International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

Artificial Intelligence and Criminal Liability in India: Rethinking Traditional Principles of Criminal Jurisprudence

Dr. Dilip M. Kathad

Abstract:

Artificial Intelligence (AI) systems now drive vehicles, screen loan applicants, recommend medical treatment, generate synthetic media and execute financial transactions with limited human supervision. When such systems cause harm, Indian criminal law, which rests on the twin pillars of *actus reus* and *mens rea* and on the premise that only a human or a juristic person created by statute can be an offender, struggles to identify who is answerable. This paper examines whether the Bharatiya Nyaya Sanhita, 2023 (BNS), the Information Technology Act, 2000 and allied laws can accommodate harms in which an autonomous, opaque and self-learning system is the proximate cause. Using a doctrinal method supported by comparative analysis of the United Kingdom, the United States, the European Union and the Council of Europe, the paper analyses the constitutional framework, the Supreme Court's jurisprudence on corporate criminal liability and statutory mens rea, and leading scholarship on machine liability. It finds that (i) AI cannot presently be an offender because it lacks voluntariness, moral agency and the capacity to be punished; (ii) existing doctrines of abetment, common intention, negligence and corporate attribution cover only a fraction of AI-related harm and leave a "responsibility gap" where the harm is unforeseeable to developers and deployers; (iii) the Indian regime is fragmented and has no AI-specific offences or duty-of-care standards; and (iv) strict human-centred accountability, rather than robot personhood, is the correct normative response. The paper's original contribution is a graduated liability model, tied to foreseeability, control and risk classification, together with specific amendments and institutional measures for the BNS, the IT Act and sectoral regulators.

Keywords: Artificial Intelligence; Criminal Liability; *Mens Rea* and *Actus Reus*; Bharatiya Nyaya Sanhita, 2023; Responsibility Gap; Corporate Criminal Liability; Legal Personhood; Algorithmic Accountability.

Introduction

Criminal law is the most coercive instrument available to the State. It condemns conduct, restricts liberty and carries moral stigma, and for that reason it has always been built around a blameworthy human mind. The maxim *actus non facit reum nisi mens sit rea*, an act does not make a person guilty unless the mind is also guilty, remains the organising principle of Indian criminal jurisprudence, even after the IPC (1860) was replaced by the BNS, the Bharatiya Nagarik Suraksha Sanhita (BNSS) and the Bharatiya Sakshya Adhiniyam (BSA), all in force from 1 July 2024.

AI challenges this architecture at its root. Classical software executes instructions written by a programmer; machine-learning systems derive their own decision rules from data, adapt after

deployment and often cannot explain their outputs (the "black box" problem). Real incidents illustrate the difficulty: the 2018 death of a pedestrian struck by an Uber test vehicle in Arizona, in which prosecutors charged only the safety driver; fatal crashes involving driver-assistance systems; deepfake-enabled impersonation and financial fraud; and discriminatory outcomes from automated decision tools. In India, AI is being adopted rapidly in policing (facial recognition), finance, healthcare, agriculture and governance, supported by policy documents such as NITI Aayog's *National Strategy for Artificial Intelligence* (2018) and *Responsible AI* papers (2021), and more recently the government's AI governance guidelines. Yet no Indian statute specifically defines AI-related offences or allocates criminal responsibility for autonomous harm.

The subject therefore matters on three levels: doctrinally (can traditional principles survive?), practically (who can be prosecuted when an AI system kills, defrauds or defames?) and constitutionally (how can accountability be secured without violating Articles 14, 20 and 21?).

Statement of Research Problem

Indian criminal law presupposes a human or juristic offender who performs a voluntary act with a culpable mental state. Autonomous AI systems (a) act without contemporaneous human instruction, (b) evolve unpredictably, (c) involve many contributors (data providers, developers, integrators, deployers, users) so that "the problem of many hands" arises, and (d) lack consciousness and intent. Consequently, when an AI system causes death, injury, fraud or discrimination, three difficulties follow: the **attribution problem** (whose act and whose mind?), the **foreseeability problem** (can harm from emergent behaviour be foreseen so as to constitute negligence or knowledge?) and the **evidentiary problem** (how can opaque algorithms be proved beyond reasonable doubt?). The core problem is a probable *responsibility gap*: serious harm for which no natural or legal person can be convicted under existing provisions, coupled with the risk of over-criminalising innocent developers or users in an attempt to fill the gap.

Review of Literature

1. Machine liability models. Gabriel Hallevy's *When Robots Kill* (2013) and *Liability for Crimes Involving Artificial Intelligence Systems* (2015) propose three models: perpetration-via-another (the AI as an innocent instrument of a human), natural-probable-consequence liability (the human is liable for foreseeable results of the system's operation) and direct liability (the AI itself satisfies actus reus and mens rea). Hallevy's third model is the most controversial because it treats a system's functional capacities as equivalent to mental elements.

2. Scepticism about robot culpability. Ryan Abbott and Alex Sarch ("Punishing Artificial Intelligence", 2019) discuss the problem of "hard AI crimes" and the limits of existing doctrine. Joanna Bryson and colleagues argue against legal personhood for machines because it may shield human manufacturers from responsibility. Gless, Silverman and Weigend ("If Robots Cause Harm, Who Is to Blame?", 2016) emphasise that self-learning systems break the traditional chain of foreseeability and propose duties of care and monitoring for those who release them. Andreas Matthias's notion of the "responsibility gap" (2004) is the foundational statement of the problem.

3. Policy and legislative documents. The European Parliament's 2017 Resolution on Civil Law Rules on Robotics floated "electronic personhood" for sophisticated robots, a proposal heavily criticised by an

open letter of AI and legal experts. The EU's Regulation (EU) 2024/1689 (the AI Act) chose instead a risk-based regulatory approach with administrative penalties and no machine personhood.

4. Indian scholarship and reports. Indian writing is largely in the form of articles in law-school journals and policy papers, which focus on civil liability, data protection, privacy, intellectual property of AI output and algorithmic bias. Government documents (NITI Aayog 2018 and 2021; MeitY advisories of 2023–24 on deepfakes and AI-generated content; the Digital Personal Data Protection Act, 2023) are regulatory and principle-based rather than penal. Commentary on corporate criminal liability (following *Standard Chartered Bank*, *Iridium* and *Sunil Bharti Mittal*) is extensive but is rarely applied to autonomous systems. Commentary on the BNS has concentrated on structural changes such as new offences (organised crime, terrorism, mob lynching) and community service, not on technology-driven harm.

Research Gap

Existing work (i) is dominated by Western jurisdictions and does not engage with the text of the BNS or Indian precedents; (ii) in India, addresses AI mainly through civil, privacy and intellectual-property lenses; (iii) treats corporate criminal liability and AI separately, without testing whether the Indian "alter ego" and identification doctrines can stretch to algorithmic decision-making; (iv) does not map the constitutional limits (Articles 14, 20 and 21) on any new regime of AI liability; and (v) offers no graduated, implementable model for Indian statutory reform. This paper attempts to fill these gaps.

Research Objectives

1. To examine whether traditional principles of actus reus, mens rea, abetment, common intention and negligence under the BNS can be applied to harm caused by AI systems.
2. To analyse the adequacy of the IT Act, 2000, the DPDP Act, 2023 and sectoral laws in addressing AI-related criminality.
3. To study how Indian courts have treated corporate criminal liability, statutory mens rea and juristic personality, and their relevance to AI.
4. To compare the approaches of the UK, USA, EU and Council of Europe to AI and criminal responsibility.
5. To identify the responsibility gap and propose a workable, constitutionally sound liability model for India.

Research Questions

1. Can an AI system be treated as an "offender" or a "person" under Indian criminal law?
2. Which natural or legal persons (developer, manufacturer, deployer, user) can be criminally liable when an AI system causes harm, and on what mental standard?
3. Are the current doctrines of abetment, conspiracy, negligence and corporate attribution sufficient for AI-related offences?
4. What lessons can India draw from foreign and international developments?
5. What reforms are needed to close the responsibility gap without compromising the principles of legality and fair trial?

Research Hypotheses

H1 (Null): The existing framework of the BNS, the IT Act and corporate criminal liability doctrines is adequate to address criminal harm caused by AI.

H2 (Alternative): The existing framework is inadequate and leaves a responsibility gap, which can be closed by a graduated, human-centred liability model without granting legal personhood to AI.

H3: Conferring criminal personhood on AI systems is neither doctrinally justified nor practically useful, because punishment cannot serve deterrence, retribution or reformation for a non-sentient entity.

Research Methodology

The study is primarily **doctrinal**: it analyses primary sources (the Constitution, the BNS, BNSS, BSA, the IT Act, the DPDP Act, and judgments of the Supreme Court and High Courts) and secondary sources (books, journal articles, government and Law Commission reports). It is also **comparative and analytical**, examining the United Kingdom, the United States, the European Union and the Council of Europe to test the transferability of foreign solutions. No empirical data were collected; where factual incidents are cited, they are drawn from publicly reported court records and reliable reports. Citations follow a uniform legal citation style. The principal limitation is the scarcity of Indian judicial authority directly on AI offences, which means that the analysis necessarily proceeds by analogy and principle.

Legal and Constitutional Framework

Constitutional provisions

Article 14 (equality and non-arbitrariness) is engaged where algorithmic decisions in policing or welfare delivery discriminate or cannot be explained.

Article 20(1) prohibits conviction under an *ex post facto* law and imposition of a greater penalty than that prescribed at the time of the offence. Any AI offence must therefore be defined in advance, which restricts judicial improvisation.

Article 20(3) (protection against self-incrimination) is relevant to compelled disclosure of source code, logs or model weights by accused persons.

Article 21 protects life, liberty and fair procedure (*Maneka Gandhi v. Union of India* (1978) 1 SCC 248) and, after *Justice K.S. Puttaswamy v. Union of India* (2017) 10 SCC 1, informational privacy, which is directly implicated in AI-driven surveillance and profiling.

Article 19(1)(a) and **19(2)** constrain the penal regulation of AI-generated speech, as the Court emphasised in *Shreya Singhal v. Union of India* (2015) 5 SCC 1.

Statutory framework

Bharatiya Nyaya Sanhita, 2023. The BNS retains the traditional structure of the IPC. Key provisions relevant to AI are:

1. General principles of intention, knowledge, negligence and recklessness, which underlie most offences.
2. The definition of "person" (BNS s. 2, corresponding to IPC s. 11), which includes a company, association or body of persons whether incorporated or not. It does not include an AI system, whether as a natural person or an artificial juridical person.
3. Common intention (BNS s. 3(5), corresponding to IPC s. 34), criminal conspiracy (BNS s. 61) and abetment (BNS Chapter V), which require at least two minds or an abettor's intention or knowledge.

4. Causing death by negligence (BNS s. 106, corresponding to IPC s. 304A) and acts endangering life or personal safety (BNS s. 125), which are the most plausible vehicles for AI harm involving autonomous vehicles, medical AI and industrial robots.
5. Cheating, personation, forgery and defamation (for example BNS ss. 318, 319, 336 and 356), which can reach deepfake fraud and impersonation committed by or through AI.
6. Offences against public tranquillity and public mischief (for example BNS ss. 196 and 353), relevant to AI-driven disinformation.

Information Technology Act, 2000. Sections 66C (identity theft), 66D (cheating by personation using a computer resource), 66E (violation of privacy), 67, 67A and 67B (obscene and sexually explicit material, including child sexual abuse material) and the intermediary regime in s. 79 and the IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 apply to AI-generated content when a human actor or intermediary is identifiable. Section 85 imposes liability on persons in charge of companies.

Digital Personal Data Protection Act, 2023. Imposes duties on data fiduciaries, with financial penalties rather than imprisonment, and applies to AI training and deployment involving personal data.

Other laws. The Consumer Protection Act, 2019 (product liability), the Motor Vehicles Act, 1988 (human driver responsibility; no framework for autonomous vehicles), the Drugs and Cosmetics Act, 1940 and Medical Device Rules, 2017 (software as a medical device) and sectoral regulations of the RBI, SEBI and IRDAI on algorithmic decision-making and trading.

Policy instruments. NITI Aayog's *National Strategy for AI* (2018) and *Principles for Responsible AI* (2021), MeitY advisories on deepfakes and unreliable AI models (2023–2024) and the India AI governance guidelines issued in 2025 articulate principles of safety, transparency and accountability but are non-binding and do not create penal liability.

Judicial Pronouncements

Mens rea and statutory offences. In *State of Maharashtra v. Mayer Hans George* (AIR 1965 SC 722), the Supreme Court held that mens rea is presumed to be an ingredient of an offence unless a statute expressly or by necessary implication excludes it. *Nathulal v. State of Madhya Pradesh* (AIR 1966 SC 43) reaffirmed that the presumption is strong. These cases mean that Indian courts are unlikely to dispense with a mental element simply because an AI was involved; the mental element must be located in a human or corporate defendant.

Corporate criminal liability. *Standard Chartered Bank v. Directorate of Enforcement* (2005) 4 SCC 530 held that a company can be prosecuted even for offences carrying mandatory imprisonment, with the fine being the only available punishment. *Iridium India Telecom Ltd. v. Motorola Inc.* (2011) 1 SCC 74 accepted that a company can be guilty of offences requiring mens rea, by attributing to it the intent of those who are its directing mind and will. *Sunil Bharti Mittal v. CBI* (2015) 4 SCC 609 clarified that the acts of an "alter ego" can be attributed to the company, but that directors are not automatically liable for company offences in the absence of statutory vicarious liability or active role. These precedents are the best-developed Indian tools for attributing AI-related harm to deploying entities.

Vicarious liability. *Sham Sunder v. State of Haryana* (1989) 4 SCC 630 and *Aneeta Hada v. Godfather Travels and Tours Pvt. Ltd.* (2012) 5 SCC 661 establish that criminal liability for another's act is exceptional and must be expressly created by statute. This is critical: absent legislation, a developer or CEO cannot be convicted simply because their product malfunctioned.

Juristic personality. The Supreme Court recognised the Hindu idol as a juristic person (*Shiromani Gurdwara Parbandhak Committee v. Som Nath Dass* (2000) 4 SCC 146; earlier, *Pramatha Nath Mullick v. Pradyumna Kumar Mullick* (1925) 52 IA 245) and the Uttarakhand High Court declared the Ganga and Yamuna rivers to be legal persons in *Mohd. Salim v. State of Uttarakhand* (2017), though that order was stayed by the Supreme Court. These decisions show that Indian courts can create legal personality by purposive reasoning, but in each case the entity acts through human guardians or representatives, and personhood serves protective rather than penal ends.

Technology, rights and judicial use of AI. *Puttaswamy* and *Anuradha Bhasin v. Union of India* (2020) 3 SCC 637 (proportionality and reasoned decision-making in restrictions on digital rights) supply the proportionality test against which any AI-based surveillance or profiling must be tested. In *Jaswinder Singh v. State of Punjab* (2023) the Punjab and Haryana High Court referred to a chatbot's response only to gain a broader perspective on bail jurisprudence, and the Delhi High Court in *Christian Louboutin SAS v. Abubaker Shaikh* (2023) held that AI output cannot be the basis of adjudication. Personality-rights orders of the Delhi and Bombay High Courts in favour of celebrities, for example *Anil Kapoor v. Simply Life India* (2023), restrain AI-based voice and image cloning. These cases are civil, but they signal judicial recognition of AI-enabled wrongs. **To date, there is no reported Supreme Court judgment on criminal liability for harm caused by an autonomous AI system.**

Critical Analysis

Can AI be an offender?

Under Indian law only a "person" can commit an offence, and AI is not one. Even if the definition were expanded, three obstacles remain. *First*, **actus reus** requires a voluntary act; an AI system's outputs are produced by statistical inference and are not volitional in the legal sense. *Second*, **mens rea** (intention, knowledge, recklessness) presupposes consciousness and the capacity to understand wrongfulness. Functional equivalence, which Hallevy proposes, can be satisfied by a thermostat; it does not capture moral blameworthiness. *Third*, **punishment** loses its purposes: retribution is meaningless against a non-sentient entity, deterrence operates only on the humans who build and deploy it, and "reformation" is merely software modification, which is a regulatory remedy rather than a penal one. Measures such as shutting down, retraining or deleting a model are better understood as civil or administrative measures (akin to forfeiture of an instrumentality) than as punishment. Conferring criminal personhood might also encourage human actors to hide behind the machine. Hypothesis H3 is therefore supported.

Perpetration-via-another and intentional misuse

Where a human deliberately uses AI to commit an offence (generating deepfake pornography, running a phishing bot, deploying an autonomous drone), the existing law functions adequately: the human is the principal offender and the AI is an instrument, analogous to a weapon. Section 66D of the IT Act and BNS provisions on cheating and personation apply. The practical difficulty lies in attribution (anonymity, cross-border infrastructure) and evidence, not in doctrine.

Negligence and the foreseeability problem

The more difficult cases involve no malicious human. An autonomous vehicle misclassifies a pedestrian; a diagnostic model recommends a lethal dose; a trading algorithm triggers a market crash. Liability must then be sought in **negligence** (BNS s. 106) or **rash and negligent acts endangering life** (s. 125). Indian courts require **gross** negligence for criminal liability, a standard higher than civil negligence. For a developer or deployer to be criminally negligent, the harm must have been reasonably foreseeable and

the precautions taken unreasonable. Yet emergent behaviour in self-learning systems may be unforeseeable even to experts, and compliance with industry standards may exist in a world without any binding standards. Courts therefore face either acquittal (the gap) or an implied lowering of the mental standard (a risk to Article 21 and the principle of legality).

Abetment, conspiracy and common intention

These doctrines presuppose the intention or knowledge of two or more minds. They are unhelpful where the "co-participant" is an AI because the system lacks intent. They might apply among the humans in a development chain only if each knowingly facilitated an offence, which is rare in ordinary product development.

Corporate attribution

Under *Iridium* and *Sunil Bharti Mittal*, a company's liability depends on identifying a human "directing mind". In AI deployment, key decisions are distributed among engineering, product, compliance and data teams, and the "mind" may be diffuse. A *systems* or "corporate culture" approach, recognised in Australia's Criminal Code (Part 2.5) and in the UK's failure-to-prevent offences, would better reflect how organisations actually manage algorithmic risk, but Indian courts have not yet adopted it. Moreover, fines against companies are often a weak deterrent unless scaled to turnover.

The responsibility gap and over-criminalisation

In practice the possible defendants are numerous: the data supplier, model developer, system integrator, deploying company, a human "in the loop" and the end user. Naming the nearest human, the so-called **"moral crumple zone"**, is unfair: a safety driver or clinician cannot be expected to out-think an opaque system at machine speed. At the same time, leaving harm unpunished would erode public trust. Liability must therefore track **control and benefit**, not mere proximity.

Evidence and procedure

Proof beyond reasonable doubt in AI cases requires access to training data, model versions, logs and testing records. The BSA, 2023 recognises electronic records and requires a certificate for their admissibility, but offers no procedure for algorithmic audit or expert explanation of model behaviour. The BNSS's provisions on search and seizure of digital devices and audio-video recording help preservation, yet trade-secret claims, cross-border data and the right against self-incrimination under Article 20(3) will raise contested questions. The explainability deficit of black-box models may also impede an accused's right to a fair defence.

Assessment of the hypotheses

The analysis supports **H2** and rejects **H1**: the existing framework deals adequately with deliberate human misuse but not with autonomous, unforeseeable harm or diffuse responsibility.

International Perspective

European Union. Regulation (EU) 2024/1689 (AI Act) classifies systems by risk (prohibited, high-risk, limited-risk and minimal-risk), imposes obligations of data governance, human oversight, transparency, logging and post-market monitoring on providers and deployers, and prescribes heavy administrative fines (up to a percentage of worldwide turnover). Criminal liability is left to Member States, and the European Parliament's 2017 suggestion of "electronic personhood" was not adopted. The revised EU Product Liability Directive (2024) extends strict liability to software and AI.

United Kingdom. Corporate criminal liability historically followed the "identification doctrine" (*Tesco Supermarkets Ltd v. Nattrass* [1972] AC 153), which was widely criticised as ineffective for large

organisations; the Economic Crime and Corporate Transparency Act 2023 widened it to senior managers for economic crimes. The Automated Vehicles Act 2024 establishes an "authorised self-driving entity" and a user-in-charge immunity, shifting responsibility for the driving task from the user to the manufacturer or operator when the vehicle is driving itself, an instructive model for India's Motor Vehicles Act.

United States. There is no federal AI criminal statute. Prosecutors rely on existing offences against humans: in the Uber Arizona case, the backup driver was charged and later pleaded guilty to a reduced charge of endangerment, while the company was not prosecuted. Driver-assistance fatalities have produced manslaughter prosecutions of human drivers. Federal and State rules address deepfakes (for example, election and non-consensual intimate imagery laws), and the Federal Trade Commission treats deceptive AI claims as unfair practices.

Council of Europe. The Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024) obliges parties to ensure accountability and remedies for human-rights violations by AI, without prescribing personhood.

Other developments. The OECD AI Principles (2019, revised 2024), UNESCO's Recommendation on the Ethics of AI (2021) and the 2017 granting of "citizenship" to the robot Sophia by Saudi Arabia (a largely symbolic act) illustrate the global debate. Across jurisdictions, the dominant trend is to **regulate the humans and organisations behind AI, not the AI itself**.

Lessons for India: adopt a risk-based classification; impose enhanced duties on providers and deployers of high-risk systems; embed logging and audit trails; use organisational (not only individual) fault for corporate liability; and clarify allocation of responsibility between users and manufacturers for highly automated systems.

Research Findings

1. **AI is not and should not be treated as a criminal "person".** It lacks voluntariness, moral agency and the capacity to experience punishment; machine personhood would dilute human accountability.
2. **Intentional misuse is largely covered** by the BNS and IT Act, subject to practical problems of attribution and evidence.
3. **A responsibility gap exists** for harm from autonomous, unforeseeable AI behaviour, because negligence, abetment and common intention depend on human foresight or intent.
4. **Corporate criminal liability doctrines are useful but limited**, since the "directing mind" test fits poorly with diffuse algorithmic decision-making and vicarious liability requires explicit statutory provision.
5. **The Indian legal framework is fragmented and non-binding in the AI domain**: policy documents set principles but create no penal duties, and no law addresses autonomous vehicles, medical AI or algorithmic decision-making by the State.
6. **Procedural law is not ready**: the BSA and BNSS do not provide for algorithmic audits, access to model documentation or court-appointed technical experts.
7. **Constitutional limits** (Articles 14, 20 and 21) bar retrospective or strict criminal liability without clear statutory definition and constrain use of AI in policing and surveillance.
8. **Comparative practice** favours risk-based regulation, strengthened organisational liability and regulatory fines over machine personhood.

Suggestions and Recommendations

A graduated liability model

Tier	Situation	Proposed liability
1	Human intentionally uses AI to commit an offence	Principal offender liability under the BNS and IT Act, with an aggravating factor where AI-enabled scale or deception is used
2	Harm arises from a known, unmitigated risk in a high-risk system	Criminal liability of the responsible deployer and provider for gross negligence or recklessness, including organisational fault
3	Harm arises from genuinely unforeseeable emergent behaviour despite compliance with prescribed standards	No criminal liability; strict civil and administrative liability, mandatory insurance and compensation fund

Legislative reforms

1. **Define "artificial intelligence system" and "high-risk AI system"** in a dedicated statute or in amendments to the IT Act, and expressly state that an AI system is not a legal person for criminal purposes.
2. **Create specific offences** in the BNS or the IT Act for: knowingly deploying a prohibited or unregistered high-risk system; failing to carry out mandatory safety testing, risk assessment and incident reporting; and creating or circulating malicious deepfakes with intent to cause harm, with graded penalties.
3. **Introduce a statutory duty of care** for providers and deployers of high-risk systems (human oversight, logging, bias testing, cybersecurity), the breach of which can be evidence of gross negligence.
4. **Adopt a "failure to prevent" offence** and a systems-based test of corporate fault, with turnover-linked fines, disqualification of directors and compliance-programme defences. Vicarious liability of senior officers should be expressly provided and limited to those in control.
5. **Amend the Motor Vehicles Act, 1988** to define the responsibility of manufacturers and operators when a vehicle is in autonomous mode, as in the UK's Automated Vehicles Act 2024.
6. **Strengthen safe-harbour conditions** in s. 79 of the IT Act for AI-generated content, linking immunity to labelling, watermarking and takedown compliance.

Procedural and evidentiary reforms

1. Mandatory **audit trails and record retention** for high-risk AI, with court-ordered access under confidentiality protections to balance trade secrets and fair trial.
2. A panel of **court-appointed technical experts** and guidelines for admissibility of algorithmic evidence under the BSA, 2023.
3. Specialised cyber-forensic units and judicial and police training on AI-enabled crime.

Institutional and policy measures

1. An independent **AI Safety and Accountability Authority** (or an empowered existing regulator) to register high-risk systems, investigate incidents, and refer cases for prosecution.

2. **Sectoral regulations** for health, finance, policing and transportation with human-in-the-loop requirements, and a ban or strict regulation of real-time facial recognition without legislative backing and proportionality review, as required by *Puttaswamy*.
3. **Compulsory insurance and a compensation fund** for victims in Tier 3 cases, so that the absence of criminal liability does not mean the absence of a remedy.
4. A **Law Commission of India study** and public consultation before enacting any AI-specific criminal provisions, to avoid vagueness and over-criminalisation.

Conclusion

The paper began by asking whether the traditional principles of Indian criminal jurisprudence can cope with harm caused by AI. The answer is: *partly*. Where AI is merely the tool of a culpable human, the BNS and the IT Act function adequately. Where harm results from autonomous, opaque or emergent behaviour, the architecture of mens rea, causation and corporate attribution leaves a genuine responsibility gap.

In response to the research questions: (1) AI cannot be an offender or a "person" under current law, and should not be made one; (2) liability should rest with developers, providers, deployers and users according to their foreseeability, control and benefit; (3) abetment, common intention and the identification doctrine are insufficient on their own; (4) comparative models support risk-based regulation and organisational fault rather than machine personhood; and (5) closing the gap requires a graduated liability model, specific duty-of-care offences, procedural reform and a dedicated regulator, all framed within Articles 14, 20 and 21.

The guiding principle should be that **technological autonomy must never become a shield against human accountability**, while criminal law, as the harshest sanction, must remain confined to demonstrable fault. Rather than abandoning the foundations of criminal jurisprudence, India should adapt them: preserving the requirement of a guilty mind, but recognising that in the age of AI, culpability may lie in the failure to anticipate, test, supervise and disclose.

REFERENCES:

Constitution and Statutes

1. The Constitution of India, arts. 14, 19, 20, 21.
2. Bharatiya Nyaya Sanhita, 2023; Bharatiya Nagarik Suraksha Sanhita, 2023; Bharatiya Sakshya Adhiniyam, 2023.
3. Indian Penal Code, 1860 (repealed).
4. Information Technology Act, 2000, and Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.
5. Digital Personal Data Protection Act, 2023.
6. Consumer Protection Act, 2019; Motor Vehicles Act, 1988; Medical Device Rules, 2017.

Cases

7. *Maneka Gandhi v. Union of India*, (1978) 1 SCC 248.
8. *State of Maharashtra v. Mayer Hans George*, AIR 1965 SC 722.
9. *Nathulal v. State of Madhya Pradesh*, AIR 1966 SC 43.
10. *Sham Sunder v. State of Haryana*, (1989) 4 SCC 630.
11. *Shiromani Gurdwara Parbandhak Committee v. Som Nath Dass*, (2000) 4 SCC 146.

12. *Standard Chartered Bank v. Directorate of Enforcement*, (2005) 4 SCC 530.
13. *Iridium India Telecom Ltd. v. Motorola Inc.*, (2011) 1 SCC 74.
14. *Aneeta Hada v. Godfather Travels and Tours Pvt. Ltd.*, (2012) 5 SCC 661.
15. *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.
16. *Sunil Bharti Mittal v. CBI*, (2015) 4 SCC 609.
17. *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1.
18. *Mohd. Salim v. State of Uttarakhand*, WPPIL No. 126 of 2014 (Uttarakhand HC, 2017).
19. *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637.
20. *Jaswinder Singh v. State of Punjab* (Punjab & Haryana HC, 2023); *Christian Louboutin SAS v. Abubaker Shaikh* (Delhi HC, 2023); *Anil Kapoor v. Simply Life India* (Delhi HC, 2023).
21. *Tesco Supermarkets Ltd v. Nattrass* [1972] AC 153 (HL).

Books and Articles

22. Abbott, R. and Sarch, A., "Punishing Artificial Intelligence: Legal Fiction or Science Fiction" (2019) 53 UC Davis Law Review 323.
23. Bryson, J., Diamantis, M. and Grant, T., "Of, For, and By the People: The Legal Lacuna of Synthetic Persons" (2017) 25 Artificial Intelligence and Law 273.
24. Elish, M.C., "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction" (2019) 5 Engaging Science, Technology, and Society 40.
25. Gless, S., Silverman, E. and Weigend, T., "If Robots Cause Harm, Who Is to Blame? Self-Driving Cars and Criminal Liability" (2016) 19 New Criminal Law Review 412.
26. Hallevy, G., *When Robots Kill: Artificial Intelligence under Criminal Law* (Northeastern University Press, 2013).
27. Hallevy, G., *Liability for Crimes Involving Artificial Intelligence Systems* (Springer, 2015).
28. Matthias, A., "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata" (2004) 6 Ethics and Information Technology 175.
29. Ratanlal and Dhirajlal, *The Indian Penal Code* and commentaries on the Bharatiya Nyaya Sanhita (latest editions).
30. Russell, S. and Norvig, P., *Artificial Intelligence: A Modern Approach* (4th edn, Pearson, 2020).

Reports, Policy Documents and International Instruments

31. NITI Aayog, *National Strategy for Artificial Intelligence: #AIforAll* (2018); *Responsible AI: Approach Document for India* (2021).
32. Ministry of Electronics and Information Technology, advisories on deepfakes and AI models (2023–2024) and India AI governance guidelines (2025).
33. European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).
34. Regulation (EU) 2024/1689 (Artificial Intelligence Act); Directive (EU) 2024/2853 (Product Liability).
35. Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (2024).
36. UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021); OECD, Recommendation of the Council on Artificial Intelligence (2019, as amended 2024).
37. United Kingdom, Automated Vehicles Act 2024; Economic Crime and Corporate Transparency Act 2023.