



The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2026

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International (CC-BY-NC-SA 4.0) License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium provided the original work is properly cited.

AI-POWERED CYBERATTACKS: THE NEED FOR A NEW LEGAL FRAMEWORK IN INDIA

~ Ujjayani Debnath

ABSTRACT

Artificial intelligence has changed the economics of cybercrime. Tasks that once demanded technical skill, time and coordination, such as drafting convincing phishing messages, cloning a voice, probing a network for weaknesses or mutating malware to avoid detection, can now be automated, scaled and delivered at negligible cost. India, with one of the largest digital populations in the world and a fast-expanding digital public infrastructure, is acutely exposed. This paper examines whether the existing Indian legal architecture, principally the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023 and the Digital Personal Data Protection Act, 2023, is adequate to address AI-powered cyberattacks. It argues that although current law can reach many of the resulting harms through technology-neutral provisions, it suffers from definitional gaps, an outdated model of attribution and intent, the absence of duties on those who build and deploy AI systems, and weak evidentiary and institutional machinery. Drawing on developments in the European Union, the United States, the United Kingdom and China, the paper proposes a calibrated, risk-based statutory framework that protects security without sacrificing innovation or constitutional freedoms.

I. INTRODUCTION

In early 2024, an employee in the Hong Kong office of a multinational engineering firm transferred approximately US\$25 million to fraudsters after attending a video conference in which the company's chief financial officer and several colleagues appeared to take part. Every other participant on the call was a digitally generated imitation.¹ The incident is instructive not merely because of its scale but because of what it reveals about the changed nature of

¹ Heather Chen & Kathleen Magramo, *Finance Worker Pays Out \$25 Million After Video Call with Deepfake 'Chief Financial Officer'*, CNN (Feb. 4, 2024).

cybercrime. No firewall was breached and no password was cracked. The attacker used artificial intelligence (AI) to counterfeit trust itself.

India's exposure to this new category of risk is considerable. The country has built a digital public infrastructure, including the Unified Payments Interface, Aadhaar-based authentication and DigiLocker, which touches the daily lives of hundreds of millions of people, many of them first-time digital users with limited experience of online deception. Official figures placed before Parliament show that the number of cyber security incidents handled by the Indian Computer Emergency Response Team (CERT-In) has risen steeply, from roughly fifty thousand in 2017 to well over a million in 2022.² At the same time, generative AI tools have moved out of research laboratories and into the hands of ordinary users. The capability that drafts a legal memorandum can just as easily draft a spear-phishing email in flawless Bengali, Hindi or English.

The legal question that follows is whether Indian law, drafted largely for a world in which a human being sits behind every cyberattack, can respond effectively to harms that are automated, scalable, adaptive and difficult to attribute. This paper contends that it cannot do so adequately. The existing statutes are technology-neutral enough to punish many end results, such as fraud, identity theft and data breach, but they say nothing about the distinctive features of AI-enabled attacks and place no obligations on the actors best placed to prevent them. The analysis is doctrinal and comparative. It examines statutes, rules, executive directions and judicial decisions, and compares the Indian position with selected foreign and international regimes. Part II sets out the conceptual framework; Part III surveys the Indian threat landscape; Part IV evaluates existing law and identifies its gaps; Part V draws comparative lessons; Part VI proposes a new framework; and Part VII concludes.

II. UNDERSTANDING AI-POWERED CYBERATTACKS

A. CONCEPTUAL FRAMEWORK

There is no settled legal definition of an "AI-powered cyberattack". For the purposes of this paper, the expression refers to any unauthorised or malicious operation against, or through, a computer resource in which an AI system, and in particular a machine-learning model, is used to plan, execute, scale or conceal the operation. Three situations deserve to be distinguished. In the first, AI is an *instrument* that amplifies a human attacker's capabilities. In the second, AI is

² Press Info. Bureau, Ministry of Electronics & Info. Tech., *Cyber Security Incidents* (replies to Parliamentary questions, various years) (reporting CERT-In incident figures).

the *target*, as where an adversary poisons training data or manipulates a model's outputs. In the third, AI operates with a degree of *autonomy*, selecting targets or adapting its methods without step-by-step human direction. The taxonomy matters because each situation strains a different part of the legal framework.³

B. PRINCIPAL MODALITIES

The first and most widespread modality is AI-assisted social engineering. Large language models can generate grammatically polished, contextually tailored messages at scale, removing the spelling errors and awkward phrasing that once helped recipients identify fraud. Europol has warned that such models lower the barrier to entry for phishing and fraud, while underground services marketed as WormGPT and FraudGPT have been advertised specifically for business email compromise.⁴⁵

The second modality is synthetic media. Voice-cloning and face-swapping tools allow an attacker to impersonate a relative, an employer or a public official using only a few seconds of source audio or a handful of photographs. The Hong Kong episode described above is the best-known example, but the same technique is routinely used in “family emergency” frauds and in the creation of non-consensual intimate imagery. The third modality is adaptive technical attack. The United Kingdom's National Cyber Security Centre has assessed that AI will almost certainly increase the volume and impact of cyberattacks over the coming years, chiefly by improving reconnaissance, vulnerability discovery and the evasion of detection tools.⁶ The fourth is the set of attacks against AI systems themselves, including data poisoning, adversarial inputs, prompt injection and model theft, which threaten the integrity of the AI-driven systems on which banks, hospitals and public authorities increasingly rely.

C. WHY THESE ATTACKS CHALLENGE LEGAL CATEGORIES

Four features set AI-powered attacks apart from conventional cybercrime. The first is *scale and speed*: a single operator can run thousands of tailored attacks at once. The second is *accessibility*: attacks that formerly required specialist skill are now available to unsophisticated actors through consumer-grade tools. The third is *opacity*: machine-learning systems often behave in ways that even their developers cannot fully explain, complicating questions of

³ Miles Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* 18–25 (2018), <https://arxiv.org/abs/1802.07228>.

⁴ Europol, *ChatGPT: The Impact of Large Language Models on Law Enforcement* 6–10 (Mar. 2023).

⁵ Daniel Kelley, *WormGPT – The Generative AI Tool Cybercriminals Are Using to Launch Business Email Compromise Attacks*, SlashNext (July 13, 2023).

⁶ Nat'l Cyber Sec. Ctr. (U.K.), *The Near-Term Impact of AI on the Cyber Threat* (Jan. 2024).

foreseeability and causation.⁷ The fourth is *dual-use character*: the same model that powers a legitimate customer-service chatbot can be repurposed to deceive customers, and the line between a lawful tool and an instrument of crime frequently turns on use rather than design.⁸ Legal systems built around identifiable human perpetrators, discrete acts and clear intent find each of these features difficult to accommodate.

III. THE INDIAN THREAT LANDSCAPE

India presents an unusually attractive environment for attackers. Its digital population is vast and heterogeneous, with wide variation in digital literacy. Real-time payments have made fraud proceeds instantly transferable, and multilingual content creates a demand for AI translation and voice tools that attackers can exploit with equal ease. Financial fraud, “digital arrest” scams in which victims are coerced by persons impersonating law-enforcement officers over video calls, and the circulation of manipulated images and videos of public figures and private individuals have become familiar features of the Indian cyber landscape.

Critical sectors are also at risk. Power grids, telecommunications, banking, healthcare and transport increasingly depend on networked and AI-assisted systems, and a successful attack on any of them could produce cascading consequences. Public concern over deepfakes reached a peak in late 2023, when manipulated videos of well-known individuals circulated widely, prompting the Ministry of Electronics and Information Technology (MeitY) to issue an advisory reminding intermediaries of their due-diligence obligations.⁹ The episode revealed a structural truth: India’s response to AI-enabled harm has so far been reactive and instrument-specific, assembled from advisories and the stretching of older provisions, rather than the product of a coherent legislative design.

IV. THE EXISTING LEGAL FRAMEWORK AND ITS LIMITATIONS

A. THE INFORMATION TECHNOLOGY ACT, 2000

The Information Technology Act, 2000 (IT Act) remains the principal cyber statute.¹⁰ Section 43 imposes civil liability for unauthorised access, introduction of computer contaminants, denial of access and related acts, while section 66 criminalises such acts when done dishonestly

⁷ Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 Harv. J.L. & Tech. 353, 363–73 (2016).

⁸ Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 Calif. L. Rev. 513, 538–45 (2015).

⁹ Ministry of Elecs. & Info. Tech., Gov’t of India, *Advisory on Due Diligence by Intermediaries/Platforms Under the Information Technology Rules, 2021* (Dec. 26, 2023).

¹⁰ Information Technology Act, 2000, No. 21 of 2000, India Code (2000) [hereinafter IT Act].

or fraudulently.¹¹ Sections 66C and 66D punish identity theft and cheating by personation using a computer resource, and are the provisions most frequently invoked in deepfake and voice-cloning frauds.¹² Section 66F, which addresses cyber terrorism, could in principle capture an AI-assisted attack on critical information infrastructure.¹³ The Act also has extraterritorial reach where the offending conduct involves a computer, computer system or network located in India.¹⁴ Sections 84B and 84C extend liability to abetment and attempt.¹⁵

These provisions are useful, but they were drafted in 2000 and amended in 2008 with a human intruder in mind. Four weaknesses stand out. First, the Act contains no definition of, or reference to, artificial intelligence, automated systems or synthetic media, so prosecutors must fit AI-enabled conduct into categories designed for different problems. Second, the penalties are modest by comparison with the harm that automated attacks can inflict. Imprisonment under section 66D, for example, extends to three years, which does not differentiate between a single act of deception and an industrialised campaign targeting thousands. Third, the Act is silent on the creation, distribution or possession of tools designed to commit cyber offences, leaving the developers and sellers of malicious AI tools to be reached, if at all, through the general law of abetment and conspiracy. Fourth, the intermediary safe harbour in section 79, as read down in *Shreya Singhal v. Union of India*, depends on actual knowledge acquired through a court order or government notification, which is poorly suited to harms such as deepfakes that spread within hours.¹⁶

B. THE BHARATIYA NYAYA SANHITA, 2023

The Bharatiya Nyaya Sanhita, 2023 (BNS) supplements the IT Act through its general offences. Cheating, cheating by personation, forgery and the use of forged electronic records as genuine can all apply to AI-enabled fraud.¹⁷¹⁸ The BNS also introduces the offence of organised crime, which expressly includes cyber-crimes committed by a group acting in concert, and a corresponding offence of petty organised crime.¹⁹ This is a welcome development, because AI-enabled fraud is often carried out by structured syndicates operating across state and national borders.

¹¹ IT Act §§ 43, 46, 66.

¹² *Id.* §§ 66C, 66D.

¹³ *Id.* § 66F.

¹⁴ IT Act §§ 1(2), 75.

¹⁵ *Id.* §§ 84B, 84C.

¹⁶ IT Act § 79; *Shreya Singhal v. Union of India*, (2015) 5 SCC 1 (India).

¹⁷ Bharatiya Nyaya Sanhita, 2023, No. 45 of 2023, India Code (2023) [hereinafter BNS].

¹⁸ BNS §§ 318, 319, 336, 340.

¹⁹ *Id.* §§ 111, 112.

The difficulty lies in the doctrinal structure of criminal liability. Indian criminal law proceeds from the presumption that guilt requires a guilty mind.²⁰ That presumption works smoothly where a human attacker deploys a tool with fraudulent intent. It becomes strained where harm arises from the conduct of an autonomous system that the human operator did not specifically direct, or from a general-purpose model that a developer released without any criminal purpose but with insufficient safeguards. Scholars have proposed several models of criminal responsibility for AI-mediated harm, including liability for natural-probable-consequence and direct-liability models, but none has been adopted in Indian law.²¹ The result is an accountability gap: the human who causes the harm may be unidentifiable, and the human who made the harm possible may fall outside the definition of any existing offence.

C. THE DIGITAL PERSONAL DATA PROTECTION ACT, 2023

AI-powered attacks frequently target personal data, both as a prize and as raw material for further attacks. The Digital Personal Data Protection Act, 2023 (DPDP Act) requires a data fiduciary to protect personal data in its possession by taking reasonable security safeguards to prevent a personal data breach, to notify the Data Protection Board and each affected data principal when a breach occurs, and exposes a fiduciary that fails in the first duty to a penalty of up to ₹250 crore.²² Once fully in force, the Act will also replace section 43A of the IT Act, the existing provision on compensation for failure to protect sensitive personal data.²⁴

The DPDP Act is a significant step, but it is a data-protection statute and not a cyber-security or AI statute. Its security obligation is framed in general terms (“reasonable security safeguards”) and does not stipulate what reasonableness requires in an environment where adversaries use AI. The Act regulates the holder of personal data, not the developer of the tool used to attack it. It also empowers the Central Government to exempt State instrumentalities from its provisions in the interests of security and public order, which limits its protective reach in relation to public-sector systems.²⁵

D. EXECUTIVE DIRECTIONS, POLICY AND INSTITUTIONS

The IT Act establishes CERT-In as the national agency for incident response and empowers it to issue directions.²⁶ Under that power, CERT-In’s directions of April 2022 require service

²⁰ Nathulal v. State of Madhya Pradesh, AIR 1966 SC 43 (India).

²¹ Gabriel Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law* 25–36 (2013).

²² Digital Personal Data Protection Act, 2023, No. 22 of 2023, India Code (2023) [hereinafter DPDP Act].

²³ *Id.* § 8(5)–(6) & sched. (item 1).

²⁴ *Id.* § 44(2).

²⁵ *Id.* § 17(2).

²⁶ IT Act §§ 70, 70A, 70B.

providers, intermediaries, data centres and bodies corporate to report specified cyber incidents within six hours of noticing them, and to retain logs for 180 days.²⁷ The National Critical Information Infrastructure Protection Centre protects systems designated as critical, and the National Cyber Security Policy of 2013 articulated a broad strategy, though it pre-dates the generative AI era by a decade.²⁸

On AI specifically, India has so far preferred soft-law instruments. The advisory on deepfakes noted above, proposed amendments to the Intermediary Guidelines and Digital Media Ethics Code Rules, 2021 on the labelling of synthetically generated content, the IndiaAI Mission, and the India AI Governance Guidelines of November 2025 collectively reflect a principles-based, innovation-friendly approach that relies on existing law.²⁹³⁰ There is considerable merit in regulatory restraint in a fast-moving field. However, advisories lack the force of law, are inconsistently complied with, and cannot create offences, impose penalties or allocate liability. Moreover, the CERT-In directions are general and do not distinguish AI-related incidents or require disclosure of AI involvement in an attack.

E. EVIDENTIARY AND PROCEDURAL GAPS

Even where a substantive offence exists, proving it in the AI context is demanding. The Supreme Court has held that electronic records are admissible only on production of the certificate required by the evidence statute, a requirement now carried into section 63 of the Bharatiya Sakshya Adhiniyam, 2023, which additionally calls for a hash value and an expert's certificate.³¹³² Synthetic media creates a distinct problem: as manipulated content becomes harder to distinguish from authentic content, courts will increasingly face the "liar's dividend", in which genuine evidence is disputed as fabricated, alongside the risk that fabricated evidence is accepted as genuine. The Bharatiya Nagarik Suraksha Sanhita, 2023 improves matters by recognising electronic communication in procedure and by requiring forensic investigation in serious offences, but it does not prescribe standards for detecting AI-generated content or for

²⁷ Indian Comput. Emergency Response Team, Directions Under Sub-Section (6) of Section 70B of the Information Technology Act, 2000 Relating to Information Security Practices, Procedure, Prevention, Response and Reporting of Cyber Incidents for Safe & Trusted Internet, No. 20(3)/2022-CERT-In (Apr. 28, 2022).

²⁸ Dep't of Elec. & Info. Tech., Gov't of India, *National Cyber Security Policy 2013* (2013).

²⁹ Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (India).

³⁰ Ministry of Elecs. & Info. Tech., Gov't of India, *India AI Governance Guidelines* (Nov. 2025); see also Press Info. Bureau, *Cabinet Approves Comprehensive National-Level IndiaAI Mission* (Mar. 7, 2024).

³¹ Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473 (India); Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1 (India).

³² Bharatiya Sakshya Adhiniyam, 2023, No. 47 of 2023, India Code (2023), § 63 & sched.

attributing an attack to a particular system or actor.³³ Investigative agencies and trial courts also face significant capacity constraints in digital forensics.

F. SUMMARY OF GAPS

The analysis yields five principal gaps. There is a *definitional gap*, since AI is absent from the statutory vocabulary. There is a *responsibility gap*, since developers, deployers and platforms bear no specific duties of care. There is a *proportionality gap*, since penalties do not reflect the aggravated harm of automated, large-scale attacks. There is an *evidentiary gap*, since attribution and authentication standards for AI-generated material are undeveloped. And there is an *institutional gap*, since no body combines technical expertise, regulatory authority and coordination functions in relation to AI-enabled threats. These gaps cannot be closed by advisory or by judicial creativity alone.

V. COMPARATIVE AND INTERNATIONAL PERSPECTIVES

A. THE EUROPEAN UNION

The European Union has adopted the most comprehensive approach. Its AI Act classifies AI systems by risk, requires high-risk systems to achieve an appropriate level of accuracy, robustness and cybersecurity, obliges providers of general-purpose models with systemic risk to carry out adversarial testing and ensure adequate cybersecurity protection, and imposes transparency duties on deepfakes.³⁴ This operates alongside the NIS2 Directive, which sets baseline cyber-security risk-management and incident-reporting duties for essential and important entities, and the Cyber Resilience Act, which requires security by design in digital products.³⁵ The European model demonstrates that obligations can be tied to the *risk profile* of the system and not to the technology as such. Its drawbacks, namely heavy compliance burdens and the danger of regulatory overreach, are real, and a developing economy would need to adapt it rather than transplant it.

B. THE UNITED STATES

The United States has followed a more fragmented, sector-based and voluntary path. The National Institute of Standards and Technology's AI Risk Management Framework offers non-

³³ Bharatiya Nagarik Suraksha Sanhita, 2023, No. 46 of 2023, India Code (2023), §§ 94, 105, 176(3).

³⁴ Regulation 2024/1689, of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence, 2024 O.J. (L 2024/1689) 1, arts. 15, 50, 55 [hereinafter EU AI Act].

³⁵ Directive 2022/2555, of the European Parliament and of the Council of 14 December 2022 on Measures for a High Common Level of Cybersecurity Across the Union, 2022 O.J. (L 333) 80; Regulation 2024/2847, of the European Parliament and of the Council of 23 October 2024 on Horizontal Cybersecurity Requirements for Products with Digital Elements, 2024 O.J. (L 2024/2847) 1.

binding guidance, and the federal executive order on safe AI development issued in 2023 was revoked in January 2025.³⁶ Criminal law continues to rest on the Computer Fraud and Abuse Act, while Congress has legislated narrowly against particular harms, such as the non-consensual publication of intimate imagery, including AI-generated imagery, through the TAKE IT DOWN Act of 2025.³⁷ The American experience illustrates both the flexibility of a sectoral approach and its tendency to leave gaps.

C. THE UNITED KINGDOM AND CHINA

The United Kingdom's Computer Misuse Act 1990 is notable for criminalising not only unauthorised access and impairment of systems but also the making, supplying or obtaining of articles for use in such offences.³⁸ This addresses precisely the gap that India's IT Act leaves open in relation to malicious AI tools. China, by contrast, has adopted binding, targeted rules on deep synthesis services and generative AI, requiring conspicuous labelling of synthetic content, identity verification of users and security assessments for certain services.³⁹ The Chinese model offers a lesson in specificity and enforceability, although its breadth of State control would be difficult to reconcile with the constitutional protections that apply in India.

D. INTERNATIONAL INSTRUMENTS

At the international level, the Budapest Convention on Cybercrime of 2001 requires states to criminalise illegal access, interference with data and systems, computer-related fraud and the misuse of devices, and to facilitate cross-border cooperation.⁴⁰ India has so far not acceded to it. The United Nations General Assembly adopted a new Convention against Cybercrime in December 2024, which establishes a broader framework for criminalisation, evidence-sharing and international assistance.⁴¹ Because AI-powered attacks are inherently transnational, with models trained in one country, hosted in another and used against victims in a third, India's domestic framework must be designed to interoperate with these instruments, and its position on participation merits fresh consideration.

³⁶ Nat'l Inst. of Standards & Tech., *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (Jan. 2023); Exec. Order No. 14110, 88 Fed. Reg. 75191 (Oct. 30, 2023), *revoked by* Exec. Order No. 14148, 90 Fed. Reg. 8237 (Jan. 20, 2025).

³⁷ Computer Fraud and Abuse Act, 18 U.S.C. § 1030; TAKE IT DOWN Act, Pub. L. No. 119-12, 139 Stat. 55 (2025).

³⁸ Computer Misuse Act 1990, c. 18, §§ 1–3A (U.K.).

³⁹ Provisions on the Administration of Deep Synthesis of Internet-Based Information Services (promulgated by the Cyberspace Admin. of China, Nov. 25, 2022, effective Jan. 10, 2023); Interim Measures for the Management of Generative Artificial Intelligence Services (promulgated by the Cyberspace Admin. of China et al., July 10, 2023, effective Aug. 15, 2023) (China).

⁴⁰ Convention on Cybercrime arts. 2–6, Nov. 23, 2001, E.T.S. No. 185.

⁴¹ G.A. Res. 79/243, United Nations Convention Against Cybercrime (Dec. 24, 2024).

VI. TOWARDS A NEW LEGAL FRAMEWORK FOR INDIA

A. THE CASE FOR REFORM AND ITS VEHICLE

It may be objected that technology-neutral laws are preferable because they do not date quickly, and that AI-powered attacks are merely old crimes committed by new means. There is force in this view, and any new framework should preserve technology neutrality in its core offences. Yet the analysis in Part IV shows that the problem is not only one of novel means but of altered scale, altered attribution and altered responsibility. These call for targeted additions. The Digital India Act, which MeitY has discussed in public consultations as a successor to the IT Act, had not been enacted at the time of writing and could serve as the legislative vehicle; alternatively, a dedicated statute or a package of amendments to the IT Act and the BNS could be adopted sooner.⁴² What matters is the substance of the reform, to which this paper now turns.

B. CORE PROPOSALS

First, statutory definitions. The law should define “artificial intelligence system”, “synthetic media” and “AI-enabled cyber offence” in technology-neutral and functional terms, so that the definitions survive technical change. Functional definitions, focused on what a system does and not how it is built, are more durable than technical ones.

Second, aggravated offences. Rather than creating a separate offence for every AI-related harm, the statute should treat the use of AI to automate, scale or conceal an existing cyber offence as an aggravating circumstance attracting enhanced punishment. This respects proportionality, since the additional penalty tracks the additional harm, and avoids duplicating existing offences. Sections 66, 66C and 66D of the IT Act and the corresponding BNS provisions are natural candidates for such enhancement.

Third, an offence of malicious tool-making. Drawing on article 6 of the Budapest Convention and section 3A of the UK Computer Misuse Act, India should criminalise the development, supply or possession of AI tools whose primary purpose is to commit cyber offences, where the person acts with intent or knowledge that the tool will be so used. A carefully drafted exemption for good-faith security research, red-teaming and academic study is essential, so that the offence does not chill legitimate work. The requirement of intent or knowledge ensures that developers of lawful general-purpose models are not punished for downstream misuse they neither intended nor knew of.

⁴² Ministry of Elecs. & Info. Tech., Gov’t of India, *Proposed Digital India Act, 2023: Presentation on Public Consultation* (2023).

Fourth, graduated duties of care for developers and deployers. For high-risk and highly capable AI systems, the law should impose civil and regulatory duties rather than criminal liability: pre-deployment security testing, monitoring for abuse, maintenance of audit logs, reasonable safeguards against misuse, and prompt reporting of serious incidents. Breach of these duties should attract administrative penalties and civil liability in negligence, and not criminal sanctions, save where recklessness or knowledge is established. This model fills the responsibility gap without importing the logical difficulties of attributing a criminal mind to a developer who never intended harm. Obligations should be proportionate to risk and capability, with lighter requirements for small enterprises and startups, so that India's innovation ecosystem is not disadvantaged.

Fifth, synthetic media provenance and rapid response. Platforms and AI service providers should be required to implement technically feasible labelling and provenance measures for AI-generated content, and to act on credible complaints of harmful impersonation within a short, defined period. Because the existing knowledge-based safe harbour is too slow for viral harms, a dedicated notice-and-takedown procedure for non-consensual intimate imagery and fraudulent impersonation, with procedural safeguards against abuse, is warranted. The statute should also give statutory recognition to personality rights, which Indian courts have so far protected case by case, as in the Delhi High Court's orders restraining the misuse of an actor's name, likeness and voice.⁴³

Sixth, incident reporting and information sharing. CERT-In's reporting directions should be supplemented by an AI-specific schedule requiring disclosure, within a defined timeframe, of incidents in which AI played a material role, including attacks on AI systems themselves. Good-faith disclosure and cooperation should be rewarded with mitigation of penalties, so that the regime encourages transparency instead of concealment.

Seventh, evidence and procedure. The Bharatiya Sakshya Adhiniyam and the Bharatiya Nagarik Suraksha Sanhita should be supported by rules or guidelines prescribing minimum forensic standards for the detection and authentication of AI-generated material, including chain-of-custody requirements, accredited forensic laboratories and the use of provenance metadata. Judges, prosecutors and police officers need systematic training, and consideration should be given to specialised benches or designated courts for complex cyber matters. Without these, even the best-drafted offences will remain difficult to prove.

⁴³ Anil Kapoor v. Simply Life India, 2023 SCC OnLine Del 6914 (India).

Eighth, institutional design. A fragmented institutional landscape is ill-suited to a fast-evolving threat. The proposed framework should confer statutory status and technical resources on a coordinating authority, which may be built upon CERT-In, with a mandate to set AI-security standards, supervise compliance by high-risk system providers, maintain a national repository of AI-related incidents and coordinate with the National Critical Information Infrastructure Protection Centre and the Indian Cyber Crime Coordination Centre. The authority should be subject to judicial review and parliamentary oversight, and its decisions should be accompanied by reasons.

Ninth, international cooperation. India should strengthen mutual legal assistance arrangements, pursue bilateral and multilateral cooperation for cross-border evidence and take-down requests, and reconsider its engagement with the Budapest Convention and the new United Nations instrument, in a manner consistent with its sovereignty concerns. Without such cooperation, domestic legislation will reach only the victims and not the perpetrators.

C. CONSTITUTIONAL AND POLICY SAFEGUARDS

Any new framework must be consistent with the Constitution. The right to privacy, recognised in *Justice K.S. Puttaswamy (Retd.) v. Union of India*, requires that any interference be backed by law, pursue a legitimate aim and satisfy the test of proportionality.⁴⁴ The Supreme Court has applied a structured proportionality analysis in cases such as *Modern Dental College*, and has insisted in *Anuradha Bhasin v. Union of India* that restrictions affecting speech and access to the internet be necessary, least restrictive and subject to review.^{45,46} The invalidation of section 66A in *Shreya Singhal* is a salutary reminder that vague, overbroad cyber offences will not survive scrutiny under Articles 14, 19(1)(a) and 21.⁴⁷

Three design principles follow. Offences must be precise, with clear mental elements, so as to satisfy the requirement of certainty in penal law under Article 20(1) and Article 21. Powers of monitoring, takedown and data access must be accompanied by independent oversight and remedies. And the framework should include a mechanism for periodic review, so that provisions can be recalibrated as technology evolves, perhaps through a statutory review within a fixed number of years of commencement. Regulatory sandboxes and consultation with

⁴⁴ *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1 (India).

⁴⁵ *Modern Dental College & Rsch. Ctr. v. State of Madhya Pradesh*, (2016) 7 SCC 353 (India).

⁴⁶ *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637 (India).

⁴⁷ *Shreya Singhal*, (2015) 5 SCC 1.

industry, academia and civil society would further help to keep the law grounded in technical reality.⁴⁸

VII. CONCLUSION

AI-powered cyberattacks represent a qualitative shift in the threat environment, and not merely a quantitative increase in cybercrime. They enable deception at scale, lower the skill required to attack, blur the lines of attribution and expose weaknesses in the AI systems on which society increasingly depends. India's existing legal framework, comprising the IT Act, the BNS and the DPDP Act, is not irrelevant to these threats; it supplies a foundation on which a response can be built. But it is neither designed for, nor adequate to, the distinctive problems that AI introduces.

This paper has argued that India should move from reactive, advisory-driven responses to a deliberate statutory framework. Such a framework would define the relevant concepts, treat AI-enabled automation as an aggravating factor, criminalise the knowing creation and supply of malicious tools, impose proportionate duties of care on those who build and deploy high-risk systems, strengthen evidentiary and institutional capacity, and engage with international cooperation. It would do so within the limits set by proportionality, legality and the protection of fundamental rights. The aim is not to regulate AI out of existence but to ensure that its benefits are not purchased at the cost of the security and trust on which India's digital future depends. The window for thoughtful, evidence-based legislation is open, but it will not remain so for long.

⁴⁸ Ryan Abbott, *The Reasonable Robot: Artificial Intelligence and the Law* 1–10 (2020).