



ISSN: 2583-7753

LAWFOYER INTERNATIONAL JOURNAL OF DOCTRINAL LEGAL RESEARCH

[ISSN: 2583-7753]

Volume 4 | Issue 3

2026

DOI: <https://doi.org/10.70183/lijdlr.2026.v04.369>

© 2026 LawFoyer International Journal of Doctrinal Legal Research

Follow this and additional research works at: www.lijdlr.com

Under the Platform of LawFoyer – www.lawfoyer.in

After careful consideration, the editorial board of LawFoyer International Journal of Doctrinal Legal Research has decided to publish this submission as part of the publication.

In case of any suggestions or complaints, kindly contact (info.lijdlr@gmail.com)

To submit your Manuscript for Publication in the LawFoyer International Journal of Doctrinal Legal Research, To submit your Manuscript [Click here](#)

AI AS AN “ACCUSED”: CAN ARTIFICIAL INTELLIGENCE HAVE CRIMINAL LIABILITY?

Gagan Sankhla¹ & Kanan Yadav²

I. ABSTRACT

Artificial intelligence has reshaped contemporary society by assuming cognitive functions that were once performed by human beings alone. As these systems become increasingly autonomous in healthcare, transport, finance and law enforcement, questions of accountability for the harm they occasion have grown more pressing. This paper examines whether criminal liability can attach to an artificial intelligence system, and if not, upon whom it must fall. Criminal liability in the Indian tradition rests upon two elements: the wrongful act (actus reus) and the guilty mind (mens rea). Both elements presuppose a subject capable of consciousness, moral evaluation and choice. An artificial intelligence system possesses none of these attributes, and it is therefore neither a legal person nor a bearer of criminal intention. The paper addresses three questions. First, can an artificial intelligence system be recognised as a person in criminal law? Secondly, if it can, is it capable of satisfying the standards of criminal culpability? Thirdly, do the existing Indian statutes adequately govern harm caused by autonomous systems? The study is doctrinal and comparative. It analyses the Bharatiya Nyaya Sanhita, 2023, the Information Technology Act, 2000 and the Digital Personal Data Protection Act, 2023 alongside the regulatory responses of the United States, the European Union and the United Kingdom. The paper concludes that Indian law does not treat artificial intelligence as a punishable subject, that the existing cyber-law framework is inadequate to address autonomous decision-making, and that liability must be allocated among developers, deployers and users through targeted statutory amendment rather than through the creation of electronic personhood.

¹ B.A. LL.B. (Hons.), 4th Year, University Institute of Legal Studies, Chandigarh University (India).
Email: gaganchaudhary3131@gmail.com

² B.A. LL.B. (Hons.), 4th Year, University Institute of Legal Studies, Chandigarh University (India).
Email: kananyadav29@gmail.com

II. KEYWORDS

Artificial Intelligence, Criminal Liability, Legal Personality, Information Technology Act 2000, Bharatiya Nyaya Sanhita 2023.

III. INTRODUCTION

Artificial intelligence is no longer a prospect of the future. It governs the operation of vehicles that drive themselves, the software that identifies persons of interest to a police officer, and the diagnostic systems upon which clinicians increasingly rely. Decisions that once required human cognition are now delegated to machines. As these systems acquire greater autonomy, a serious legal question arises: who bears responsibility when the outcome is harmful?

Consider a medical services provider whose diagnostic system misclassifies a patient's condition, and the patient dies as a consequence. Can the system itself be punished? The criminal law was not framed for this situation, and its established categories are beginning to strain.

Two elements must be established before a person is convicted of an offence: the commission of a wrongful act (*actus reus*) and the existence of a guilty mind (*mens rea*). An artificial intelligence system has no mind capable of forming an intention. It holds no conception of right and wrong, and it experiences neither remorse nor guilt. It executes instructions and identifies patterns in data. When it produces a harmful outcome, it does so without any awareness of the moral quality of that outcome.

The difficulty is compounded by opacity. Many systems operate as "black boxes" whose developers cannot state with precision how a particular output was reached. Where harm follows, the question of fault becomes difficult to answer. Does responsibility rest with the developer who wrote the code, the vendor who marketed the system, the enterprise that deployed it, or the user who relied upon it? The existing law supplies no settled answer, and victims are left without a clear route to redress.

One body of opinion holds that existing offences should be adapted to accommodate autonomous systems; another contends that new offences should be created for harm caused by artificial intelligence. Any attempt to hold a machine criminally liable

would require a reconsideration of what the law means by a “person”, and conferring that status upon a system that cannot be punished would hollow out the concept rather than extend it.

The research problem addressed in this paper is therefore as follows. Indian criminal law allocates liability on the basis of individual human fault, yet autonomous systems generate harm through processes that cannot readily be traced to the intention or negligence of any identifiable person. The result is an accountability gap. This paper examines whether that gap should be closed by extending criminal personhood to artificial intelligence, or by reallocating responsibility among the human and corporate actors who design, deploy and profit from these systems.

A. Research Objectives

The objectives of this study are as follows:

1. to determine whether an artificial intelligence system can be recognised as a person capable of bearing criminal liability under Indian law, and in particular under the Bharatiya Nyaya Sanhita, 2023;
2. to examine whether such a system can satisfy the mental element required for criminal liability, and to assess the leading doctrinal models by which liability for AI-related harm has been theorised;
3. to assess the adequacy of the existing Indian statutory framework, including the Information Technology Act, 2000, the Digital Personal Data Protection Act, 2023 and the subordinate instruments made under them, in addressing harm caused by autonomous systems;
4. to compare the Indian position with the regulatory responses of the United States, the European Union and the United Kingdom, and to identify the reforms best suited to Indian conditions.

B. Research Questions

The study addresses three questions:

1. Can an artificial intelligence system be legally recognised as a person for the purposes of criminal law?

2. If such recognition were extended, could an artificial intelligence system satisfy the standards of criminal culpability, in particular the requirement of mens rea?
3. Is there existing legislation in India capable of governing harm caused by autonomous artificial intelligence systems, and if not, what form should reform take?

C. Research Methodology

The study is doctrinal and comparative. The doctrinal component analyses primary legal materials to establish the position under Indian law. The primary sources comprise the Bharatiya Nyaya Sanhita, 2023, the Information Technology Act, 2000, the Digital Personal Data Protection Act, 2023 and the Digital Personal Data Protection Rules, 2025, the Consumer Protection Act, 2019, and the decisions of the Supreme Court of India and of the High Courts on corporate criminal liability, criminal negligence and juristic personality.

The comparative component examines Regulation (EU) 2024/1689 (the Artificial Intelligence Act) as amended, Directive (EU) 2024/2853 on liability for defective products, the Automated Vehicles Act 2024 (UK) and the Automated and Electric Vehicles Act 2018 (UK), together with the policy instruments issued by the Ministry of Electronics and Information Technology. Secondary sources comprise peer-reviewed scholarship on the criminal liability of artificial intelligence entities, doctoral research on machine accountability, and institutional reports of the United Nations and the European Commission.

Three jurisdictions have been selected for comparison, and the selection is deliberate. The United States has been included because it allocates liability for AI-related harm through the general law of torts, without a federal statute, and therefore illustrates the capacity and the limits of an unamended common law system. The European Union has been included because it is the only jurisdiction to have enacted a comprehensive, risk-based statutory regime for artificial intelligence, and because its legislative history records a considered rejection of electronic personhood.

The United Kingdom has been included because its legal system shares its common law foundations with India, because it has adopted a sector-specific rather than a horizontal approach, and because it is the only comparator to have enacted a statute that expressly reallocates criminal responsibility for the conduct of an autonomous machine. Jurisdictions that regulate artificial intelligence principally through administrative guidance, and those whose legal systems differ materially from the Indian system, have been excluded.

The study is subject to three limitations. First, it is confined to criminal liability and treats civil and regulatory liability only so far as they bear upon the criminal question. Secondly, it does not undertake empirical or comparative-quantitative analysis of enforcement outcomes. Thirdly, the law is stated as it stood on 19 October 2026; instruments notified after that date, including any revision of the commencement schedule under the Digital Personal Data Protection Act, 2023, have not been considered.

IV. LEGAL PERSONHOOD AND ARTIFICIAL INTELLIGENCE

The threshold question is whether an artificial intelligence system can be a legal subject in criminal law. Orthodox criminal theory confines culpability to those who are conscious, capable of moral reasoning and able to act with intention. As autonomous systems assume decision-making roles in medicine, transport and finance, that confinement has been questioned.

A. The Concept of Legal Personhood

Legal personality is not restricted to human beings. In law, a person is any individual, organisation or body recognised as capable of holding rights and bearing duties. The definition is functional rather than biological: it responds to what society requires, not to natural characteristics. A legal person possesses rights, in that it may own property, contract and sue; duties, in that it is bound to observe the law and answer for breach; recognition, in that the law acknowledges its status; and capacity, in that it may act

directly or through agents.³ Legal personhood is accordingly a legal fiction adopted for practical purposes, and not the recognition of a moral reality.⁴

Criminal law recognises two classes of person. Natural persons are human beings, and they bear the full measure of criminal responsibility because they are conscious, self-aware, capable of moral judgment and able to apprehend the significance of punishment. Juristic persons are entities to which the law attributes personality for defined purposes. The capacity of the latter to commit offences has been established by a line of Indian authority to which this paper now turns.

B. Corporate Criminal Liability in Indian Law

The starting point is *State of Maharashtra v Syndicate Transport Co (P) Ltd*, in which the Bombay High Court recognised corporate criminal liability in principle while excluding offences that only a natural person can commit and offences punishable solely with imprisonment.⁵ That exclusion no longer represents the law. In *Standard Chartered Bank v Directorate of Enforcement*, a Constitution Bench of the Supreme Court held that a company may be prosecuted and convicted even where the offence carries a mandatory sentence of imprisonment, the court imposing the fine alone.⁶

The Court thereafter addressed the attribution of a mental state to a corporation. In *Iridium India Telecom Ltd v Motorola Inc*, it held that a company may be prosecuted for offences requiring mens rea, and that the intention of those who control its affairs to such a degree that they constitute its “directing mind and will” is attributed to the company on the alter ego principle.⁷ In *Sunil Bharti Mittal v CBI*, the Court clarified the direction in which attribution operates: the criminal intent of the alter ego is imputed to the company, but the converse does not follow, and a director cannot be made vicariously liable for an offence of the company unless there is sufficient

³ Athina Sachoulidou, “AI Systems and Criminal Liability: A Call for Action” (2024) 11(1) Oslo Law Review 1, 2-4, DOI 10.18261/olr.11.1.3.

⁴ “Beyond Personhood: The Evolution of Legal Personhood and AI Governance” (2025) TechReg, 7 September 2025, available at <https://techreg.org/article/view/22555>.

⁵ *State of Maharashtra v Syndicate Transport Co (P) Ltd* AIR 1964 Bom 195.

⁶ *Standard Chartered Bank v Directorate of Enforcement* (2005) 4 SCC 530.

⁷ *Iridium India Telecom Ltd v Motorola Inc* (2011) 1 SCC 74.

evidence of his active role coupled with criminal intent, or a statutory provision expressly imposing such liability.⁸

Taken together, these authorities establish that Indian criminal law is capable of reaching a non-human entity, but only because that entity acts through identifiable human beings whose minds the law may attribute to it. The sharper question for the present study is therefore not whether a machine may be punished, but whether an artificial intelligence system can ever be treated as the alter ego of the enterprise that deploys it, so that the outputs of the system are imputed to the company as its own mental state. That question is examined in Part 8.6.

C. Non-Human Juristic Persons in Indian Law

Indian law has conferred juristic personality upon entities other than corporations, and those precedents are instructive. In *Yogendra Nath Naskar v Commissioner of Income Tax*, the Supreme Court held that a Hindu idol is a juristic person capable of holding property and of being taxed, its affairs being conducted by a shebait.⁹ In *Shiromani Gurdwara Parbandhak Committee v Som Nath Dass*, the Court held that the Guru Granth Sahib is a juristic person, while emphasising that legal personality is conferred by the law to serve a practical purpose and does not depend upon the capacity to act.¹⁰ In *Mohd Salim v State of Uttarakhand*, the Uttarakhand High Court declared the rivers Ganga and Yamuna to be legal persons and appointed named officers as their guardians in loco parentis;¹¹ that order was stayed by the Supreme Court on 7 July 2017.¹²

Two features are common to each of these entities. Each acts exclusively through a human guardian, trustee or committee, and none is capable of committing an offence or of being punished. Indian law has therefore conferred personality where it was necessary to secure the holding of property or the protection of an interest, and never

⁸ *Sunil Bharti Mittal v CBI* (2015) 4 SCC 609.

⁹ *Yogendra Nath Naskar v Commissioner of Income Tax* (1969) 1 SCC 555.

¹⁰ *Shiromani Gurdwara Parbandhak Committee v Som Nath Dass* (2000) 4 SCC 146.

¹¹ *Mohd Salim v State of Uttarakhand* (Uttarakhand High Court, 20 March 2017), Writ Petition (PIL) No 126 of 2014.

¹² *State of Uttarakhand v Mohd Salim*, SLP (C) No 16879 of 2017, order dated 7 July 2017 (Supreme Court of India).

in order to render the entity criminally answerable. That distinction supports the thesis of this paper: recognition as a legal person for a limited civil purpose does not entail, and has never entailed, capacity to bear criminal liability.

D. Electronic Personality in International Debate

The question has also been raised outside India. In its Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, the European Parliament invited the Commission to consider creating “a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons”.¹³ The proposal attracted an open letter signed by more than 150 experts in robotics, artificial intelligence, law and ethics, who objected that electronic personhood rested upon an overvaluation of the actual capabilities of such systems and would permit manufacturers to displace their own responsibility.¹⁴

The Commission did not act upon the proposal, and the Artificial Intelligence Act subsequently enacted allocates obligations to providers and deployers without conferring personality upon the system. The international trajectory has therefore been away from electronic personhood and towards the allocation of responsibility along the value chain.

E. Statutory Definitions under the Bharatiya Nyaya Sanhita, 2023

The Indian statutory position is governed by the definitions in Chapter I of the Bharatiya Nyaya Sanhita, 2023. Section 2(26) provides that “person” includes “any company or association or body of persons, whether incorporated or not”.¹⁵ The definition is inclusive, but each limb of it presupposes either an incorporated body or an aggregation of persons. An artificial intelligence system is neither. It is not a company, because it is not incorporated under any law; it is not an association or body

¹³ European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) [2018] OJ C252/239, para 59(f).

¹⁴ Open Letter to the European Commission on Artificial Intelligence and Robotics (2018), available at <http://www.robotics-openletter.eu>.

¹⁵ Bharatiya Nyaya Sanhita 2023, s 2(26). The provision corresponds to s 11 of the Indian Penal Code, 1860, which the Bombay High Court applied in *Syndicate Transport* (n 4).

of persons, because it is not composed of persons at all. It is an artefact owned and operated by persons, and it falls on the object side of the distinction the section draws.

Section 2(33) provides that a person causes an effect “voluntarily” when he causes it by means whereby he intended to cause it, or by means which, at the time of employing those means, he knew or had reason to believe to be likely to cause it.¹⁶ Section 2(29) in turn provides that a person has “reason to believe” a thing if he has sufficient cause to believe that thing, and not otherwise.¹⁷ Both definitions are expressed in terms of intention, knowledge and belief, which are states of mind. An artificial intelligence system holds no beliefs and forms no intentions; it therefore cannot act voluntarily within the meaning of the Sanhita, and the voluntariness upon which the general part of the criminal law is constructed is unavailable in its case.

It follows that the extension of criminal personhood to artificial intelligence could not be achieved by judicial construction. It would require Parliament to amend section 2(26) so as to include a defined class of autonomous systems, and to make consequential provision for the attribution of a mental state under section 2(33) and for a sanction capable of operating against a non-sentient subject. No such amendment has been proposed, and for the reasons set out below none is warranted.

F. The Inapplicability of Punishment

A further reason why artificial intelligence cannot be a person in criminal law is that punishment would be incapable of achieving its purpose. Criminal sanction presupposes a subject capable of suffering deprivation. Imprisonment operates upon liberty; a fine operates upon assets; probation operates upon conduct. A corporation, though it cannot be imprisoned, holds property and reputation and may therefore be fined and disqualified.

None of the established modes of punishment translates to an artificial intelligence system. There is no analogue to imprisonment, and the deletion of software is not a criminal sanction but a technical measure. Fines presuppose a patrimony, which the

¹⁶ Bharatiya Nyaya Sanhita 2023, s 2(33).

¹⁷ Bharatiya Nyaya Sanhita 2023, s 2(29).

system does not possess. Probation presupposes a capacity to comply, which it does not possess either. A machine may be deactivated, modified or retrained, but each of these is an engineering decision rather than a penal one. Criminal punishment is also expressive: it communicates censure to an agent capable of receiving it. A system that cannot experience shame, remorse or deterrence cannot receive that communication.¹⁸ Where the entity cannot be punished, the designation of that entity as an accused serves no legal purpose.

G. Proposals for Limited Personhood

Several scholars have nonetheless argued for a restricted form of personality as a means of closing the accountability gap, the difficulty being that where an autonomous decision produces harm, no single human actor may be identifiable as the author of that decision.¹⁹ Three models recur in the literature. Electronic personhood would confer limited rights and duties upon the system for defined purposes. Functional personhood would recognise the system only within specified domains, such as healthcare, transport or finance. A hybrid model would treat the system as a bearer of duties for the purpose of civil compensation while leaving criminal responsibility with the human actors behind it.²⁰²¹

Each model addresses compensation rather than culpability. None supplies an answer to the objection developed above, namely that a subject incapable of suffering punishment cannot be the object of a criminal sanction. The hybrid model is the most defensible of the three, precisely because it confines the innovation to the civil sphere and leaves the criminal law intact. This paper adopts that position.

¹⁸ Kamil Mamak, "AI Personhood, Criminal Law, and Punishment" in Oxford Intersections: AI in Society (Oxford University Press 2025), DOI 10.1093/9780198945215.003.0043.

¹⁹ Sachoulidou (n 2) 5-7; Alice Giannini, Criminal Behavior and Accountability of Artificial Intelligence Systems (PhD thesis, Maastricht University and University of Florence, Eleven Publishers 2023) 141-168.

²⁰ Andreas Nanos, "Criminal Liability of Artificial Intelligence" (SSRN, 6 November 2023), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4623126.

²¹ Beyond Personhood (n 3); Giannini (n 18) 169-192.

H. Mens Rea and Artificial Intelligence

Mens rea is a foundational element of criminal liability. The law considers not only the effect that the accused produced upon another, but the state of mind in which he produced it. It is this element that distinguishes an accident, or an honest mistake, from a crime. The concept becomes difficult to apply where the immediate cause of harm is a machine.

1. Mens Rea and Human Consciousness

In classical criminal theory, mens rea is bound up with consciousness. A person who knows the nature of his conduct, who is able to contemplate its consequences, and who nonetheless chooses to proceed, is held responsible for the result. Punishment is justified because the person could have acted otherwise and did not. Moral agency is therefore the premise of the mental element, and not merely a desirable feature of it.

2. Degrees of Culpability under the Bharatiya Nyaya Sanhita, 2023

Indian criminal law recognises graduated mental states. Intention denotes conduct directed towards a desired result; knowledge denotes awareness that a result is practically certain to follow; recklessness denotes the conscious acceptance of a known risk; and negligence denotes a failure to exercise the care that a reasonable person would have exercised. The Bharatiya Nyaya Sanhita gives statutory form to the first three through the definitions of “voluntarily” in section 2(33) and “reason to believe” in section 2(29),²² and to the fourth through the offences of rash or negligent conduct, principally section 106.²³ Where this paper refers to the Indian Penal Code, 1860, it does so only for historical context, the Sanhita having come into force on 1 July 2024.

The Supreme Court has repeatedly affirmed that these states must be proved and may not be presumed. In *Nathulal v State of Madhya Pradesh* the Court held that mens rea is an essential ingredient of a criminal offence unless it is excluded, expressly or by necessary implication, by the statute creating it.²⁴ In *Jacob Mathew v State of Punjab*

²² Bharatiya Nyaya Sanhita 2023, ss 2(29) and 2(33).

²³ Bharatiya Nyaya Sanhita 2023, s 106.

²⁴ *Nathulal v State of Madhya Pradesh* AIR 1966 SC 43.

the Court held that the negligence required for criminal liability must be of so high a degree as to be described as gross, or as amounting to recklessness.²⁵

None of these mental states can be predicated of an artificial intelligence system. The system does not pursue a purpose, because it optimises an objective function supplied by a designer. It does not know, because it holds no propositional attitudes. It does not consciously accept a risk, because it has no awareness of risk as such. It does not fail to take care, because the standard of care is a normative expectation addressed to an agent. The system may produce outputs that correspond in form to each of these categories, but the correspondence is behavioural and not mental.

3. Models of Criminal Liability for Artificial Intelligence Entities

The leading doctrinal framework for this question is that proposed by Gabriel Hallevy, who identified three models by which criminal liability might be imposed in respect of an artificial intelligence entity.²⁶ Each merits consideration against Indian doctrine.

The first is the perpetration-via-another model, under which the system is treated as an innocent agent and liability attaches to the person who used it to bring about the offence. This model maps directly onto Indian law. A person who deploys a system to cause harm is the principal offender, and the analysis under sections 3(5) and 49 of the Bharatiya Nyaya Sanhita, relating to common intention and abetment, proceeds as it would where the instrument employed was any other inanimate object. The model is doctrinally sound and requires no legislative change.

The second is the natural-probable-consequence model, under which a programmer or user is liable for an offence that was a natural and probable consequence of the system's ordinary operation, even though it was not intended. This model corresponds to the Indian treatment of rashness and negligence, and to the standard of gross negligence established in *Jacob Mathew*. It is the model most likely to govern

²⁵ *Jacob Mathew v State of Punjab* (2005) 6 SCC 1 [48] ("Conclusions summed up", conclusion (6)); also reported at AIR 2005 SC 3180.

²⁶ Gabriel Hallevy, "The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control" (2010) 4(2) *Akron Intellectual Property Journal*, art 1, available at <https://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1>.

the typical case of AI-related harm in India, in which a developer or deployer released or continued to operate a system whose propensity to cause injury was foreseeable.

The third is the direct liability model, under which the system itself is treated as satisfying both *actus reus* and *mens rea*. This paper rejects that model. It rests upon an equivalence between the functional states of a machine and the mental states of a person that the criminal law does not support, and it cannot survive the objection developed in Part 5.6: a subject that cannot experience censure or deprivation cannot be punished, and a conviction that produces no penal consequence is a declaration rather than a sentence. The model also carries a practical danger. By locating fault in the machine, it would furnish developers and deployers with a defence where the law should instead be tightening their obligations.

The first two models therefore account for the Indian position without any extension of personhood, and the proposals for limited personhood considered in Part 5.7 are best understood as responses to the compensation gap that the third model fails to close.

4. Autonomy, Free Will and Machine Decision-Making

It is sometimes argued that advanced systems are autonomous, and that autonomy suffices to ground culpability. The argument trades upon an ambiguity in the word. What appears to be a decision is the execution of a program, the inference of a model and the processing of data. The machine exercises no free will. Its behaviour may exceed the expectations of its designers, but unpredictability is not freedom, and a process may be opaque without being volitional.

5. Artificial Intelligence as a Source of Harm, not a Moral Agent

The sounder characterisation is that an artificial intelligence system is a source of harm and not a moral offender. The two are not equivalent. A system may be causally implicated in an injurious outcome without supplying the mental element of an offence. The question the law should ask is whether a human being was negligent, reckless or otherwise blameworthy in the design, testing, deployment or supervision

of the system. That approach preserves the structure of criminal liability while leaving the victim with a remedy.

The criminal law accordingly searches for the human link. A developer who designs a system for an unlawful purpose may be liable as a principal. An operator who deploys it carelessly may be liable in negligence. An enterprise that places it in service without adequate safety assessment may be liable through its controlling officers. In each case the system is the instrument, and the inquiry remains directed at the conduct of the person who made, supplied or used it. Where artificial intelligence is employed deliberately for fraud, unlawful surveillance, cybercrime or violence, the user is answerable in the ordinary way, and the autonomy of the instrument is immaterial.

That approach is consistent with the international ethical consensus. The UNESCO Recommendation on the Ethics of Artificial Intelligence provides that Member States should ensure that it is always possible to attribute ethical and legal responsibility for the conduct of AI systems to physical persons or to existing legal entities, that an AI system can never displace ultimate human responsibility and accountability, and that ultimate responsibility must always lie with natural or legal persons, AI systems not being given legal personality themselves.²⁷ The instrument therefore locates fault in the human actor rather than in the machine, and it does so in terms that align with the position of Indian law.

V. THE RESPONSIBILITY GAP AND THE BLACK BOX PROBLEM

A. The Responsibility Gap

The responsibility gap arises where damage, injury or loss results from the operation of an artificial intelligence system and no human being can be held accountable. The criminal law assumes that harm proceeds from human conduct, whereas an autonomous system may generate an outcome that cannot be traced to the act or omission of any identifiable individual. The gap is most apparent in autonomous

²⁷ UNESCO, Recommendation on the Ethics of Artificial Intelligence (adopted 23 November 2021) paras 35, 36 and 68.

vehicles, diagnostic systems, algorithmic trading, predictive policing and industrial robotics.

Three elements of the ordinary analysis become difficult to establish. *Mens rea* is difficult to prove because the human actor may neither have intended the harm nor foreseen it. Causation is weakened because the proximate cause of the harm is the operation of the system. Negligence is difficult to establish where the outcome was not reasonably foreseeable at the time of design or deployment.

B. The Black Box Problem

These difficulties are aggravated by opacity. A black box system receives an input and produces an output without exposing the intermediate reasoning, and the problem is most pronounced in deep learning architectures whose decision paths traverse many layers of parameters. A diagnostic system may recommend an inappropriate course of treatment, and an autonomous vehicle may execute a hazardous manoeuvre, without anyone, including the developers, being able to reconstruct the reasoning that produced the result.

The consequence for the criminal law is direct. Where the decision path cannot be reconstructed, it is difficult to prove that any human being knew of a risk, or ought to have known of it. Each participant can plausibly disclaim responsibility: the developer may say that the system was designed only to learn from data; the deployer may say that the output was unforeseeable; the user may say that he relied upon the system as it was supplied. Harm is evident, and fault is not.

C. Transparency, Explainability and Accountability

Transparency denotes the capacity to observe the operation of a system, including the data upon which it was trained and the manner in which that data is used. Explainability denotes the capacity of the system to furnish a reason for its output in terms intelligible to a human being. Both are legally significant because the criminal law requires proof of intention, knowledge or negligence, and an opaque system defeats that proof. Accountability is therefore contingent upon transparency, and a

regulatory requirement of explainability in high-risk applications is a precondition of effective criminal enforcement rather than a merely technical standard.

VI. THE INDIAN LEGAL FRAMEWORK

Indian criminal law continues to allocate liability by reference to human purpose and human direction. The Bharatiya Nyaya Sanhita, 2023 and the Information Technology Act, 2000 address human actors rather than autonomous systems, and neither makes specific provision for harm occasioned by a system acting without immediate human direction. This Part consolidates the Indian position, which is cross-referred in the comparative and reform discussion that follows.

A. The Bharatiya Nyaya Sanhita, 2023

The Sanhita proceeds upon the premise that the subject of the criminal law is capable of thought, choice and action. Its general part is constructed around the definitions examined in Part 5.5, and its offences of rash and negligent conduct presuppose a standard of care addressed to a human agent. No provision of the Sanhita contemplates an autonomous system as either the author of an offence or a distinct category of instrument, and the Government has indicated that amendment, rather than reinterpretation, is the appropriate route to addressing AI-generated harm.

B. The Information Technology Act, 2000

The Information Technology Act addresses artificial intelligence only indirectly, and as a tool in human hands. Section 43 provides for civil compensation where a computer resource is damaged, and treats the system as an instrument rather than as a liable entity. Section 66 makes dishonest or fraudulent conduct in relation to the acts described in section 43 a criminal offence, and expressly requires dishonesty or fraud, which an artificial intelligence system cannot possess. Section 43A imposes liability upon a body corporate that is negligent in implementing reasonable security practices in relation to sensitive personal data, and so proceeds upon the premise that the human controller, and not the system, is answerable.

Section 85 provides that where a contravention is committed by a company, every person who at the time was in charge of and responsible to the company for the

conduct of its business is liable along with the company, subject to the defence of absence of knowledge and the exercise of due diligence. The provision is a statutory vicarious liability of the kind contemplated in *Sunil Bharti Mittal*, and is considered further in Part 8.6. An artificial intelligence system is not a company or body corporate and possesses no legal personality, and so falls outside the section.

C. The Digital Personal Data Protection Act, 2023 and the Rules of 2025

The Digital Personal Data Protection Act, 2023, together with the Digital Personal Data Protection Rules, 2025, now governs the processing of digital personal data, including processing carried out by automated systems. The Act imposes obligations of purpose limitation, notice, security safeguards and breach notification upon the data fiduciary, and enforces them through financial penalties imposed by the Data Protection Board rather than through criminal sanction.

Its significance for the present inquiry is twofold. First, it confirms the legislative preference for allocating duties to the human or corporate entity that determines the purpose and means of processing, and not to the processing system. Secondly, by providing for the omission of section 43A of the Information Technology Act, it consolidates the regulation of data security within a single administrative regime, which narrows the civil avenues available where an AI system causes harm through the misuse of personal data.

D. The India AI Governance Guidelines, 2025

On 5 November 2025 the Ministry of Electronics and Information Technology released the India AI Governance Guidelines.²⁸ The Guidelines decline to recommend a standalone artificial intelligence statute at this stage. They proceed instead upon the view that the existing framework of information technology, data protection, consumer, competition and criminal law already reaches developers and deployers, and they call for a comprehensive review to identify gaps, together with targeted amendment of existing legislation.

²⁸ Ministry of Electronics and Information Technology, India AI Governance Guidelines (5 November 2025).

They identify the Information Technology Act as requiring amendment so that the classification of AI systems is made clear, question whether the definitions of “intermediary” and “publisher” and the safe harbour in section 79 are apt for systems that generate or modify content, and contemplate amendment of the Bharatiya Nyaya Sanhita in relation to AI-generated content. On accountability, the Guidelines provide that responsibility is to be assigned according to the function performed by each actor in the value chain and the risk of harm arising from it, supported by transparency reporting, auditing and grievance redressal, and that sectoral regulators are to prescribe specific requirements within their respective domains.

The Government’s stated position bears directly upon the reform proposals advanced in this paper, and it must be engaged with rather than passed over. Two observations follow. First, the Guidelines’ preference for amending existing law weakens the case for a dedicated criminal statute on artificial intelligence, and this paper accordingly recommends amendment of the Sanhita and of the Information Technology Act in preference to a new penal code for autonomous systems. Secondly, a functional allocation of liability across the value chain is an administrative technique, and it does not translate directly into the criminal law, which attributes fault to a specific person in respect of a specific act or omission. The allocation model is therefore best deployed to define the duties whose breach supplies the predicate for criminal negligence, rather than to create criminal liability of its own force.

E. The Intermediary Guidelines Amendment Rules, 2026

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 were notified on 10 February 2026 and came into force on 20 February 2026.²⁹ The Rules address synthetically generated information. They require such information to be labelled and to carry embedded metadata identifying it as synthetically generated, impose verification and declaration obligations upon significant social media intermediaries, and shorten the period within which specified unlawful content must be removed.

²⁹ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2026, notified 10 February 2026, in force 20 February 2026.

The Rules are the first Indian instrument to regulate generative output directly, and they confirm the legislative technique identified in the Guidelines: obligations are placed upon identifiable intermediaries and deployers, and no obligation is placed upon the system itself. Their relevance to criminal liability is that a documented failure to comply supplies evidence of the knowledge or negligence that the criminal law requires.

F. Corporate Attribution and the Liability of Providers

Enterprises, rather than individual employees, develop, train, procure and deploy artificial intelligence systems, and the question of corporate criminal liability is therefore central. The analysis is governed by the attribution doctrine set out in Part 5.2. Under *Iridium*, the mental state of those who constitute the directing mind and will of the company is attributed to the company; under *Sunil Bharti Mittal*, that attribution does not run in reverse, and a director is liable only upon evidence of his active role coupled with criminal intent or under an express statutory provision such as section 85 of the Information Technology Act.

Applying that doctrine to artificial intelligence discloses a structural difficulty. Where the harmful decision originates in an algorithm and not in a senior officer, there may be no directing mind whose intention can be attributed to the company. The output is the product of training data, model architecture and deployment conditions, each the responsibility of a different function within the enterprise, and no one of which amounts to a decision of the controlling officers.

Diamantis has argued that the law should respond by treating the algorithmic systems through which a corporation processes information as part of an extended corporate mind, so that what the system knew is treated as what the company knew.³⁰ That argument is attractive in that it prevents a company from insulating itself by delegating judgment to a machine. It is also doctrinally unorthodox, because it attributes to the corporation a mental state that no human officer ever held.

³⁰ Mihailis E. Diamantis, "The Extended Corporate Mind: When Corporations Use AI to Break the Law" (2020) 98 North Carolina Law Review 893.

The preferable course under Indian law is narrower. Rather than deeming the company to possess the system's knowledge, the court should examine the knowledge actually held by the controlling officers as to the system's risks: what the internal records disclosed, what the testing showed, what warnings were received and disregarded, and what was done once a risk was identified. Where those officers knew of an unsafe system and authorised its deployment, their intention is attributable to the company on ordinary Iridium principles, and no extension of doctrine is required.

Two matters of principle should be stated plainly. The burden of proof lies upon the prosecution, and not upon the accused company, to establish beyond reasonable doubt that the company, through its directing mind and will, knew of the risk or was grossly negligent in failing to apprehend it. Further, the fact that an outcome was generated by an autonomous system is not a defence. A company cannot discharge its responsibility by asserting that the system acted on its own, because the decisions to select the model, to train it on particular data, to disregard a warning and to omit monitoring are all corporate decisions capable of attribution.

VII. PRODUCT LIABILITY AND CRIMINAL NEGLIGENCE

Where an artificial intelligence system causes harm, two distinct bodies of law may be engaged. Product liability is a civil remedy directed at compensation where a defective product causes damage. Criminal liability is directed at censure and punishment where the conduct is sufficiently culpable. The central question is whether the developer of a system that causes harm without any purpose of doing so can be criminally negligent, and the answer turns upon the degree of the departure from the standard of care.

A. Product Liability as a Civil Remedy

Product liability attaches where a defect in a product causes damage. In the case of an artificial intelligence system, the defect may lie in the design, in a defective component, or in the failure to warn users of a known risk. The claimant must establish that the defect caused the injury, death or loss; no wrongful intent need be proved. It is therefore the more accessible route in AI cases, and its object is to restore the victim rather than to censure the defendant. In India the remedy is available under

Chapter VI of the Consumer Protection Act, 2019, which prescribes the liability of product manufacturers, service providers and sellers.³¹

B. When Criminal Negligence Arises

Criminal liability may arise in three situations. The first is where the developer is aware of a defect and nonetheless releases or continues to supply the system. The second is where the design is careless to a degree that would have been apparent to a reasonable person in that position. The third is where the developer fails to monitor a system knowing that it is capable of causing harm.

Thus a company that knows that its vehicle system performs unreliably in heavy rain and continues to sell vehicles fitted with that system may be reckless. A clinical system released without adequate validation may evidence gross negligence. By contrast, where a system produces a harmful output that was genuinely unforeseeable, the case for criminal liability is correspondingly weaker, and the civil remedy is the appropriate one.

C. The Indian Standard: Jacob Mathew and Section 106

The distinction between ordinary and gross negligence is well established in India. In *Jacob Mathew v State of Punjab*, a three-Judge Bench of the Supreme Court held on 5 August 2005 that the negligence required to found criminal liability must be of so high a degree as to be described as gross, or as amounting to recklessness, and that a mere error of judgment or an act of ordinary negligence is not sufficient.³² The Court framed that principle in the specific context of medical professionals, holding that a doctor should not be exposed to prosecution for an error of judgment, and prescribing procedural safeguards before a criminal case is registered against a medical practitioner.

The principle is capable of extension to the developers and deployers of artificial intelligence systems, and the analogy is close. Both act under conditions of irreducible uncertainty; both exercise specialised judgment that a court is not equipped to second-

³¹ Consumer Protection Act 2019, ss 2(34) and 84-87.

³² *Jacob Mathew* (n 24) [48]. The judgment is reported at (2005) 6 SCC 1 and AIR 2005 SC 3180.

guess with hindsight; and in both cases the imposition of criminal liability for ordinary error would deter the very activity the law wishes to encourage. The extension is also limited in the same way. The protection of the Jacob Mathew standard is available only to a defendant who adhered to the practices accepted as proper by a responsible body of professional opinion, and a developer who disregarded established testing, validation or monitoring practice falls outside it.

The statutory provision has since been recast. Section 304A of the Indian Penal Code, 1860 has been replaced by section 106 of the Bharatiya Nyaya Sanhita, 2023, which raises the maximum punishment for causing death by a rash or negligent act to five years. Section 106(1) now carries a proviso under which a registered medical practitioner who causes death by a rash or negligent act in the course of a medical procedure is punishable with imprisonment of up to two years and a fine.³³ The proviso is significant for the present inquiry in two respects.

It confirms that Parliament is prepared to graduate criminal exposure according to the professional context in which a risk is taken, which supplies a legislative model for the treatment of AI developers. It also underlines the limits of the analogy: the statutory concession is confined to registered practitioners, and a developer whose system contributes to a patient's death enjoys no corresponding protection.

As at 19 October 2026, and to the authors' knowledge, no reported Indian decision has assigned criminal liability for harm caused by an artificial intelligence system. That statement is based upon searches of SCC Online, Manupatra, Indian Kanoon and the judgment portal of the Supreme Court of India. In the absence of authority, the general principles of negligence, knowledge and recklessness continue to govern, and the law has yet to settle the circumstances in which the developer of such a system is criminally answerable.

³³ Bharatiya Nyaya Sanhita 2023, s 106(1), first proviso, which defines a registered medical practitioner by reference to the National Medical Commission Act 2019 and the National Medical Register or a State Medical Register.

VIII. COMPARATIVE LEGAL APPROACHES

States have responded to the legal risks of artificial intelligence in materially different ways. Most frameworks continue to rest upon human liability, while a few are developing dedicated regimes for harm caused by autonomous systems. The comparative question is where responsibility is placed: upon the developer or manufacturer, upon the operator or deploying enterprise, or upon a no-fault regime that relieves the victim of the burden of proving fault.

A. United States

The United States allocates liability for AI-related harm through the existing law of torts, principally product liability, negligence and vicarious liability, and has enacted no comprehensive federal statute on the subject.³⁴ Courts and legislatures accordingly address AI harm through categories developed for conventional products. The approach is incremental and flexible, and it permits the law to respond case by case. It also leaves gaps where harm arises from a system that does not fit the established classifications, in particular where the alleged defect lies in a model that continues to learn after sale.

B. European Union

The European Union regulates artificial intelligence through two instruments of distinct character. Regulation (EU) 2024/1689, the Artificial Intelligence Act, is a regulatory statute that imposes obligations upon providers and deployers and enforces them through administrative fines under Article 99; it confers no right to compensation upon a victim.³⁵ Its obligations for high-risk systems apply from 2 December 2027 for stand-alone systems listed in Annex III, and from 2 August 2028

³⁴ Law Commission (England and Wales), *Artificial Intelligence and the Law: a discussion paper* (31 July 2025) paras 2.14-2.41, surveying the common law treatment of AI-related harm; see also *Restatement (Third) of Torts: Products Liability* (American Law Institute 1998) ss 1-2.

³⁵ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [2024] OJ L2024/1689, arts 99 and 113.

for systems embedded as safety components in products already covered by sectoral legislation, those dates having been deferred by Regulation (EU) 2026/1744.³⁶

Compensation is instead provided under Directive (EU) 2024/2853 on liability for defective products, a no-fault regime that expressly covers software and artificial intelligence systems and applies to products placed on the market after 9 December 2026.³⁷ The proposed Artificial Intelligence Liability Directive, which would have adapted non-contractual fault-based liability rules to artificial intelligence, was announced for withdrawal in Annex IV to the Commission Work Programme 2025 and formally withdrawn by notice published in the Official Journal on 6 October 2025.³⁸

Neither instrument renders an artificial intelligence system criminally liable. Criminal law remains a competence of the Member States, and responsibility continues to rest with the natural and legal persons behind the system, which supports the argument advanced in this paper.

C. United Kingdom

The United Kingdom has adopted a sector-specific approach, regulating artificial intelligence through existing regulators in domains such as healthcare, financial services and transport rather than through a single horizontal statute.³⁹ That approach has not prevented targeted legislation where a sector required it, and the most significant example for present purposes concerns automated vehicles, which supply the recurring illustration in this paper.

The Automated Vehicles Act 2024 received Royal Assent on 20 May 2024.⁴⁰ The Act establishes an authorisation regime under which every authorised automated vehicle

³⁶ Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulation (EU) 2024/1689, Regulation (EU) 2018/1139 and Regulation (EU) 2023/1230 [2026] OJ L, 24 July 2026, in force 27 July 2026.

³⁷ Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L2024/2853, arts 4 and 22.

³⁸ Commission Work Programme 2025, COM(2025) 45 final, Annex IV; Withdrawal of Commission proposals [2025] OJ C/2025/5423, 6 October 2025.

³⁹ Law Commission (n 34) paras 1.9-1.23.

⁴⁰ Automated Vehicles Act 2024 (c 10).

must have a designated authorised self-driving entity, ordinarily the manufacturer, which bears continuing responsibility for ensuring that the vehicle satisfies the self-driving test, and it provides for the licensing of no-user-in-charge operators for vehicles operated without a driver in the vehicle.⁴¹ Critically, where an authorised automated feature is engaged, the user-in-charge is immune from criminal and civil responsibility for the manner in which the vehicle drives, his duties reviving on a transition demand.⁴²

The Act also creates new offences, including the misuse of restricted terms such as “self-driving” in relation to unauthorised vehicles, the making of communications likely to confuse end users as to a vehicle’s autonomous capability, and the provision of false or misleading information in support of an authorisation application.⁴³ Civil compensation continues to be governed by the Automated and Electric Vehicles Act 2018, under which the insurer is primarily liable to the injured party where an accident is caused by an automated vehicle driving itself.⁴⁴

The United Kingdom model is instructive because it demonstrates, in statutory form, precisely the reallocation for which this paper argues. Faced with a machine that drives itself, Parliament did not confer personality upon the vehicle. It identified the corporate entities best placed to control the risk, imposed authorisation and licensing obligations upon them, relieved the human occupant of responsibility for conduct he no longer controlled, and created offences directed at the corporate conduct that generates the risk. Criminal responsibility was moved along the value chain and not transferred to the machine.

D. Other Jurisdictions

Elsewhere the position remains unsettled. Several jurisdictions continue to rely upon human fault and corporate personality, while others are developing stricter liability regimes or compulsory insurance. Comparative scholarship proposes variously that the developer should answer for defects in design, that the operator should answer

⁴¹ Automated Vehicles Act 2024, ss 6, 7 and 12.

⁴² Automated Vehicles Act 2024, ss 47-49.

⁴³ Automated Vehicles Act 2024, ss 24-27, 78 and 79.

⁴⁴ Automated and Electric Vehicles Act 2018 (c 18), s 2.

for the conditions of deployment, and that a no-fault fund should answer where neither can be identified.⁴⁵ The shared difficulty is how to secure accountability without suppressing innovation, and the comparative material suggests not a single correct answer but a set of institutional choices from which India may select.

IX. AUTONOMOUS WEAPONS SYSTEMS AND INTERNATIONAL LAW

Autonomous weapons present the accountability problem in its most acute form. A system capable of selecting and engaging a target without human intervention raises a legal and moral difficulty, because where such a system causes unlawful harm it may not be apparent who is answerable.

A. Autonomous Weapons and the Accountability Gap

Autonomous weapons identify, track and engage targets without human intervention at the point of engagement. The most contentious category comprises lethal autonomous weapons systems, which apply force without a human taking the final decision, and such systems are already present in drone technology, missile defence and border surveillance. Proponents argue that they detect threats faster and reduce human error; the principal concern is that they may kill without the exercise of human judgment.

Harm may arise in several ways. A system trained on defective or unrepresentative data may misclassify civilians as combatants. A system trained on biased data may discriminate in its targeting. An adversary may manipulate the system into engaging an unintended target. A machine learning system may also acquire behaviours its designers never intended, producing results that in armed conflict may amount to unlawful killing, destruction of protected objects or serious violations of international humanitarian law.

The accountability gap follows. The programmer may say that he did not intend the harm; the commander may say that the weapon was certified as safe; the armed force

⁴⁵ R. Farajpour, "Legal and Comparative Analysis of Civil Liability of Artificial Intelligence in Automated Decision-Making" (2025) AI and Tech in Behavioral and Social Sciences.

may say that the system malfunctioned. Since criminal liability requires a guilty mind, and no human being directed the particular engagement, proof becomes difficult, and it is in this sense that autonomous weapons are said to threaten the premise of individual criminal responsibility.

B. International Humanitarian Law and the Rome Statute

International humanitarian law protects civilians in armed conflict through the principles of distinction, proportionality and military necessity. Those rules are addressed to human decision-makers, and they assume that a person applies judgment to each attack. Article 36 of Additional Protocol I to the Geneva Conventions requires a State, in the study, development, acquisition or adoption of a new weapon, means or method of warfare, to determine whether its employment would in some or all circumstances be prohibited by international law.⁴⁶ The obligation is one of weapons review, and it operates upon the State and its officials rather than upon the system.

The Rome Statute of the International Criminal Court supplies the individual-responsibility framework that the accountability-gap argument tends to overlook. Article 25 establishes individual criminal responsibility, including for ordering, soliciting, inducing or otherwise contributing to the commission of a crime.⁴⁷ Article 28 establishes the responsibility of commanders and other superiors for crimes committed by forces under their effective command and control where they knew, or owing to the circumstances at the time should have known, that the forces were committing or about to commit such crimes, and failed to take all necessary and reasonable measures within their power to prevent or repress them.⁴⁸ Article 30 provides that, unless otherwise provided, a person is criminally responsible only where the material elements are committed with intent and knowledge.⁴⁹

⁴⁶ Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3, art 36.

⁴⁷ Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3, art 25.

⁴⁸ Rome Statute, art 28.

⁴⁹ Rome Statute, art 30.

Articles 28 and 30 read together answer a substantial part of the accountability-gap objection. Article 30 confirms that the mental element attaches to a person and not to a machine, which is consistent with the position taken throughout this paper. Article 28 supplies the mechanism by which responsibility is located notwithstanding the absence of a direct order, because the standard of “should have known” reaches the commander who deployed a system he knew to be unreliable, or who failed to institute measures capable of preventing unlawful engagements. The gap, properly analysed, is therefore less a gap in the law than a difficulty of proof, and it is narrowed by any measure that improves the recording and auditability of autonomous engagements.⁵⁰

C. The Current International Process

The international position has developed materially and must be stated accurately. Following Resolution 79/62 of 2024, the General Assembly adopted Resolution 80/57 on lethal autonomous weapons systems on 1 December 2025 at its 52nd plenary meeting, by a recorded vote of 164 in favour to 6 against, with 7 abstentions, the States voting against being Belarus, Burundi, the Democratic People’s Republic of Korea, Israel, the Russian Federation and the United States.⁵¹ The Resolution affirms the need for a broad and inclusive multilateral approach and calls upon the High Contracting Parties to the Convention on Certain Conventional Weapons to complete the elements of an instrument being developed under the mandate of the Group of Governmental Experts, with a view to future negotiations.

Within that Group, the Chair’s rolling text of 18 December 2025 is the working basis for any future instrument, and the mandate of the Group is to be revisited at the Seventh Review Conference of the Convention in November 2026, which is expected to determine whether formal negotiations are to commence. India has consistently maintained that the Group of Governmental Experts is the appropriate forum, has opposed the creation of parallel processes, and has supported a politically binding

⁵⁰ Antonio Coco, “Modes of Liability for AI-Enabled Crimes in International Criminal Law” (2026) 107 *International Law Studies* 228.

⁵¹ UNGA Res 80/57 (1 December 2025) UN Doc A/RES/80/57, adopted at the 52nd plenary meeting; see also UNGA Res 79/62 (2024).

instrument founded upon the Group's guiding principles as a first step, while declining to endorse negotiations for a legally binding prohibition at this stage. India does not appear among the States recorded as voting against Resolution 80/57 or as abstaining.

Two conclusions follow for the present inquiry. The international community has not proposed that an autonomous weapon should itself be an accused; the entire process is directed to preserving human control and human responsibility. The measures under discussion, principally a requirement of meaningful human control over the use of force, rules on testing and transparency, and obligations of record-keeping, are the same measures that make human criminal responsibility provable. Indian criminal law, which under the *Bharatiya Nyaya Sanhita* predicates liability upon intention or negligence, would require proof that a human being acted rashly or recklessly in deploying such a system, and is at present insufficiently developed to address violations arising from AI-enabled warfare.

X. SUGGESTIONS AND RECOMMENDATIONS

The proposals that follow are consolidated in a single Part and are ranked according to the legal instrument each requires. They proceed upon a settled position on the question of liability standard. For the criminal law, this paper rejects strict liability and adopts a negligence-based standard for high-risk artificial intelligence, requiring proof of gross negligence or recklessness in accordance with Jacob Mathew. Strict liability is retained, but confined to the civil sphere, where no-fault product liability secures compensation without attributing blame.

That division is deliberate: criminal punishment without fault would be inconsistent with the moral foundation of the *Sanhita* and would deter the development of beneficial systems, whereas no-fault civil liability relieves the victim of an evidential burden he cannot discharge against an opaque system. The two regimes are complementary and not alternatives, and product liability accordingly remains the principal route to compensation while criminal liability is reserved for culpable conduct.

A. Measures Available Under the Existing Law

First, the attribution doctrine established in *Iridium India Telecom Ltd v Motorola Inc* and clarified in *Sunil Bharti Mittal v CBI* is already sufficient to reach an enterprise whose controlling officers knowingly deploy an unsafe system, and no amendment is required for that purpose. Prosecutors should be guided to examine internal testing records, risk registers, warnings received and post-deployment monitoring as the evidence from which knowledge or gross negligence may be inferred.

Secondly, section 106 of the *Bharatiya Nyaya Sanhita, 2023* already reaches a developer or deployer whose rash or negligent conduct causes death, and the *Jacob Mathew* standard supplies the threshold. Courts may apply the existing standard to AI-related harm without legislative intervention, by asking whether the defendant departed from the practices accepted as proper by a responsible body of professional opinion.

Thirdly, section 85 of the *Information Technology Act, 2000* and Chapter VI of the *Consumer Protection Act, 2019* already allocate responsibility to companies and to product manufacturers respectively, and should be invoked in preference to proposals for novel forms of liability.

B. Measures Requiring Amendment of Existing Legislation

First, the *Bharatiya Nyaya Sanhita, 2023* should be amended to insert a definition of a high-risk automated system and an offence of causing death or grievous hurt through the grossly negligent development, deployment or supervision of such a system, framed on the model of section 106 and subject to the same gross-negligence threshold. The amendment should be confined to defined high-risk applications, namely clinical decision support, autonomous vehicles, critical infrastructure and the use of automated systems in law enforcement.

Secondly, the *Information Technology Act, 2000* should be amended, as the *India AI Governance Guidelines* themselves recommend, to clarify the classification of artificial intelligence systems, to settle whether the safe harbour in section 79 extends to systems that generate or modify content, and to impose statutory duties of pre-

deployment testing, logging and explainability upon providers and deployers of high-risk systems. Those duties matter for the criminal law because breach of a defined statutory duty supplies the predicate from which gross negligence may be established, which is otherwise difficult to prove against an opaque system.

Thirdly, the Consumer Protection Act, 2019 should be amended to place beyond doubt that software and artificial intelligence systems supplied as, or incorporated within, a product fall within the definition of a product for the purposes of product liability, and to address defects arising from post-sale learning. The revised European product liability regime under Directive (EU) 2024/2853 furnishes a drafting precedent.

Fourthly, the sectoral regulators identified in the India AI Governance Guidelines, including the Reserve Bank of India, the Securities and Exchange Board of India and the National Medical Commission, should prescribe domain-specific standards of validation, human oversight and record-keeping, so that the content of the duty of care is ascertainable in advance.

C. Measures Requiring New Legislation

First, a dedicated criminal statute on artificial intelligence is not recommended. The Government's stated preference for amending existing law is sound, because a separate penal code for autonomous systems would duplicate the general part of the Sanhita and would create classification disputes at the margin without closing any gap that amendment cannot close.

Secondly, a new statute is warranted for one purpose only. Parliament should enact a no-fault compensation scheme for harm caused by designated high-risk automated systems, funded by compulsory insurance maintained by the deployer and administered through an existing adjudicatory forum. The scheme addresses the residual case in which harm is established, the system is opaque and no human fault can be proved; it secures the victim's remedy without requiring the criminal law to abandon the fault principle. The Automated and Electric Vehicles Act 2018 of the United Kingdom supplies the model.

Thirdly, legal personality, whether electronic, functional or hybrid, should not be conferred upon artificial intelligence systems. The European Parliament's proposal of 2017 was not adopted, the Artificial Intelligence Act allocates obligations to providers and deployers instead, and the Indian authorities on juristic personality confer status only where an entity holds an interest requiring protection through a human guardian. Conferring personality upon a system that cannot be punished would create a locus for blame without a locus for sanction, and would operate to shield the human actors behind it.

Fourthly, in relation to autonomous weapons, India should support the retention of meaningful human control over the use of lethal force, and should press within the Group of Governmental Experts for obligations of testing, transparency and engagement logging. Those obligations serve the criminal law directly, because they generate the evidence by which command responsibility under Article 28 of the Rome Statute, and rashness or recklessness under the *Bharatiya Nyaya Sanhita*, may be proved.

XI. CONCLUSION

This paper has examined whether artificial intelligence can be an accused in criminal law, and has answered that it cannot. Three findings support that conclusion.

First, an artificial intelligence system falls outside the definition of "person" in section 2(26) of the *Bharatiya Nyaya Sanhita*, 2023, which extends only to companies, associations and bodies of persons; and the Indian authorities conferring juristic personality upon idols, scriptures and rivers confer it for the holding of interests through human guardians, never for the purpose of criminal answerability.

Secondly, such a system cannot satisfy the mental element, because the definitions of "voluntarily" and "reason to believe" in sections 2(33) and 2(29) are framed in terms of intention, knowledge and belief, and because punishment presupposes a subject capable of suffering deprivation. On Hallevy's framework, the perpetration-via-another and natural-probable-consequence models account for AI-related harm within existing doctrine, while the direct liability model must be rejected.

Thirdly, the existing Indian statutes are inadequate, though not for the reason usually advanced: the Bharatiya Nyaya Sanhita and the Information Technology Act are not deficient in their attribution of fault to human actors, but in the absence of defined duties of testing, logging and explainability whose breach would make that fault provable against an opaque system.

The practical difficulty is therefore evidential rather than conceptual. The accountability gap arises because the decision path of a black box system cannot be reconstructed, and not because the law lacks a subject upon whom to fix responsibility. It follows that the remedy lies in transparency obligations and in the allocation of defined duties along the value chain, which is the approach taken by the India AI Governance Guidelines of 2025, by the Intermediary Guidelines Amendment Rules of 2026, by the European Artificial Intelligence Act, and, in the clearest statutory form, by the Automated Vehicles Act 2024 of the United Kingdom.

The recommendations in Part 12 proceed accordingly. Criminal liability for high-risk artificial intelligence should rest upon gross negligence and not upon strict liability; strict liability belongs to the civil sphere, where a no-fault compensation scheme supported by compulsory insurance should address the residual case. Targeted amendment of the Bharatiya Nyaya Sanhita, the Information Technology Act and the Consumer Protection Act is preferable to a dedicated criminal statute, and legal personality should not be conferred upon artificial intelligence in any form. Legislative intervention is required, but its object should be to make human fault provable rather than to invent a machine defendant. The law is stated as at 19 October 2026.

XII. REFERENCES

A. Cases

1. Iridium India Telecom Ltd v Motorola Inc (2011) 1 SCC 74.
2. Jacob Mathew v State of Punjab (2005) 6 SCC 1; AIR 2005 SC 3180.
3. Mohd Salim v State of Uttarakhand (Uttarakhand High Court, 20 March 2017), Writ Petition (PIL) No 126 of 2014.
4. Nathulal v State of Madhya Pradesh AIR 1966 SC 43.

5. *Shiromani Gurdwara Parbandhak Committee v Som Nath Dass* (2000) 4 SCC 146.
6. *Standard Chartered Bank v Directorate of Enforcement* (2005) 4 SCC 530.
7. *State of Maharashtra v Syndicate Transport Co (P) Ltd* AIR 1964 Bom 195.
8. *State of Uttarakhand v Mohd Salim, SLP (C) No 16879 of 2017*, order dated 7 July 2017 (Supreme Court of India).
9. *Sunil Bharti Mittal v Central Bureau of Investigation* (2015) 4 SCC 609.
10. *Yogendra Nath Naskar v Commissioner of Income Tax* (1969) 1 SCC 555.

B. Indian Legislation and Subordinate Instruments

1. Bharatiya Nyaya Sanhita 2023.
2. Consumer Protection Act 2019.
3. Digital Personal Data Protection Act 2023.
4. Digital Personal Data Protection Rules 2025.
5. Indian Penal Code 1860.
6. Information Technology Act 2000.
7. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules 2026, notified 10 February 2026.
8. Ministry of Electronics and Information Technology, Notification G.S.R. 843(E), 13 November 2025 (commencement of the Digital Personal Data Protection Act 2023).
9. Ministry of Electronics and Information Technology, India AI Governance Guidelines (5 November 2025).

C. Foreign and International Instruments

1. Automated and Electric Vehicles Act 2018 (c 18) (UK).
2. Automated Vehicles Act 2024 (c 10) (UK).
3. Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC [2024] OJ L2024/2853.

4. European Commission, Commission Work Programme 2025, COM(2025) 45 final, Annex IV; Withdrawal of Commission proposals [2025] OJ C/2025/5423, 6 October 2025.
5. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) [2018] OJ C252/239.
6. Open Letter to the European Commission on Artificial Intelligence and Robotics (2018), available at <http://www.robotics-openletter.eu>.
7. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3.
8. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [2024] OJ L2024/1689.
9. Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 [2026] OJ L, 24 July 2026.
10. Restatement (Third) of Torts: Products Liability (American Law Institute 1998).
11. Rome Statute of the International Criminal Court (adopted 17 July 1998, entered into force 1 July 2002) 2187 UNTS 3.
12. UNESCO, Recommendation on the Ethics of Artificial Intelligence (adopted 23 November 2021).
13. UNGA Res 79/62 (2024) on lethal autonomous weapons systems.
14. UNGA Res 80/57 (1 December 2025) UN Doc A/RES/80/57 on lethal autonomous weapons systems.
15. Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, Chair's rolling text (18 December 2025).

D. Books, Theses and Journal Articles

1. Coco, Antonio, "Modes of Liability for AI-Enabled Crimes in International Criminal Law" (2026) 107 International Law Studies 228.
2. Diamantis, Mihailis E., "The Extended Corporate Mind: When Corporations Use AI to Break the Law" (2020) 98 North Carolina Law Review 893.
3. Farajpour, R., "Legal and Comparative Analysis of Civil Liability of Artificial Intelligence in Automated Decision-Making" (2025) AI and Tech in Behavioral and Social Sciences.
4. Giannini, Alice, Criminal Behavior and Accountability of Artificial Intelligence Systems (PhD thesis, Maastricht University and University of Florence, Eleven Publishers 2023).
5. Hallevy, Gabriel, "The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control" (2010) 4(2) Akron Intellectual Property Journal, art 1.
6. Mamak, Kamil, "AI Personhood, Criminal Law, and Punishment" in Oxford Intersections: AI in Society (Oxford University Press 2025).
7. Sachoulidou, Athina, "AI Systems and Criminal Liability: A Call for Action" (2024) 11(1) Oslo Law Review 1, DOI 10.18261/olr.11.1.3.

E. Reports and Other Sources

1. Law Commission (England and Wales), Artificial Intelligence and the Law: a discussion paper (31 July 2025).
2. Nanos, Andreas, "Criminal Liability of Artificial Intelligence" (SSRN, 6 November 2023), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4623126.
3. "Beyond Personhood: The Evolution of Legal Personhood and AI Governance" (2025) TechReg, 7 September 2025, available at <https://techreg.org/article/view/22555>.