CINEFORUM
ISSN: 0009-7039
Vol. 66. No. S8, 2026 (Special Issue S8)

Artificial Intelligence and Criminal Liability in India: Accountability and Attribution under the Bharatiya Nyaya Sanhita, 2023

Dr. Rehana Shawl[1]; Deeba Ashraf[2]; Divya Sehrawat[3]
[1]Assistant Professor ,School of Legal Studies, Central University of Kashmir
ORCID : https://orcid.org/0009-0004-3789-1959
Email: rehana.shawl@gmail.com
[2]Legal Consultant @ Law Commission of India
ORCID : https://orcid.org/0009-0005-6216-4523 ,
Email: ddeebaashraf@gmail.com
[3]Email : divyasehrawat29@gmail.com

ABSTRACT
The Bharatiya Nyaya Sanhita, 2023 (BNS) can already punish almost every harm committed through artificial intelligence, but it cannot coherently attribute harms that emerge from an AI system's autonomous behaviour without a human who intended or foresaw them. This paper calls that space the attribution gap and maps its boundaries.
Using doctrinal analysis of the BNS, the Information Technology Act, 2000, the Bharatiya Sakshya Adhiniyam, 2023 and the Bharatiya Nagarik Suraksha Sanhita, 2023, read with Supreme Court authority on corporate mens rea, the paper tests Gabriel Hallevy's three models of AI criminal liability against Indian law. It situates the analysis in empirical context: NCRB recorded 1,01,928 cybercrime cases in 2024, the first year above one lakh, and citizens reported losses of ₹22,845.73 crore to cyber fraud in 2024 alone.
The paper finds that (i) the "perpetration-via-another" model is fully available because an AI system is not a "person" under section 2(26) and so functions as a tool; (ii) the "natural-probable-consequence" model is only partly available, because BNS abetment and conspiracy provisions presuppose a human or juristic co-actor; and (iii) direct liability of the AI system itself is neither possible nor desirable. The gap falls hardest on developers and deployers of autonomous systems, and on companies whose harm arises from diffuse organisational fault rather than any one directing mind.
It proposes a graded attribution framework consistent with MeitY's India AI Governance Guidelines (November 2025): a deeming rule for deployer causation, a reckless-deployment offence for high-risk systems, and a corporate "failure to prevent" duty with a due-diligence defence.

Keywords: artificial intelligence; criminal liability; attribution; mens rea; Bharatiya Nyaya Sanhita 2023; corporate criminal liability; deepfakes; synthetically generated information; India AI Governance Guidelines

1 INTRODUCTION

Indian criminal law was recodified for the first time in 164 years when the BNS replaced the Indian Penal Code, 1860 on 1 July 2024. The new code modernised definitions of "document" and forgery to include electronic records and added organised crime that expressly names "cyber-crimes". It did not, however, address the agent that now sits between many offenders and their victims: an AI system that generates, decides or acts with some degree of autonomy.
The question is not academic. Voice-cloned "digital arrest" calls, deepfake investment endorsements, synthetic intimate imagery and AI-assisted phishing now account for a growing share of reported cyber fraud. Police routinely register these cases under BNS sections on cheating, personation, forgery and defamation together with sections 66C, 66D and 67 of the IT Act. Those charges work when a human chose the harmful output. They strain when nobody did.

1.1 Statement of the problem

Criminal liability in India rests on a voluntary act (or illegal omission) by a **person**, accompanied by a culpable mental state that the offence requires. AI systems challenge both pillars. They are not persons, so they cannot carry liability; and their outputs may be neither intended nor specifically foreseen by any person, so human liability may fail for want of mens rea. The result can be harm without a culpable author, a "responsibility gap" in the sense first described by Matthias (2004).

1.2 Research questions

To what extent can existing BNS provisions attribute AI-mediated harm to a human or juristic person?
Which liability model (perpetration-via-another, natural-probable-consequence, or direct liability) does Indian law support, and where does each fail?

OPEN ACCESS © CINEFORUM

How should liability be apportioned among developers, deployers, users and intermediaries in the AI value chain? What targeted legislative reform would close the attribution gap without chilling innovation or diluting the requirement of personal fault?

1.3 Hypothesis

The BNS is **adequate for AI as instrument** but **inadequate for AI as quasi-autonomous agent**. The inadequacy is concentrated in three places: mens rea for unforeseen emergent outputs, causation where many actors each contribute a small risk, and corporate liability where fault is organisational rather than personal.

1.4 Methodology

The study is primarily **doctrinal**, reading the text of the BNS, IT Act, BSA and BNSS with leading Supreme Court decisions. It is **comparative**, drawing on the EU AI Act, recent UK and US offences on synthetic intimate imagery, and China's deep-synthesis and labelling rules. It uses **secondary empirical data** from NCRB Crime in India (2023 and 2024 editions) and Ministry of Home Affairs replies to Parliament on I4C fraud reporting, to establish the scale of the problem. Data are reported as published; NCRB figures count FIRs registered, not convictions, and record no separate AI category.

1.5 Scope and limitations

The paper addresses substantive criminal liability. It touches procedure and evidence only where they decide whether attribution is provable. It does not address civil liability, intellectual property, or autonomous weapons. Because neither NCRB nor I4C publishes an "AI-enabled" category, empirical claims about AI's share of crime are necessarily inferential.

2. LITERATURE REVIEW AND THEORETICAL FRAMEWORK

The scholarship divides on one question: should the law stretch human-centred doctrines to reach AI-mediated harm, or build new ones? Most writers favour stretching for now; a minority argue that sufficiently autonomous systems will eventually need direct regulation of their own.

2.1 The responsibility gap

Matthias (2004) argued that learning machines create situations where neither the manufacturer nor the operator can predict the machine's behaviour, and so neither can fairly be held responsible. Santoni de Sio and Mecacci (2021) refined this into four distinct gaps: culpability, moral accountability, public accountability and active responsibility. For criminal law the **culpability gap** is decisive, because punishment without fault offends both Article 21 and the general scheme of the BNS.

2.2 Hallevy's three models

Hallevy (2010) offered the framework most often applied in comparative work.

Model	Core idea	Who is liable	Mental element required
Perpetration-via-another	AI is an innocent agent, like a child or a tool	Programmer or user who directed it	Intent or knowledge of the directing human
Natural-probable-consequence	Human did not intend the offence but it was a foreseeable result of the system's use	Programmer or user	Negligence or foreseeability
Direct liability	AI itself satisfies actus reus and a functional mens rea	The AI entity (with human liability in parallel)	Functional "knowledge" attributed to the system

The first two models keep liability with humans. The third requires treating the machine as a legal subject, which Hallevy accepted but most later commentators reject.

2.3 Critiques of direct liability

Abbott and Sarch (2019) concluded that punishing AI is conceivable but rarely justified, since its expressive and deterrent benefits can usually be achieved by punishing humans or reforming civil law. Lima (2018) reached a similar view. The European Parliament's 2017 resolution floating "electronic personhood" drew wide criticism and was not adopted into the EU AI Act. Bryson, Diamantis and Grant (2017) warned that machine personhood would let human actors shelter behind a judgment-proof legal entity, a concern that applies with special force in India, where corporate criminal enforcement is already weak.

2.4 Indian scholarship and policy

OPEN ACCESS © CINEFORUM

Indian writing has been largely policy-led. NITI Aayog's Responsible AI for All papers (2021) emphasised accountability but did not address criminal attribution. MeitY's India AI Governance Guidelines (5 November 2025) recommend a **graded liability system** linked to function performed, level of risk and due diligence observed, and call for clear definitions of developers, deployers and users. The Guidelines create no offences; they rely on existing law and acknowledge that causation and foreseeability become harder as autonomy grows. Doctrinal commentary on the BNS has concentrated on deepfakes and on the breadth of section 111, not on the general part of criminal law.

2.5 Analytical framework adopted

This paper treats attribution as two separate questions. **Causal attribution** asks whether a person's conduct can be said to cause the AI-mediated harm. **Fault attribution** asks whether that person had the mental state the specific offence requires. Hallevy's models are then tested against BNS text on both questions, for each actor in the AI value chain. The framework is summarised in Figure 3 (section 5).

3. EMPIRICAL CONTEXT: THE SCALE OF AI-ADJACENT CYBERCRIME IN INDIA

Cybercrime is the fastest-growing category of registered crime in India, and its growth is concentrated in fraud by deception, the category most transformed by generative AI. NCRB recorded 1,01,928 cybercrime cases in 2024, up 17.9% on 2023, in a year when total cognisable crime fell 6%.

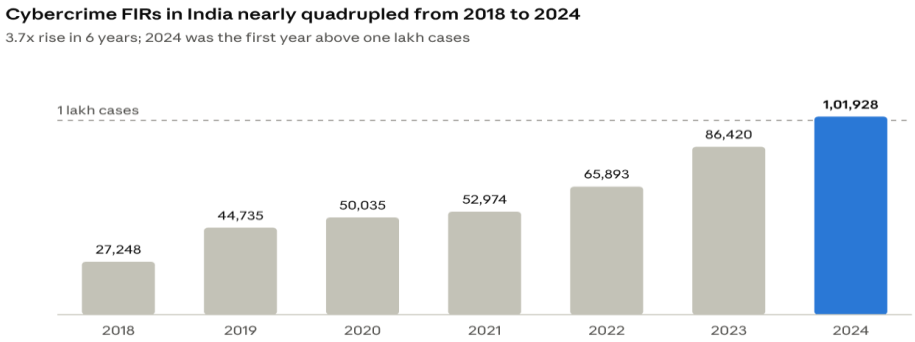


Figure 1. NCRB Crime in India, 2018 to 2024 editions · cybercrime cases registered

Fraud is the dominant motive: 68.9% of 2023 cybercrime cases (59,526 of 86,420) and 72.6% in 2024 (73,987 cases). Registration is also uneven. Karnataka (21,889) and Telangana (18,236) together accounted for 46% of 2023 cases, while charge-sheeting rates in those states were only 18.1% and 20.9%, which points to an investigative bottleneck rather than only a crime wave.

Complaint data from I4C show a much larger problem than FIRs capture. Citizens reported 36.37 lakh financial fraud incidents in 2024 against 24.43 lakh in 2023, and losses rose from ₹7,465.18 crore to ₹22,845.73 crore, a 206% increase. Of losses in the first nine months of 2024, the categories most amenable to AI-generated personas and endorsements dominate.

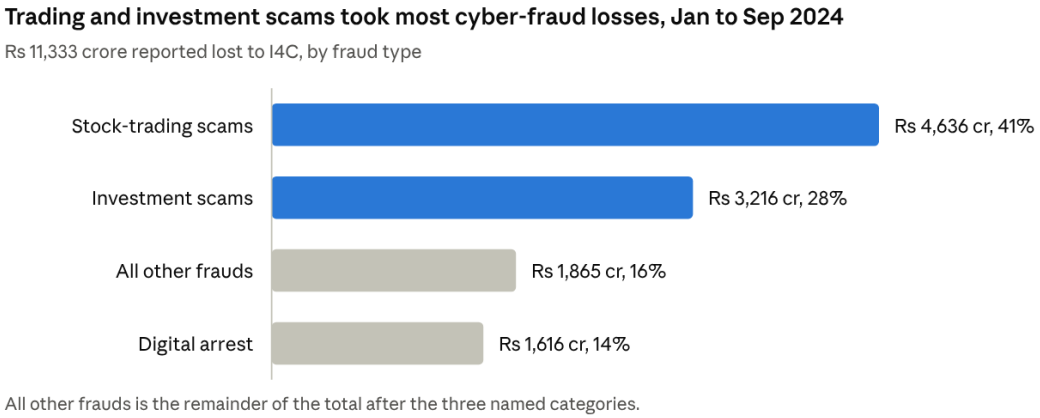


Figure 2. I4C / CFCFRMS data, January to September 2024, as reported by The Wire

OPEN ACCESS © CINEFORUM

Interpretive caution. Neither NCRB nor I4C records whether AI was used. Investment-scam "advisers", fake broker apps and "digital arrest" video calls increasingly rely on synthetic voices, faces and scripted chat, but their AI share cannot be measured from public data. This measurement gap is itself a finding: the law cannot calibrate AI-specific liability without AI-specific statistics (see recommendation 8.3).

4. DOCTRINAL ANALYSIS UNDER THE BNS, 2023

The BNS reaches AI-mediated harm through its general definitions, its rules on joint and inchoate liability, and a cluster of specific offences. It works well when a human chooses the harmful output and poorly when the output emerges from the system.

4.1 Legal personhood: section 2(26)

Section 2(26) defines "person" to include any company or association or body of persons, whether incorporated or not. The definition is inclusive, but nothing in the text, its legislative history, or the Supreme Court's corporate cases suggests it extends to software. An AI system therefore cannot be an accused, an abettor or a conspirator. Hallevy's direct-liability model is unavailable in India as a matter of positive law, and this paper argues below that it should stay that way.

4.2 Actus reus: AI as instrument

Because the AI is not a person, its outputs are legally the acts of whoever used it, in the same way that a forged cheque is the act of the forger and not the printer. Section 2(33) treats a person as causing an effect "voluntarily" when it is caused by means the person intended to use, or knew or had reason to believe were likely to cause it. A user who prompts a model to clone a relative's voice and then calls the victim commits cheating by personation under section 319 directly; the model is the means.

This is Hallevy's perpetration-via-another model, and Indian law supports it without strain. Explanation 3 to section 46 confirms that the person through whom an offence is committed need not be capable of committing it. For AI the point is stronger still: there is no second "person" at all, so the human is a principal rather than an abettor.

4.3 Mens rea: the core difficulty

Most BNS offences relevant to AI require a specific mental state: "dishonestly" (s. 2(7)) or "fraudulently" (s. 2(9)) for cheating and forgery, intention or knowledge of likely harm to reputation for defamation (s. 356), or intention for criminal intimidation (s. 351). Three situations arise.

Directed misuse. The user intends the harmful output. Mens rea is easy to prove; the AI is irrelevant to fault.

Foreseeable misuse by others. A developer releases a voice-cloning tool knowing it is widely used for fraud. Knowledge of general misuse is not knowledge of a particular victim. Unless the developer joined the scheme, no specific-intent offence is made out against them.

Emergent harm. An autonomous trading or content agent produces a defamatory statement or a manipulative transaction that no human requested. No one had the specific mental state. Only negligence-based offences remain.

The negligence offences are few. Section 106(1) punishes causing death by a rash or negligent act, and section 125 punishes rash or negligent acts that endanger life or personal safety. Following Jacob Mathew v. State of Punjab (2005) 6 SCC 1, criminal negligence must be gross, not a mere failure of reasonable care. There is no BNS offence of negligently causing financial loss, reputational harm or a privacy violation. Emergent AI harms of those kinds therefore fall outside the criminal law entirely.

4.4 Joint liability, abetment and conspiracy

Section 3(5) (common intention), sections 45 to 60 (abetment) and section 61 (criminal conspiracy) are the natural tools for reaching developers and platforms. Each, however, requires a meeting of minds or an intentional aid between persons. Section 51, which makes an abettor liable for a different act that was a probable consequence of the abetment, is the closest the BNS comes to Hallevy's natural-probable-consequence model. It is unavailable unless a human abetment exists in the first place.

A developer who markets a model as "undetectable for KYC bypass" may be an abettor by intentional aid. A developer who merely releases a general-purpose model is not. This is the right result, but it leaves the space between those poles unregulated.

4.5 Organised crime: section 111

Section 111 brings "cyber-crimes" and "economic offence" (including mass-marketing fraud) within organised crime when committed as continuing unlawful activity by a syndicate. It carries severe punishment and attaches liability to members, abettors and those who harbour or possess proceeds. AI-scaled fraud networks fit comfortably. The term "cyber-crimes" is undefined, however, and courts have stressed strict construction.

OPEN ACCESS © CINEFORUM

Whether a technology vendor whose tools are reused by a syndicate is a "member" or facilitator will turn on proof of knowledge, not mere supply.

4.6 Specific offences engaged by AI misuse

AI-enabled conduct	BNS provisions	Typical IT Act companion
Voice-clone or video "digital arrest" fraud	s. 318 cheating; s. 319 cheating by personation; s. 308 extortion	s. 66C identity theft; s. 66D personation by computer resource
Deepfake defamation of a public figure	s. 356 defamation; s. 336 forgery; s. 340 using forged electronic record; s. 353 public mischief	s. 66D
Synthetic intimate imagery	s. 77 voyeurism; s. 78 stalking; s. 79 insulting modesty; s. 294 obscene material	s. 66E; s. 67; s. 67A
AI-generated disinformation threatening unity	s. 152 acts endangering sovereignty; s. 196 promoting enmity; s. 197 imputations	s. 69A blocking
Autonomous system causing physical harm	s. 106(1) death by negligence; s. 125 endangering life	—
Syndicated AI-scaled fraud	s. 111 organised crime; s. 61 conspiracy	ss. 43, 66

Recent FIRs follow exactly this pattern. Deepfake videos of political leaders in 2025 and 2026 were registered under sections 336, 340(2), 352, 356(2) and 61(2), and an AI-generated imagery case against a Kannada actress in June 2026 combined sections 75(3), 79, 336, 356, 351, 352 and 294 with sections 66C and 66D of the IT Act. Every such FIR names unknown human users. None names a developer.

4.7 Corporate criminal liability

Companies can be prosecuted for offences requiring mens rea. Standard Chartered Bank v. Directorate of Enforcement (2005) 4 SCC 530 held that a company may be prosecuted even where imprisonment is mandatory, with fine substituted. Iridium India Telecom v. Motorola (2011) 1 SCC 74 adopted the identification doctrine: the mens rea of those who are the company's "directing mind and will" is attributed to it. Sunil Bharti Mittal v. CBI (2015) 4 SCC 609 held the reverse: a director is not liable for the company's offence merely because of office, absent a statutory deeming provision.

This framework fits AI poorly. AI harms in large firms usually arise from distributed decisions on data, model design, testing and deployment. No single senior officer may know enough to supply the company's mens rea. The identification doctrine then shields precisely the organisations best placed to prevent harm. Section 85 of the IT Act does deem officers in charge liable for IT Act offences committed by a company, with a due-diligence defence, but it does not extend to BNS offences.

4.8 General exceptions

Two defences will be pleaded by developers and deployers. Section 14 and section 17 (mistake of fact in good faith) may cover a deployer who reasonably believed an AI's output was accurate. Section 18 (accident in doing a lawful act with proper care) may cover unforeseeable malfunction. Both turn on **proper care and caution**, which is why documented testing, red-teaming and labelling become, in practice, the measure of criminal exposure.

5. ATTRIBUTION ACROSS THE AI VALUE CHAIN

Applying the two-step test (causal attribution, then fault attribution) produces a single decision path. Liability is secure only on the first two branches, where a human intended the output or acted with another person. Everything else depends on negligence, and the BNS punishes negligence only for physical harm.

OPEN ACCESS © CINEFORUM

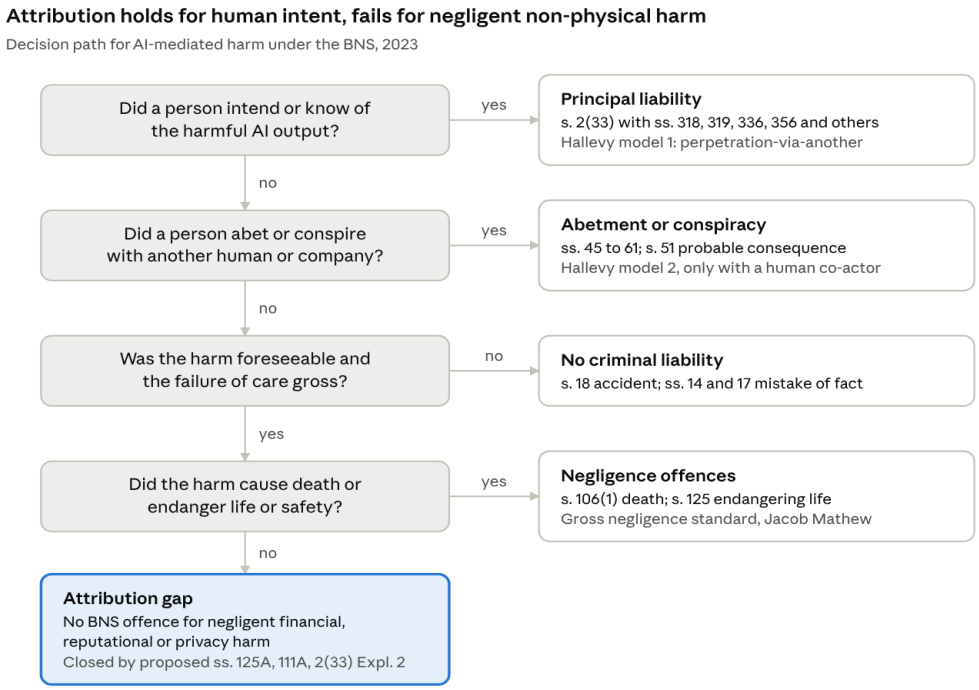


Figure 3. Attribution decision path, author's synthesis of BNS ss. 2(33), 14, 17, 18, 45 to 61, 106, 125

The accented box is where most emergent AI harms land: a model or agent produces fraud, defamation or a privacy violation that no one specifically intended, and the harm is not physical.

5.1 Actor-wise attribution matrix

Actor	Typical AI-related conduct	Strongest BNS route	Main hurdle	Gap rating
End user	Prompts a model to clone a voice or create a deepfake, then uses it	Principal liability, ss. 318, 319, 336, 356	Identification and proof of authorship	Low
Deployer (bank, hospital, platform operator)	Puts an AI system into a customer-facing or safety role	ss. 106(1), 125 if physical harm; corporate liability via Iridium	No negligence offence for financial or data harm; diffuse corporate fault	High
Developer	Builds and releases a general-purpose or specialised model	Abetment by intentional aid (s. 45) if marketed for crime	Knowledge of general misuse is not knowledge of a specific offence	High
Intermediary	Hosts and distributes AI-generated content	Loss of IT Act s. 79 safe harbour, then underlying offence	Actual-knowledge standard after Shreya Singhal	Medium
Syndicate	Uses AI at scale for continuing fraud	s. 111 organised crime; s. 61 conspiracy	Undefined "cyber-crimes"; strict construction	Low to medium
AI system	Generates the harmful output	None: not a "person" under s. 2(26)	Not a legal subject	Not applicable

The matrix shows an inverse relationship: the actors with the greatest capacity to prevent AI harm at scale, developers and large deployers, face the weakest criminal exposure. That is the attribution gap in its institutional form.

6. INTERPLAY WITH ALLIED STATUTES AND POLICY

The BNS does not operate alone. India's AI-related criminal exposure is built from five instruments, and the most important recent change, the 2026 synthetic-content rules, sits outside the BNS altogether.

OPEN ACCESS © CINEFORUM

6.1 Information Technology Act, 2000 and the 2026 SGI Rules

The IT Act supplies the technology-specific offences (ss. 66C, 66D, 66E, 67, 67A, 67B) and the intermediary safe harbour in section 79. After Shreya Singhal v. Union of India (2015) 5 SCC 1, an intermediary loses safe harbour only on actual knowledge through a court order or government notification.

On 10 February 2026 MeitY notified the IT (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (G.S.R. 120(E)), in force from 20 February 2026. They define **synthetically generated information** (SGI) as audio, visual or audio-visual content artificially or algorithmically created or altered so as to appear real. Platforms must label such content prominently, attach provenance metadata, and deploy technical measures against unlawful SGI. Text-only output falls outside the definition.

The Rules matter for criminal attribution in two ways. First, a platform that ignores its SGI duties risks losing section 79 protection and becoming exposed to the underlying BNS offence. Second, labels and metadata create evidentiary trails that make proof of human authorship more feasible. They do not, however, regulate model developers who are not intermediaries.

6.2 Digital Personal Data Protection Act, 2023

The DPDP Act imposes civil penalties of up to ₹250 crore and contains no imprisonment. It governs the training data and personal data processed by AI systems but creates no criminal offence. Misuse of personal data to build a deepfake is prosecuted, if at all, under the BNS and IT Act.

6.3 Bharatiya Nagarik Suraksha Sanhita, 2023

The BNSS makes forensic examination at the crime scene mandatory for offences punishable with seven years or more (s. 176(3)) and permits summons for electronic records (s. 94). For AI cases the bottleneck is capacity: state forensic laboratories rarely have validated deepfake-detection tools, and model logs typically sit with foreign providers outside the reach of a section 94 summons.

6.4 Bharatiya Sakshya Adhiniyam, 2023

Section 63 of the BSA (formerly s. 65B of the Evidence Act) governs admissibility of electronic records through a certificate. Proving who prompted a model, what the model generated, and whether it was later altered requires chain-of-custody evidence that ordinary certificates were not designed to capture. Provenance metadata under the 2026 Rules may help, if courts accept it as reliable.

6.5 India AI Governance Guidelines, November 2025

The Guidelines are non-binding. They recommend defining roles across the AI value chain, a graded liability system proportionate to function and risk, and targeted amendments to existing laws on classification of digital entities, apportionment of liability and safe harbour. They expressly create no new penalties. This paper's proposals in section 8 are designed to give that policy direction a criminal-law form.

Instrument	Status	Bearing on attribution
BNS, 2023	In force 1 Jul 2024	Substantive offences; person; mens rea; corporate liability
IT Act, 2000	In force	Cyber offences; s. 79 safe harbour; s. 85 officers' liability
IT Amendment Rules, 2026 (SGI)	In force 20 Feb 2026	Labelling, metadata, platform due diligence
DPDP Act, 2023	Phased commencement	Data duties; civil penalties only
BNSS, 2023 / BSA, 2023	In force 1 Jul 2024	Forensics, electronic evidence, admissibility
India AI Governance Guidelines	Issued 5 Nov 2025, non-binding	Graded liability; value-chain roles

7. COMPARATIVE PERSPECTIVES

No major jurisdiction has made AI systems criminally liable. All four examined here keep liability with humans and organisations, and differ mainly in how far they regulate upstream developers.

OPEN ACCESS © CINEFORUM

Jurisdiction	Main instrument	Approach to liability	Lesson for India
European Union	AI Act, Regulation (EU) 2024/1689; revised Product Liability Directive (EU) 2024/2853	Risk-tiered ex ante duties on providers and deployers; administrative fines; criminal law left to member states; proposed AI Liability Directive withdrawn in 2025	Role definitions (provider, deployer) and risk tiers give criminal law a duty to anchor negligence to
United Kingdom	Online Safety Act 2023; Data (Use and Access) Act 2025; Economic Crime and Corporate Transparency Act 2023	Specific offences for sharing and creating sexually explicit deepfakes; corporate "failure to prevent fraud" offence with reasonable-procedures defence	A failure-to-prevent model bypasses the identification doctrine without strict liability
United States	TAKE IT DOWN Act (2025); state deepfake statutes	Federal offence for publishing non-consensual intimate imagery including AI-generated depictions; platform takedown duty	Narrow, harm-specific offences avoid overbreadth concerns
China	Deep Synthesis Provisions (2023); AI-Generated Content Labelling Measures (in force 1 Sep 2025)	Mandatory labelling, real-name verification and provider security duties, enforced administratively and through criminal law	India's 2026 SGI Rules follow a similar labelling logic but stop at intermediaries

Two lessons stand out. First, the EU's role vocabulary of provider and deployer gives criminal negligence a concrete standard of care to measure against; India's Guidelines point the same way but have no statutory footing yet. Second, the UK's failure-to-prevent offence, in force for fraud since 1 September 2025, offers a tested route around the identification doctrine. It makes a large organisation liable when an associated person commits fraud for its benefit, unless it had reasonable prevention procedures. Adapted to AI deployment, that structure targets organisational fault directly.

8. GAPS AND RECOMMENDATIONS

Five gaps follow from the analysis, and each can be closed without granting AI personhood or imposing strict criminal liability.

#	Gap	Where it bites	Recommended response
G1	No negligence offence for non-physical harm	Emergent AI fraud, defamation, privacy harm	Offence of reckless deployment of a high-risk AI system
G2	Identification doctrine shields organisations	Large developers and deployers	Corporate failure-to-prevent duty with due-diligence defence
G3	Causation diluted across the value chain	Multi-actor harms	Deeming rule on deployer causation, rebuttable
G4	Section 111 "cyber-crimes" undefined	Over- and under-inclusion of AI syndicates	Statutory definition cross-referencing IT Act offences
G5	Forensic and evidentiary capacity	Proof of authorship and alteration	Provenance standards admissible under BSA s. 63; FSL deepfake units

8.1 Proposed model provisions

The following text is offered as a drafting basis for discussion, structured as amendments to the BNS.

Section 2(26A) — "AI system" means a machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as content, predictions, recommendations or decisions that can influence physical or virtual environments.

OPEN ACCESS © CINEFORUM

Section 2(33), Explanation 2 — Where an effect is caused by an AI system, a person who deployed or operated that system for the purpose in the course of which the effect was caused shall be deemed to have caused it, unless that person proves that the effect could not have been prevented by the exercise of proper care and caution.
Section 125A. Reckless deployment of a high-risk AI system — Whoever, being a developer or deployer, places in operation an AI system notified as high-risk, with knowledge of a substantial and unjustifiable risk that it will cause death, grievous hurt, or wrongful loss exceeding a notified threshold, and without taking measures notified under the IT Act, shall be punished with imprisonment up to three years, or fine, or both.
Section 111A. Failure to prevent AI-enabled offence by an organisation — Where an offence under this Sanhita is committed by means of an AI system developed or deployed by a company, and a person associated with the company intended to benefit it, the company shall be guilty of an offence punishable with fine, unless it proves that it had reasonable procedures in place to prevent such use.

8.2 Design safeguards

Fault is preserved. Every proposal requires knowledge, recklessness or a failure of proper care; none imposes strict liability on natural persons.
Risk is graded. Only notified high-risk systems attract section 125A, mirroring the Guidelines' graded liability.
Defences are built in. Reasonable procedures and proper care are complete defences, rewarding documented testing and labelling.
No machine personhood. The AI system remains an instrument; liability attaches only to persons in the value chain.
Open-source carve-out. Release of model weights without deployment should not by itself satisfy "deploys", to avoid chilling research.

8.3 Institutional recommendations

NCRB should add an "AI-enabled" flag to its cybercrime tables so that policy can be evidence-based.
I4C should publish annual statistics on deepfake and voice-clone complaints separately from generic fraud.
The Law Commission of India should undertake a reference on AI and the general part of criminal law, as its counterpart in England and Wales has begun to do.
State FSLs should receive validated deepfake-detection capacity within the BNSS s. 176(3) forensic mandate.

9. CONCLUSION AND DIRECTIONS FOR FUTURE RESEARCH

The hypothesis is confirmed. The BNS handles AI as an instrument with little difficulty: section 2(26) keeps the machine outside personhood, section 2(33) makes its outputs the voluntary acts of those who use it, and the cheating, personation, forgery and defamation offences map directly onto deepfake and voice-clone harms. The FIRs registered since 2024 show that police already use this toolkit.
The BNS fails where AI behaves as a quasi-autonomous agent. Specific-intent offences cannot reach developers and deployers who foresaw misuse only in general; negligence offences cover physical harm alone; abetment and conspiracy need a human co-actor; and the identification doctrine lets organisations escape when fault is distributed. Hallevy's first model works in India, his second works only partly, and his third should be rejected.
The answer is not machine personhood but a sharper allocation of human and organisational duty. A rebuttable causation rule for deployers, a reckless-deployment offence confined to notified high-risk systems, and a corporate failure-to-prevent duty would give the Guidelines' graded liability a criminal-law form while keeping fault at the centre.
Future research should pursue four lines:
Empirical. Case-file studies of AI-related FIRs and charge-sheets, measuring how often developers or platforms are investigated and why prosecutions fail.
Evidentiary. The reliability of provenance metadata and AI-detection tools under BSA s. 63 and the Daubert-like scrutiny Indian courts apply to expert evidence.
Agentic systems. Liability where AI agents act across multiple platforms and jurisdictions, including under section 48 (abetment outside India).
Constitutional. Whether deemed-causation and reverse-burden provisions satisfy Articles 14, 20(3) and 21.

REFERENCES

1. Statutes, rules and policy
2. Bharatiya Nyaya Sanhita, 2023 (Act 45 of 2023).
3. Bharatiya Nagarik Suraksha Sanhita, 2023 (Act 46 of 2023).
4. Bharatiya Sakshya Adhiniyam, 2023 (Act 47 of 2023).
5. Information Technology Act, 2000 (Act 21 of 2000).
6. Digital Personal Data Protection Act, 2023 (Act 22 of 2023).

OPEN ACCESS © CINEFORUM

7. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E), 10 February 2026. Summaries: Business Today; Nishith Desai Associates.
8. Ministry of Electronics and Information Technology, India AI Governance Guidelines (5 November 2025). Analysis: Khaitan & Co ERGO; Mondaq.
9. NITI Aayog, Responsible AI for All, Parts 1 and 2 (2021).
10. Regulation (EU) 2024/1689 (Artificial Intelligence Act); Directive (EU) 2024/2853 (Product Liability).
11. Online Safety Act 2023 (UK); Data (Use and Access) Act 2025 (UK); Economic Crime and Corporate Transparency Act 2023 (UK).
12. TAKE IT DOWN Act, Pub. L. 119-12 (US, 2025).
13. Provisions on the Administration of Deep Synthesis of Internet Information Services (PRC, 2023); Measures for Labelling AI-Generated Synthetic Content (PRC, 2025).
14. Cases
15. Jacob Mathew v. State of Punjab (2005) 6 SCC 1.
16. Standard Chartered Bank v. Directorate of Enforcement (2005) 4 SCC 530.
17. Iridium India Telecom Ltd v. Motorola Inc. (2011) 1 SCC 74.
18. Aneeta Hada v. Godfather Travels & Tours (2012) 5 SCC 661.
19. Sunil Bharti Mittal v. CBI (2015) 4 SCC 609.
20. Shreya Singhal v. Union of India (2015) 5 SCC 1.
21. Data sources
22. National Crime Records Bureau, Crime in India 2023 (released 30 September 2025): Khan Global Studies summary; Millennium Post.
23. National Crime Records Bureau, Crime in India 2024 (released 6 May 2026): SPM IAS summary; The Wire.
24. Ministry of Home Affairs, Lok Sabha written reply on cyber fraud losses (22 July 2025): Angel One; YourStory.
25. I4C data on January–September 2024 fraud categories: The Wire.
26. Deepfake FIRs: Tribune India; WION; The Star.
27. Section 111 commentary: Bar & Bench.
28. Secondary literature
29. Abbott, R. and Sarch, A. (2019) 'Punishing Artificial Intelligence: Legal Fiction or Science Fiction', UC Davis Law Review 53, 323.
30. Bryson, J., Diamantis, M. and Grant, T. (2017) 'Of, for, and by the people: the legal lacuna of synthetic persons', Artificial Intelligence and Law 25, 273.
31. Hallevy, G. (2010) 'The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control', Akron Intellectual Property Journal 4, 171.
32. Lima, D. (2018) 'Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law', South Carolina Law Review 69, 677.
33. Matthias, A. (2004) 'The responsibility gap: Ascribing responsibility for the actions of learning automata', Ethics and Information Technology 6, 175.
34. Santoni de Sio, F. and Mecacci, G. (2021) 'Four Responsibility Gaps with Artificial Intelligence', Philosophy & Technology 34, 1057.

OPEN ACCESS © CINEFORUM