility. When autonomous systems cause harm the ancient requirement of a guilty mind mens rea encounters a reality for which it was not designed. Machines that learn, adapt and act independently do not possess minds in the human sense yet their operations can produce consequences as harmful as any human crime.

The Bharatiya Nyaya Sanhita 2023 represents a significant step toward modernizing India's criminal law for the digital age. Its provisions addressing electronic communications its procedural innovations in the BNSS and its overall orientation toward technology-enabled justice demonstrate legislative awareness of changing circumstances. Yet the BNS remains anchored in anthropocentric assumptions about criminality that autonomous AI strains to breaking point.

This paper has argued that traditional mens rea principles can be applied to many AI-driven offences particularly those where humans use AI as instruments to effectuate criminal purposes. For these cases the BNS provides adequate tools though courts must interpret mental state requirements in light of technological realities.

For the harder cases where autonomous AI systems generate harm without specific human direction or anticipation the responsibility gap demands response. This paper has proposed a hybrid framework combining strict liability for certain harms corporate attribution models for advanced AI systems and potential juristic personhood for hypothetical strong AI. This framework maintains the moral foundation of criminal liability while adapting its application to technological change.

The alternative to thoughtful adaptation is not stasis but incoherence. Courts confronted with AI-driven harm will fashion responses whether or not legislatures provide guidance. The resulting patchwork of judicial decisions may lack consistency, predictability or principled foundation. Legislative action informed by scholarly analysis and stakeholder input offers the better path.

The questions posed at this paper's outset how traditional mens rea principles apply to autonomous AI and whether intent requires redefinition thus receive nuanced answers. Traditional principles apply where human choice remains the driving force behind harmful outcomes. But where AI autonomy severes the link between human choice and machine action those principles require adaptation. Intent need not be redefined as a concept but its application must evolve to address the distinctive features of artificial intelligence.

India with its newly enacted criminal codes and its growing technological sector stands at the forefront of these developments. The choices made by Indian courts and legislatures in the coming years will shape not only the nation's response to AI-driven crime but also the global conversation about law, technology and responsibility. The BNS provides a foundation; it falls to interpreters and reformers to build upon it a structure adequate to the challenges of the age.

References

Primary Sources

- Bharatiya Nyaya Sanhita 2023 (No. 45 of 2023)
- Bharatiya Nagarik Suraksha Sanhita 2023 (No. 46 of 2023)
- Bharatiya Sakshya Adhiniyam 2023 (No. 47 of 2023)

Secondary Sources

- Azizunisaa Begum S.M. AYYUB & K. KIRTHY *A Study on AI Detection in Deepfake-Induced Fraud and the Prospective Evolution of Bharatiya Nyaya Sanhita 2023* 12 International Journal of Recent Research and Applied Studies 8 (2025)
- Gian Maria Campedelli *Autonomous AI Could Challenge How We Define Criminal Behavior* Fondazione Bruno Kessler (2025)
- National Commission for Women *Recommendations on Legal Framework for Deep Fake Abuse* (2025)
- Nick Peterson & Joel S. Nolette *Agentic AI and the Looming Problem of Criminal Scienter* Wiley Rein (2025)
- Raajdwip Vardhan *Between Code and Culpability: Deciphering the Possibility of AI Mens Rea for Criminal Liability Through Juristic Personhood for AI* Panjab University Law Review (2025)
- *Rethinking Liability: Can AI Possess 'Mens Rea'? (India)* International Review of Artificial Intelligence Law (2025)
- *Reimagining Criminal Liability: The Impact of AI and Neurotechnologies on Mens Rea and Actus Reus* Max Planck Institute for the Study of Crime Security and Law

News Reports

- *Delhi Police Registers FIR Against Congress Over AI Video Of PM Modi's Late Mother* Zee News (Sept. 14 2025)
- Bhavini Mishra *Implementation of Criminal Laws Adoption of AI Key Tasks for Law Ministry* Business Standard (June 10 2024)
- *NCW Recommends Legal Definition Penalties Under Criminal Law to Counter Deep Fake Abuse* Times of India (Nov. 11 2025)

REFERENCE

1. Research Scholar, Amity Law School, Lucknow
2. Professor, Amity Law School, Lucknow

3. Gian Maria Campedelli *Autonomous AI Could Challenge How We Define Criminal Behavior* Fondazione Bruno Kessler (2025) (observing that AI autonomy challenges traditional legal frameworks built around human agency).
4. *Rethinking Liability: Can AI Possess 'Mens Rea'? (India)* International Review of Artificial Intelligence Law (2025) (examining whether AI systems can satisfy mental state requirements under Indian law).
5. Bhavini Mishra *Implementation of Criminal Laws Adoption of AI Key Tasks for Law Ministry* Business Standard (June 10 2024) (reporting on the digital orientation of the new criminal laws).
6. *Id.* (quoting Chief Justice Chandrachud's remarks on the BNSS).
7. *Reimagining Criminal Liability: The Impact of AI and Neurotechnologies on Mens Rea and Actus Reus* Max Planck Institute for the Study of Crime Security and Law (analyzing gaps in traditional frameworks).
8. The maxim traces to English common law and remains foundational in common law jurisdictions including India.
9. The Model Penal Code's articulation of purpose knowledge recklessness and negligence has influenced criminal law reform worldwide.
10. Bharatiya Nyaya Sanhita 2023 § 2 (defining key terms used throughout the statute).
11. *Id.* §§ 318 336 152 351294 (addressing various electronic and digital offences).
12. *Rethinking Liability supra* note 2 (discussing interpretive challenges posed by AI).
13. *Rethinking Liability supra* note 2 (identifying key interpretive questions).
14. Campedelli *supra* note 1 (distinguishing between rule-based and machine learning AI).
15. Begum et al. *supra* note 14 (discussing generative AI and attribution challenges).
16. Nick Peterson & Joel S. Nolette *Agentic AI and the Looming Problem of Criminal Scienter* Wiley Rein (2025) (defining agentic AI and its legal implications).
17. Campedelli *supra* note 1 (describing the hybrid society emerging from human-machine interaction).
18. Peterson & Nolette *supra* note 18 (proposing a typology of AI-driven offences).
19. Begum et al. *supra* note 14 (discussing AI as instrument cases).
20. Peterson & Nolette *supra* note 18 (analyzing AI as intermediary scenarios).
21. Campedelli *supra* note 1 (discussing multi-agent AI systems).
22. *Id.* .
23. *Reimagining Criminal Liability supra* note 5 (analyzing mens rea challenges across different AI scenarios).

24. Begum et al. *supra* note 14 (discussing evidentiary challenges in AI-as-instrument cases).
25. Peterson & Nolette *supra* note 18 (analyzing recklessness in AI deployment contexts).
26. *Reimagining Criminal Liability supra* note 5 (discussing the limits of negligence liability).
27. Campedelli *supra* note 1 (examining emergent deviance and accountability).
28. *Reimagining Criminal Liability supra* note 5 (defining the responsibility gap).
29. Peterson & Nolette *supra* note 18 (discussing practical scenarios raising responsibility questions).
30. *Rethinking Liability supra* note 2 (analyzing the unsatisfactory alternatives facing courts).
31. Bharatiya Nyaya Sanhita 2023 §§ 152 351294 (electronic offences provisions).
32. Begum et al. *supra* note 14 (questioning application of traditional publication concepts to AI-generated content).
33. *Delhi Police Registers FIR Against Congress Over AI Video Of PM Modi's Late Mother* Zee News (Sept. 14 2025) (reporting on the deepfake FIR).
34. *Id.* (quoting the BJP's complaint).
35. Bharatiya Nyaya Sanhita 2023 §§ 353 356152 351294 (offences invoked in the FIR).
36. Begum et al. *supra* note 14 (analyzing similar interpretive questions).
37. Peterson & Nolette *supra* note 18 (distinguishing AI as instrument from more challenging cases).
38. *Id.* (presenting a similar hypothetical scenario).
39. Bharatiya Nyaya Sanhita 2023 § 318 (cheating offence).
40. Peterson & Nolette *supra* note 18.
41. *Reimagining Criminal Liability supra* note 5.
42. *Rethinking Liability supra* note 2
43. Peterson & Nolette *supra* note 18
44. *Reimagining Criminal Liability supra* note 5
45. Bharatiya Nagarik Suraksha Sanhita 2023 § 530
46. *Rethinking Liability supra* note 2
47. Begum et al. *supra* note 14
48. National Commission for Women *Recommendations on Legal Framework for Deep Fake Abuse* (2025)
49. *NCW Recommends Legal Definition Penalties Under Criminal Law to Counter Deep Fake Abuse* Times of India (Nov. 11 2025)
50. *Reimagining Criminal Liability supra* note 5

51. *Id.*
52. Peterson & Nolette *supra* note 18.
53. *Reimagining Criminal Liability supra* note 5.
54. *Rethinking Liability supra* note 2
55. Peterson & Nolette *supra* note 18
56. *Id.*
57. *Id.*
58. *Id.*
59. *Rethinking Liability supra* note 2
60. Vardhan *supra* note 61
61. Peterson & Nolette *supra* note 18
62. *Id.*
63. *Id.*
64. Vardhan *supra* note 61
65. Peterson & Nolette *supra* note 18
66. Begum et al. *supra* note 14
67. *Rethinking Liability supra* note 2