INDIAN JOURNAL OF LEGAL AFFAIRS AND RESEARCH

VOLUME 3 ISSUE 1

Peer-reviewed, open-access, refereed journal

IJLAR

+91 70421 48991
editor@ijlar.com
www.ijlar.com

DISCLAIMER

The views and opinions expressed in the articles published in the Indian Journal of Legal Affairs and Research are those of the respective authors and do not necessarily reflect the official policy or position of the IJLAR, its editorial board, or its affiliated institutions. The IJLAR assumes no responsibility for any errors or omissions in the content of the journal. The information provided in this journal is for general informational purposes only and should not be construed as legal advice. Readers are encouraged to seek professional legal counsel for specific legal issues. The IJLAR and its affiliates shall not be liable for any loss or damage arising from the use of the information contained in this journal.

Introduction

Welcome to the Indian Journal of Legal Affairs and Research (IJLAR), a distinguished platform dedicated to the dissemination of comprehensive legal scholarship and academic research. Our mission is to foster an environment where legal professionals, academics, and students can collaborate and contribute to the evolving discourse in the field of law. We strive to publish high-quality, peer-reviewed articles that provide insightful analysis, innovative perspectives, and practical solutions to contemporary legal challenges. The IJLAR is committed to advancing legal knowledge and practice by bridging the gap between theory and practice.

Preface

The Indian Journal of Legal Affairs and Research is a testament to our unwavering commitment to excellence in legal scholarship. This volume presents a curated selection of articles that reflect the diverse and dynamic nature of legal studies today. Our contributors, ranging from esteemed legal scholars to emerging academics, bring forward a rich tapestry of insights that address critical legal issues and offer novel contributions to the field. We are grateful to our editorial board, reviewers, and authors for their dedication and hard work, which have made this publication possible. It is our hope that this journal will serve as a valuable resource for researchers, practitioners, and policymakers, and will inspire further inquiry and debate within the legal community.

Description

The Indian Journal of Legal Affairs and Research is an academic journal that publishes peer-reviewed articles on a wide range of legal topics. Each issue is designed to provide a platform for legal scholars, practitioners, and students to share their research findings, theoretical explorations, and practical insights. Our journal covers various branches of law, including but not limited to constitutional law, international law, criminal law, commercial law, human rights, and environmental law. We are dedicated to ensuring that the articles published in our journal adhere to the highest standards of academic rigor and contribute meaningfully to the understanding and development of legal theories and practices.

"CRIMINAL LIABILITY FOR AI-GENERATED HARM IN INDIA: RETHINKING MENS REA IN AUTONOMOUS SYSTEMS"

AUTHORED BY - TAVLEEN KAUR
Institution: Chandigarh University, Gharuan, Punjab

ABSTRACT

The rapid advancement of artificial intelligence (AI) has introduced complex challenges to traditional principles of criminal liability, particularly the requirement of *mens rea*. Autonomous AI systems, capable of making decisions without direct human intervention, disrupt the foundational assumption that criminal responsibility is premised on a culpable mental state. This paper examines the limitations of applying conventional doctrines of *actus reus* and *mens rea* to AI-generated harm within the Indian legal framework.

Through a doctrinal and analytical approach, the study examines the limitations of AI systems in possessing intention, knowledge, or awareness in the legal sense, thereby creating a significant gap in attributing criminal liability. It evaluates alternative models of liability, including developer liability, user responsibility, and strict liability, and highlights their respective limitations in addressing the complexities of autonomous systems. The paper further analyses the fragmented nature of the existing Indian legal framework, which lacks a comprehensive mechanism to regulate AI-related harm.

In response, the study proposes a hybrid liability model that attributes responsibility to human actors within the AI lifecycle based on factors such as control, foreseeability, and negligence, while also recognizing the need for new categories of liability tailored to artificial intelligence. The paper concludes that a balanced and adaptive legal framework is essential to ensure accountability without undermining technological innovation in the evolving landscape of AI.

SECTION 1: INTRODUCTION

When a machine causes harm, who should be held criminally liable? The rapid advancement of artificial intelligence (AI) has significantly transformed contemporary society, enabling machines to perform tasks that traditionally required human cognition. From autonomous vehicles to algorithm-driven decision-making systems in healthcare and finance, AI technologies are increasingly capable of functioning with minimal or no human intervention. While these developments enhance efficiency and innovation, they simultaneously introduce complex legal challenges, particularly in instances where such systems cause harm.

Historically, criminal liability has been premised on the coexistence of two essential elements: *actus reus* (the wrongful act) and *mens rea* (the guilty mind). The foundational principle encapsulated in the maxim *actus non facit reum nisi mens sit rea* underscores that liability arises only when a culpable mental state accompanies a prohibited act. This framework, however, is inherently human-centric, as it presupposes intention, knowledge, or recklessness attributes that artificial intelligence systems do not possess in the legal or moral sense.

The emergence of autonomous AI systems challenges this traditional paradigm. Unlike conventional tools, these systems can process vast amounts of data, learn from patterns, and make independent decisions without direct human control. Consequently, when harm results from such autonomous actions, such as accidents involving self-driving vehicles or erroneous algorithmic decisions, the attribution of criminal liability becomes uncertain and contested.

Existing legal scholarship has explored the broader implications of AI in law, often focusing on regulatory frameworks, ethical considerations, and civil liability. However, a significant gap remains in the application of core criminal law doctrines, particularly the concept of *mens rea*, to AI-generated harm within the Indian legal context. The absence of a clear legal framework addressing such scenarios highlights the inadequacy of traditional principles in dealing with emerging technological realities.

In light of these challenges, this paper seeks to examine the limitations of existing criminal liability doctrines in the context of autonomous AI systems. It aims to analyze whether the requirement of *mens rea* can be meaningfully applied to AI-related offences and to explore alternative models of liability, including developer liability, user responsibility, and strict liability approaches. The scope of this study is confined to a doctrinal analysis of Indian criminal law, supplemented by limited comparative insights, with the objective of proposing a more coherent and adaptable legal framework for addressing AI-generated harm.

SECTION 2: UNDERSTANDING AI AND AUTONOMOUS SYSTEMS

Artificial intelligence (AI) refers to the development of computer systems capable of performing tasks that traditionally require human intelligence, including reasoning, learning, decision-making, and perception. It encompasses a range of technologies such as machine learning, deep learning, and natural language processing, which enable systems to analyze data, identify patterns, and improve performance over time without explicit programming. These technologies have facilitated the widespread integration of AI into various sectors, significantly transforming the manner in which decisions are made and tasks are executed.

In contemporary society, AI is embedded in numerous everyday applications. Search engines utilize sophisticated algorithms to deliver relevant results, while recommendation systems employed by digital platforms analyze user behavior to suggest personalized content. Similarly, AI plays a crucial role in sectors such as healthcare, where it assists in diagnostics and treatment planning, and in finance, where it is used for fraud detection and risk assessment. These applications demonstrate the increasing reliance on AI systems in decision-making processes that have real-world consequences.

A more advanced form of artificial intelligence is autonomous AI, which is capable of making decisions and taking actions without direct human intervention. Unlike traditional systems that operate based on predefined instructions, autonomous AI systems learn from data, adapt to changing circumstances, and function independently within dynamic environments. Examples include self-driving vehicles and automated decision-making systems that operate in real time. The

defining characteristic of such systems is their ability to act without continuous human oversight, thereby reducing direct human control over outcomes.

This autonomy raises significant legal concerns, particularly in situations where harm is caused as a result of AI-driven decisions. The independent functioning of such systems challenges traditional assumptions of control and accountability, as the outcomes cannot always be directly traced to a specific human intention. Consequently, the increasing deployment of autonomous AI systems necessitates a re-evaluation of existing legal frameworks, especially in the context of criminal liability.

SECTION 3: TRADITIONAL PRINCIPLES OF CRIMINAL LIABILITY

Criminal liability in Indian law is fundamentally premised on the coexistence of two essential elements: *actus reus* (the guilty act) and *mens rea*[1] (the guilty mind). This principle is encapsulated in the well-established maxim *actus non facit reum nisi mens sit rea*, which signifies that an act alone does not render a person criminally liable unless it is accompanied by a culpable mental state. The requirement of both physical and mental elements ensures that criminal punishment is imposed only in cases involving moral blameworthiness, thereby distinguishing criminal offences from mere accidents or civil wrongs.

Mens rea, or the mental element of a crime, refers to the state of mind with which an act is committed and plays a crucial role in determining the degree of culpability. It may manifest in various forms, including intention, knowledge, recklessness, and negligence. Intention represents the highest level of culpability, where an individual consciously aims to bring about a specific consequence. Knowledge involves awareness that certain consequences are likely to result from one's actions, while recklessness denotes a conscious disregard of a substantial and unjustifiable risk. Negligence, in contrast, arises from a failure to exercise reasonable care, resulting in harm that could reasonably have been foreseen and avoided.

Complementing the mental element is *actus reus*, which constitutes the external or physical component of a crime. It includes voluntary acts, omissions, or conducts that result in a legally prohibited consequence. For criminal liability to be established, the act must be attributable to the

accused and must bear a causal connection to the harm caused. Thus, traditional criminal law operates on the assumption that both the act and the accompanying mental state originate from a human actor capable of intention, awareness, and control.

However, it is important to note that certain exceptions exist within criminal law where liability may be imposed without proof of *mens rea*, such as in cases of strict liability offences. These exceptions are typically justified on grounds of public welfare and regulatory necessity. Nevertheless, such instances remain limited and do not displace the general principle that criminal liability is primarily founded upon the existence of a guilty mind.

This human-centric framework of criminal liability, while effective in conventional contexts, becomes increasingly problematic when applied to artificial intelligence. Autonomous systems, which are capable of independent decision-making, do not possess intention or consciousness in the legal sense. This raises fundamental questions regarding the applicability of *mens rea* in cases involving AI-generated harm, thereby necessitating a critical re-examination of established legal doctrines.

SECTION 4: THE PROBLEM OF *MENS REA* IN AI-GENERATED HARM

Has a machine ever "intended" to do something? This question represents a fundamental legal dilemma in the age of artificial intelligence, where AI systems increasingly make decisions affecting areas such as safety, privacy, finance, and even warfare. As legal systems attempt to grapple with these developments, a central issue emerges: whether artificial intelligence can possess the *mens rea* necessary for criminal responsibility. The distinction becomes particularly relevant when viewed through the lens of Alan Turing's formulation of whether machines can behave as though they think, rather than actually think. While AI systems may simulate intelligent behavior, such simulation does not equate to the existence of a conscious mental state as required under criminal law [2].

The doctrine of *mens rea* requires that the accused possess intention, knowledge, recklessness, or legally recognized negligence. However, even the most advanced AI systems do not possess awareness, do not experience consequences, and do not understand or appreciate risk in a human

sense. Their actions are guided by algorithmic processes and optimization functions rather than conscious deliberation. Consequently, equating algorithmic outputs with legal intent would amount to a categorical error, as it conflates functional behavior with genuine mental states. This creates a doctrinal gap, as the foundational requirement of criminal liability cannot be meaningfully attributed to a non-human entity lacking consciousness.

This gap raises a critical question: if artificial intelligence cannot possess *mens rea*, then upon whom should liability be imposed when harm occurs?

One possible approach is to attribute liability to human actors within the chain of creation and deployment, such as programmers, operators, or organizations. Where harm results from intentional misuse or negligent design, the mental element may be traced back to these actors. For instance, incidents involving AI systems producing harmful or offensive outputs such as the case of Microsoft's "Tay" chatbot demonstrate how failures in oversight and design can lead to undesirable outcomes. In such cases, courts may rely on principles such as reasonable foreseeability, proximate causation, and control to determine liability [3].

At a broader level, contemporary regulatory approaches, including the European Union's AI governance framework, emphasize that accountability must ultimately remain with human agents [4]. The design choices underlying AI systems such as training data, model architecture, and safety mechanisms play a decisive role in shaping outcomes. Given the inherent unpredictability of autonomous systems, there exists a growing expectation that developers and deploying entities incorporate safeguards, transparency measures, and risk mitigation strategies [5]. Where harm arises from systemic flaws rather than individual acts, responsibility may shift to institutional actors, thereby preventing the problematic attribution of liability to an autonomous system itself.

SECTION 5: MODELS OF LIABILITY FOR AI-GENERATED HARM

The inability to attribute *mens rea* to artificial intelligence necessitates a shift in focus towards alternative models of liability that can accommodate the unique nature of AI-generated harm. In the absence of a culpable mental state on the part of the machine, legal responsibility must be reassigned to human or institutional actors involved in the design, deployment, and use of such

systems. Among the most significant approaches are developer liability, user liability, and strict liability.

Developer liability places responsibility on the creators and designers of AI systems. This approach is particularly relevant where harm arises from defects in design, inadequate training data, or the failure to incorporate necessary safety mechanisms. Given that developers exercise substantial control over the architecture and functioning of AI systems, it may be argued that they bear responsibility where risks are reasonably foreseeable. However, this model faces limitations in cases where AI systems evolve through machine learning processes beyond the direct control of their creators, thereby complicating the attribution of fault.

User liability, on the other hand, focuses on the individuals or entities that deploy or operate AI systems. Liability may arise where users intentionally misuse AI or act negligently in its deployment, such as relying on AI outputs without appropriate verification in high-risk contexts. This model is more applicable in scenarios where human intervention remains significant. However, in cases involving highly autonomous systems, the degree of user control may be minimal, thereby weakening the justification for imposing liability solely on the user.

A third approach is that of **strict liability**, which dispenses with the requirement of proving *mens rea* altogether. Under this model, liability may be imposed irrespective of intention or negligence, particularly in cases involving inherently hazardous activities. By focusing on the occurrence of harm rather than the mental state of the actor, strict liability offers a practical solution to the challenges posed by AI systems [6]. However, its application in the context of artificial intelligence raises concerns regarding fairness, as it may impose liability even in situations where actors have taken reasonable precautions.

Each of these models attempts to address the gap created by the absence of *mens rea* in artificial intelligence. However, none provides a complete solution in isolation. Developer liability may fail in cases of autonomous evolution, user liability may be inadequate where control is limited, and strict liability may lead to unjust outcomes. Consequently, a balanced and hybrid approach,

incorporating elements of these models, may be necessary to effectively regulate AI-generated harm within the framework of criminal law.

SECTION 6: INDIAN LEGAL FRAMEWORK AND REGULATORY GAPS

India currently lacks a comprehensive and unified legal framework specifically governing artificial intelligence, unlike jurisdictions such as the European Union. Instead, the regulation of AI is fragmented across multiple existing statutes and sector-specific regulatory bodies. While these laws address certain aspects of AI deployment, they were not designed to deal with the unique challenges posed by autonomous systems, particularly in the context of criminal liability.

The Information Technology Act, 2000[7] governs digital activities, including cyber security and intermediary liability, and is often invoked in cases involving online harm. Similarly, the Digital Personal Data Protection Act, 2023[8] establishes a framework for the protection of personal data, imposing obligations such as consent, purpose limitation, and data security, which are highly relevant for AI systems that rely on large datasets. Other statutes, including the Consumer Protection Act, 2019[9] and the Copyright Act, 1957, address issues of product liability and intellectual property, respectively. In addition, sectoral regulators such as the Reserve Bank of India, Securities and Exchange Board of India, and Insurance Regulatory and Development Authority of India have issued guidelines governing the use of AI within their respective domains.

Despite this regulatory landscape, significant gaps remain. Existing laws primarily focus on human actors and assume the presence of intention or negligence, which are difficult to establish in the context of autonomous AI systems. For instance, offences under the Bharatiya Nyaya Sanhita, 2023[10] including fraud, defamation, and abetment require proof of *mens rea*, thereby limiting their applicability in cases where harm results from AI-driven decisions. Similarly, while the Motor Vehicles Act, 1988 may address liability in cases involving autonomous vehicles, it does not adequately account for situations where control is exercised by an algorithm rather than a human driver.

Furthermore, issues such as data privacy, algorithmic bias, and lack of transparency exacerbate the problem. AI systems often operate as "black boxes," making it difficult to trace decision-making processes and assign responsibility. While the data protection regime attempts to address concerns relating to consent and misuse of personal data, it does not resolve the deeper question of criminal accountability for harm caused by autonomous systems.

Thus, the Indian legal framework remains reactive and fragmented, addressing isolated aspects of AI without offering a coherent approach to criminal liability. This regulatory gap becomes particularly problematic in cases involving AI-generated harm, where traditional doctrines such as *mens rea* fail to provide clear answers. Consequently, there is a pressing need for a more structured and forward-looking legal framework that can effectively address the challenges posed by artificial intelligence.

SECTION 7: TOWARDS A COHERENT FRAMEWORK FOR AI LIABILITY IN INDIA

The challenges posed by artificial intelligence to traditional principles of criminal liability necessitate a forward-looking and structured legal response in India. Given the inadequacy of existing frameworks, there is a compelling need for the development of a dedicated regulatory regime that specifically addresses the complexities of AI-generated harm.

One significant step in this direction is the proposal for a comprehensive legislative framework governing artificial intelligence. Recent policy discussions in India have indicated the possibility of an Artificial Intelligence (Ethics and Accountability) framework aimed at regulating the development and deployment of AI systems. Such a framework could establish clear standards for accountability, transparency, and ethical compliance. The introduction of an independent oversight body to monitor AI systems and address issues such as algorithmic bias and misuse would further strengthen regulatory control. However, any such mechanism must be carefully balanced to avoid excessive compliance burdens and over-centralization of power.

In the context of criminal liability, the adoption of a **modified strict liability approach** may provide a practical solution. Given that artificial intelligence systems cannot possess *mens rea*, imposing liability based solely on fault may be ineffective.

A calibrated form of strict liability, particularly for high-risk AI applications such as autonomous vehicles, healthcare systems, and financial algorithms, would ensure accountability without the need to establish intent. At the same time, safeguards must be incorporated to prevent unjust outcomes, especially where all reasonable precautions have been taken.

Further, the principle of **human accountability** must remain central to any regulatory framework. Liability should be distributed across the chain of actors involved in the lifecycle of AI systems, including developers, deployers, and organizations. This may be achieved through a system of **graded responsibility**, where liability is assigned based on the degree of control, foreseeability of harm, and level of negligence involved. Such an approach would prevent the inappropriate attribution of liability to autonomous systems while ensuring that responsible human actors are held accountable.

Additionally, regulatory measures should mandate **transparency, auditability, and risk assessment** in AI systems. Developers must be required to incorporate safeguards, conduct impact assessments, and ensure traceability of decision-making processes. This would not only facilitate accountability but also assist courts in determining liability in complex cases involving AI-generated harm.

SECTION 8: CONCLUSION

In conclusion, India must move beyond a fragmented and reactive approach towards a comprehensive and adaptive legal framework for artificial intelligence. By integrating elements of strict liability, human accountability, and regulatory oversight, the legal system can better address the challenges posed by autonomous technologies while maintaining the foundational principles of criminal justice.

The rapid advancement of artificial intelligence has exposed fundamental limitations within traditional principles of criminal liability, particularly the requirement of *mens rea*. As this paper has demonstrated, the human-centric foundation of criminal law is ill-equipped to address harm caused by autonomous systems that lack consciousness, intention, and moral agency. The inability to attribute a guilty mind to artificial intelligence creates a doctrinal gap, challenging the applicability of established legal frameworks in cases of AI-generated harm.

Through an analysis of existing legal principles, models of liability, and the current Indian regulatory landscape, it becomes evident that reliance on conventional approaches is insufficient. While developer liability, user responsibility, and strict liability each offer partial solutions, none adequately resolves the complexities introduced by autonomous decision-making systems. This necessitates a shift towards a more nuanced and adaptive legal framework.

Accordingly, this study supports the adoption of a **hybrid liability model**, wherein human actors such as developers, manufacturers, and users may be held accountable under specific circumstances based on factors such as control, foreseeability, and negligence. At the same time, there is a need to conceptualize new categories of liability tailored to the unique characteristics of artificial intelligence, ensuring that accountability is neither evaded nor unjustly imposed.

Ultimately, the challenge lies in balancing technological innovation with legal accountability. As artificial intelligence continues to evolve, the law must respond proactively by rethinking traditional doctrines and embracing flexible regulatory approaches. Only through such adaptation can the criminal justice system remain effective and relevant in the age of autonomous technologies.

[1] See generally discussions on AI and criminal liability frameworks in legal scholarship.
[2] James Vincent, “Twitter taught Microsoft’s AI chatbot to be a racist,” The Verge (2016).
[3] “AI Chatbot Linked to Teen Suicide Case Under Investigation,” The Economic Times (2025)
[4] OECD Principles on Artificial Intelligence (2019); European Union, Artificial Intelligence Act, 2024.
[5] Rylands v. Fletcher, (1868) LR 3 HL 330.

[6] Information Technology Act, 2000 (India).

[7] Digital Personal Data Protection Act, 2023 (India)

[8] Consumer Protection Act, 2019 (India)

[9] Bharatiya Nyaya Sanhita, 2023 (India)

[i] Ratanlal & Dhirajlal, The Indian Penal Code (LexisNexis, latest Ed.).