# ARTIFICIAL INTELLIGENCE AS A NEW WEAPON OF CYBERCRIME: LEGAL SECURITY IN REGULATING RAPID AI DEPLOYMENT

Brish Kumar Pankaj, B.A. LL.B. (Hons.), Law College Dehradun, Uttaranchal University, Dehradun, Uttarakhand, India.

Dr. Aishwarya Singh, Assistant Professor, Law College Dehradun, Uttaranchal University, Dehradun, Uttarakhand, India.

## ABSTRACT

Artificial intelligence (AI) is being used in India as a powerful tool in cybercrime. AI acts as a force multiplier for cybercrime in India, and this increasingly challenges legal security when deployment cycles are faster than legal and forensic response. This study develops a legal- security framework for rapid AI deployment by combining doctrinal analysis of Indian cyber and criminal statutes, delegated intermediary obligations, data protection rules, and leading constitutional and evidentiary case law, with a structured mapping of the AI-enabled attack chain (data poisoning, model theft, prompt injection, and downstream laundering). The study finds that AI is mainly used to speed up the commission of traditional crimes - impersonation, cheating, extortion, invasion of privacy, delivery of malware, and payment fraud - by lowering the skills needed and scaling the multilingual persuasion, but at the same time, it increases the risks of plausible deniability and synthetic evidence. Indias regulatory architecture remains anchored in the Information Technology Act, 2000 and conditional intermediary safe harbor, but recent rulemaking has leaned toward time-critical governance: The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 define synthetically generated information, require prominent labelling and persistent metadata, and reduce actual knowledge takedown compliance to three hours. The Digital Personal Data Protection framework and the Digital Personal Data Protection Rules, 2025 add parallel duties around purpose limitation and breach intimation, while CERT-In's directions on cyber incident reporting set a six-hour operational timeline for coordinated response. Key gaps persist in attribution, admissibility, and remedies. Investigations often lack preserved prompt histories, API access logs, and platform-side provenance needed to prove intent, and courts face rising authenticity disputes as deepfakes and fabricated logs blur the "original" electronic record. The paper concludes that legal security in the AI era requires supply-chain duties that follow control over models and tool-integrations, evidence-preserving containment procedures that are auditable and judicially reviewable, and victim-centered remedies -freezing, takedown, and injunctions - that operate within hours rather than weeks.

**Keywords:** AI-enabled cybercrime; synthetically generated information (SGI); deepfakes; intermediary due diligence; electronic evidence; DPDP Rules 2025; CERT-In incident reporting; rapid injunctive relief.

## 1.1 INTRODUCTION

Artificial intelligence changed the economics of cybercrime by making it easier and cheaper for even low-skilled criminals to launch large-scale attacks, such as scams, hacks, and money laundering, without the need for much time or expensive equipment. The issue with the law is not only about the technical capability; it is more a question of governance - who has the authority, who is accountable, and what standards can be enforced both in private platforms and public enforcement. Artificial intelligence shortens the time from planning to the damage, at the same time, it increases the possibilities of denying the crime through different layers of automation. In this situation, legal security relates to the ability of the law to impose predictable obligations, distribute the risks, and safeguard the rights even under the conditions of quick deployment.[1]

Using AI in cybercrime reveals a deep imbalance in the relationship between private deployment incentives and public rule-of-law commitments. The first two, developers and deployers, are only interested in maximizing speed and user growth without even considering the State that also has to justify its coercive measures through legality, necessity, proportionality, and procedural fairness. Law enforcement agencies have offenders who can simply shift their costs by using anonymized infrastructure and cross-border platforms. Victims endure high evidentiary burdens and the double whammy of their reputation being quickly tarnished. There is a governance gap after all that which only under-enforcement can be blamed for; but rather fragmented decision-making, where the takedown, blocking, investigation, and compensation of offences are carried out without having consistent thresholds, documentation standards, or review mechanisms.[2]

In India, the challenge of governance is made more difficult by constitutional safeguards that limit the regulation of speech, privacy, and due process. This is combined with a statutory dependence on intermediary compliance and digital evidence. Legal certainty should balance prevention and legality: content restrictions have to be carefully limited to avoid ambiguity; surveillance and monitoring need to be limited to prevent unreviewable discretion; prosecution should be based on admissible electronic evidence and not on mere suspicion. The system must also deal with the differential harms, where out-of-touch users are targeted through local

[1] Leslie F. Sikos, *AI-Enabled Cybersecurity* 121 (Springer Nature Switzerland, Cham, 1st edn., 2021).
[2] Leslie F. Sikos, *AI in Cybersecurity* 87 (Springer, Cham, 1st edn., 2019).

language persuasion and where synthetic media can cause disruption of public order without clear identification of the source.

A legal security framework that facilitates quick AI deployment depends on three interconnected capabilities. One, responsible parties should be clearly pointed out throughout the AI supply chain, so that enforcement does not end up being merely symbolic if only end users are 'arrested'. Two, there must be procedural safeguards that help preserve the integrity of the evidence, as synthetic product can be created to either mislead the investigations or to set up the victims. Three, the remedies have to be functioned within hours instead of weeks, because the viral spread itself is a violation of the right to be protected. This way of thinking is fit for regulations that are capable of being not only reactive, but also administratively auditable and judicially reviewable.

## 1.2 AI-ENABLED CYBERCRIME LANDSCAPE

One can best think of AI-enabled cybercrime as an industry where models, data sources, and platform features are essentially crime accelerators for the usual criminal objectives. It is not a switch from crime to technology, but rather a shift from manual criminal labor to automated criminal throughput. Synthetic impersonation, multilingual persuasion, and rapid financial extraction through online payment systems are how Indian victims face these attacks.[3] The law enforcement side of it sees compressed timelines, increased false leads, and convincingly generated evidence on a massive scale. The regulatory puzzle lies in continuing to consider AI as a means of accelerating the commission of current crimes while simultaneously dealing with the new vectors that arise from reliance on models.[4]

### 1.2.1 Meaning and Scope

Functional clarity on AI systems is important as it affects liability and compliance obligations, which are based on foreseeable misuse, controllability, and auditability. Generative AI is a term used for models that generate new text, audio, images, or code similar to the ones they were trained on, thus capable of highly deceptive and automated content production.[5] Agentic automation means systems that design and carry out multi-step tasks by making use of the tools, thus making a prompt-to-transaction, scraping and exfiltration modus operandi.

[3] Madhumitha P S, "A Study on the Role of Artificial Intelligence in Preventing Cyber Crimes - Indian and International Perspective", 4 *Indian Journal of Integrated Research in Law* 836 (2024).

[4] Marc Schmitt, "Digital Deception: Generative AI in Social Engineering Attacks", 57 *Artificial Intelligence Review* 324 (2024).

[5] ChatGPT Report, *available at:* https://www.europol.europa.eu/publications-events/publications/chatgpt-report (last visited on February 10, 2026).

Recommender systems order and boost content, thus they can make scams or deepfakes appear normal through engagement optimization. ML-based anomaly evasion is about adversarial moves using the detection models as a loophole, thus making fraud flags and intrusion alarms less dependable.[6]

### 1.2.2 AI Attack Chain

Mapping an attack chain has important legal implications for when different stages in the chain are under the control of different actors, and negligence or complicity can be linked to different points of control. The weaponization phase starts with data acquisition and training a model, where using a poisoned or illegal dataset can result in embedding backdoors or leaking sensitive information.[7] It goes on to model access, where APIs and user interfaces become the door to misuse and where logging policies have an impact on attribution. Deployment environments then decide if safeguards can be practically implemented, including limits on rates and verification of identities. Misuse downstream consists of laundering through platforms, fast reposting, and cross-border monetization, which make jurisdiction and preservation more difficult.[8]

#### *1.2.2.1 Data Poisoning and Model Manipulation*

Data poisoning is changing model behavior in a way that it first contaminates the data of training, then it obscures the feedback signals, and/or it corrupts the evaluation benchmarks. In these scenarios, the outputs will be biased towards the attackers' needs, or the attackers' presence will be kept secret. An attacker who wants to put malicious patterns into the open datasets or wants to manipulate the reinforcement feedback by means of coordinated reporting can cause the models to normalize scam scripts or to misclassify harmful payloads as benign.[9] In cases where security tools are heavily dependent on the use of ML classifiers, poisoning can effectively decrease the sensitivity of detection to certain payload families thereby creating selective blind spots. From the viewpoint of legal security, the main problem with poisoning is that it is generally invisible until the damage is done, therefore, compliance frameworks should

[6] Cybercrime, *available at:* https://www.interpol.int/en/Crimes/Cybercrime (last visited on February 11, 2026).
[7] Yevgeniy Vorobeychik, Murat Kantarcioglu, *Adversarial Machine Learning* 103 (Morgan & Claypool Publishers, San Rafael, 1st edn., 2018).
[8] Clarence Chio, David Freeman, *Machine Learning and Security: Protecting Systems with Data and Algorithms* 142 (O'Reilly Media, Sebastopol, 1st edn., 2018).
[9] Yogi Reddy Maramreddy, Kireet Muppavaram, "Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach", 14 *Engineering, Technology & Applied Science Research* 15505 (2024).

require documentation of dataset provenance, monitoring of drift, and incident-reporting, all being triggered by behavioral anomalies.[10]

#### *1.2.2.2 Model Extraction and Theft*

Model extraction is basically a way for a criminal gang to copy what a model can do without having to pay for training, thus turning a proprietary model into criminal infrastructure that can be reused. The theft can be done by systematically asking questions that approximate decision boundaries, by leaking weights, or by insider control that defeats contracts.[11] The thief can then tailor the stolen model for creating fraud methods, hiding malware, or voice cloning, while the original developer ends up with a damaged reputation and regulatory investigations. Among the legal security issues are the definition of what reasonable security measures for model assets are, the specification of the access control responsibilities, and the keeping of forensic logs that help in attribution without 'over-collecting' personal data beyond the limits of the lawful purpose.[12]

#### *1.2.2.3 Prompt Injection and Jailbreaks*

Prompt injection exploits one weakness of AI systems that is, they obediently follow instructions which are embedded in content, tool outputs or retrieved documents. Hence, attackers can override the safety constraints. If tool-using assistants are subjected to the injected prompts, such prompts can be indicative of secret disclosure, performance of unauthorized deeds, or transaction flow manipulations.[13] Jailbreaks can be considered as bypass attempts of policy filters by using adversarial phrasing, encoding tricks, or multi-turn manipulation to get instructions that may be harmful or targeted impersonation content. The legal security issue is that the ordinary users may cause the harmful actions without knowing the cause, thus complicating mens rea and foreseeability. Risk governance is about having evident safeguards, audit trails, and clearly scoped tool permissions that are in line with the least-privilege principles to be able to manage the risk.[14]

---

[10] Zhiyi Tian, Li Chen, et.al., "A Comprehensive Survey on Poisoning Attacks and Countermeasures in Machine Learning", 55 *ACM Computing Surveys* 166 (2022).

[11] LLM10: Model Theft, *available at:* https://genai.owasp.org/llmrisk2023-24/llm10-model-theft/ (last visited on February 8, 2026).

[12] OWASP Top 10 for Large Language Model Applications, *available at:* https://owasp.org/www-project-top-10-for-large-language-model-applications/ (last visited on February 9, 2026).

[13] James Phoenix, Mike Taylor, *Prompt Engineering for Generative AI* 74 (O'Reilly Media, Sebastopol, 1st edn., 2024).

[14] Steve Wilson, *The Developer's Playbook for Large Language Model Security* 156 (O'Reilly Media, Sebastopol, 1st edn., 2024).

### 1.2.3 Core AI Cybercrime Typologies in India

Indian cybercrimes mirror a mixture of traditional offenses and AI-based scaling of the crimes. Most of the reported crimes have been around identity deception, account hacking, payment fraud, and blackmail for reputation, the majority of which were at the consumers and through social media, messaging platforms, and app ecosystems.[15] Parties to the lawsuits usually end up with the issue of intermediary takedown duties, privacy breaches, and granting of injunctive relief in cases where synthetic media is used to harm personality and livelihood. The operational enforcement is concentrating on mule accounts, device compromise, and call-center style coordination, while the AI component is still positioned upstream as an enabler of persuasion and automation. The below-mentioned typologies are significant as each requires different evidential and remedial measures.[16]

#### *1.2.3.1 Deepfake Impersonation and Identity Fraud*

Deepfakes significantly increase the danger of impersonation as they are able to produce lifelike audio-video that can easily bypass regular verification methods. Fake voice calls and video messages can trick people into making UPI transfers, changing passwords, or giving away one-time passwords through the use of panic and authority tactics. Social harm takes place when counterfeit videos are leaked in order to blackmail or harass, for instance, through extortion and workplace sabotage.[17] Political misinformation becomes a public order threat if synthetic materials are released at election time or in situations of communal tension. Legal safeguards necessitate the rapid securing of originals, the implementation at the platform level of means to trace the first upload, and solutions that not only involve takedown but also limitation on reposting without infringing the constitutional rights to free speech and without the prohibited categories being vague.[18]

#### *1.2.3.2 Automated Phishing and Social Engineering at Scale*

AI has helped to reduce the linguistic friction that was one of the limiting factors of phishing. Now it is possible to send personalized phishing messages in a variety of Indian languages and

[15] Aahana Chopra, Ananya Shukla, "Exploring the Misuse of Deepfake Technology in India: Implications for Society", 16 *Global Media Journal-Indian Edition* 1 (2024).
[16] Narendra Kumar Dhaka, Navaneet Kumar, "Cybercrime in India: Law, Policy, & Enforcement in AI Era", 15 *Indian Streams Research Journal* 1 (2025).
[17] Advisory No. 2(4)/2023-CyberLaws, *available at:* https://www.meity.gov.in/static/uploads/2024/02/c9f89809b63d22656be38a166ef14949.pdf (last visited on February 6, 2026).
[18] MeitY Issues Advisory to All Intermediaries to Comply With IT Rules Targeting Deepfakes, *available at:* https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=1990542 (last visited on February 7, 2026).

dialects. Phishing scammers use data from breaches and public profiles that they mix with generative text to create quite believable stories that can include what seems to be a reminder of an invoice, a parcel delivery, and customer care impersonation.[19] The scale effect is also significant: campaigns can test thousands of variants simultaneously and thus quickly find those that will be capable of passing spam and fraud filters. On the legality side, the security issues involve the difficulties in proving inducement and deception when the messages are auto-generated, the need to preserve message headers and metadata, and the obligation to facilitate timely platform cooperation for routing records and account identifiers within legally valid processes.[20]

#### *1.2.3.3 Malware Assistance and Vulnerability Exploitation*

AI helps the cybercriminals by providing them with an easier and faster way of discovering exploits, writing codes, and developing new tactics that can evade security measures, even when the models are limited to obtaining explicit instructions for malicious activities. A cybercriminal can utilize AI to modify the payloads, create different polymorphic variants, and develop social engineering tricks that can deliver malware through the use of documents or links. Machine learning-based anomaly evasion goes against endpoint detection systems by generating traffic patterns that behave and look like normal user behavior.[21] Legal security is a matter of matching substantive offences with the capacity of forensic: investigators need to be able to capture the most ephemeral volatile artefacts, gather and verify logs from different devices, and maintain integrity in the chain-of-custody. There is also the issue of liability which may come up for those deployers who provide access to the tool without having any adequate guardrails or monitoring.[22]

#### *1.2.3.4 Financial Fraud in Digital Payments Ecosystems*

Digital payments have significantly increased the exposure to cyber-attacks as they offer instant settlement, proxy identifiers, and layered intermediaries which result in the distribution of responsibility among banks, payment apps, and platform channels. Artificial Intelligence is being used to facilitate mule recruitment by carrying out mass outreach, preliminary screening,

[19] Markus Jakobsson, Steven Myers, *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft* 166 (John Wiley & Sons, Hoboken, 1st edn., 2007).
[20] Chris Hadnagy, *Social Engineering: The Science of Human Hacking* 198 (John Wiley & Sons, Hoboken, 2nd edn., 2018).
[21] Thanassis Avgerinos, Sang Kil Cha, et.al., "Automatic Exploit Generation", 57 *Communications of the ACM* 74 (2014).
[22] Trung Minh Doan, Nghi Khoa, et.al., "AAGAN: Android Malware Generation System Based on the Adversarial Autoencoder and GAN", 11 *Vietnam Journal of Computer Science* 275 (2024).

and even coaching through automated scripts. At the same time, invoice fraud employs synthetic documents and voice confirmation to change payment directions quickly. Synthetic KYC crime exploits fake documents, doctored photos, and even deepfake liveness attacks to open accounts that are then used for laundering and fast cash-out. The security of the law demands frozen assets fast, the bank and the platform must work together efficiently and evidentiary standards must link device, identity, and transaction intent without reliance on informal screenshots or unverifiable message forwards. Since victims are often caught up with irrevocable transfers and broken resorting mechanisms across several service providers, the solutions need to be focused on down streams and should consider restitution to victims.[23]

Fig. 2 Fig. 2 illustrates the yearly number of complaints and the amounts advised through the National Cyber Crime Reporting Portal and the Citizen Financial Cyber Fraud Reporting and Management System, which together basically reflect the actual scale of the financial cyber fraud reports. The statistics advocate the requirement of time-limited freezing and preservation measures that operate within very short time intervals because delay increases the damage and deterioration of the track record.[24]

[23] Unified Payments Interface Safety Shield, *available at:* https://www.npci.org.in/safety-feature (last visited on February 5, 2026).

[24] SMS on Customer Liability in Unauthorised Electronic Banking Transactions, *available at:* https://www.rbi.org.in/commonman/english/Scripts/SMSLimitedliability.aspx (last visited on February 4, 2026).

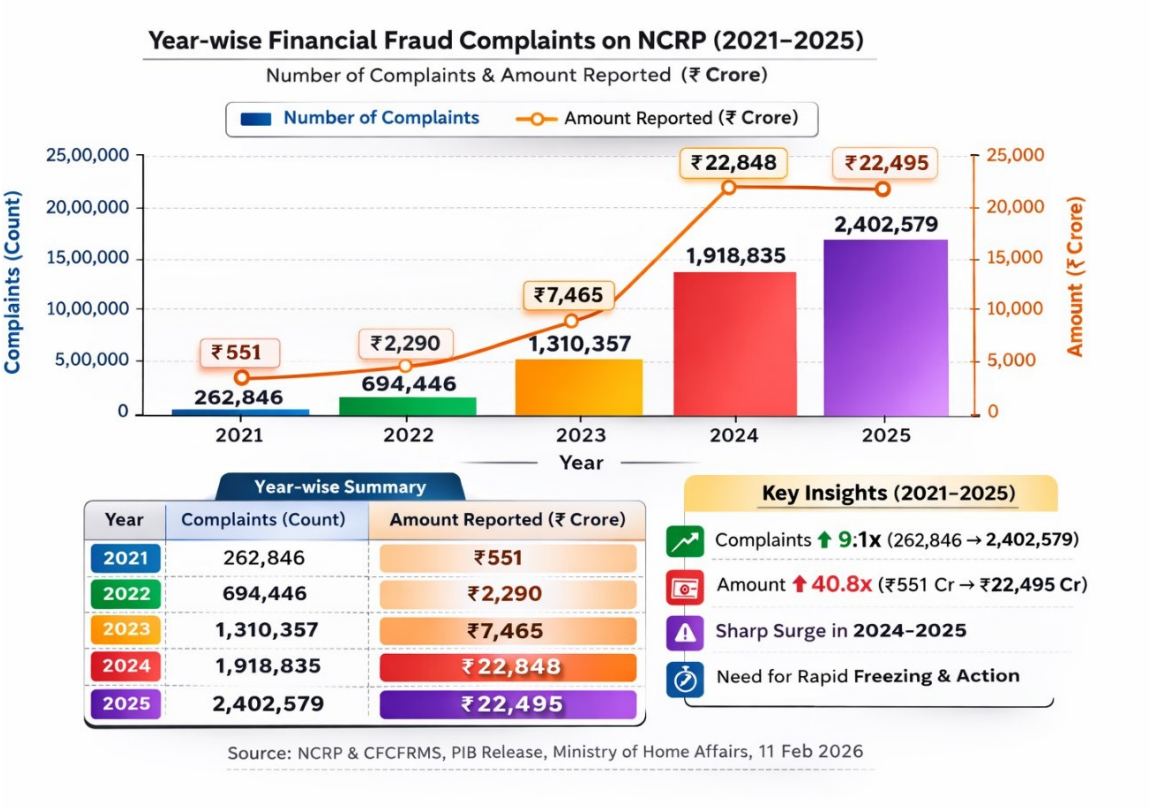


**Figure 1.** Year-wise NCRP financial fraud complaints and amounts reported by citizens.[25]

### 1.2.4 Impact and Risk Drivers

Scale, speed, anonymity, and cross, border infrastructure turn common fraud into systemic risk by drastically shortening the period between planning, execution, and harm. AI helps to bring down the marginal cost of repetition and allows for high, volume targeting even before complaint channels or platform processes get activated. Attribution is made difficult through the use of multiple accounts, rented infrastructure, and synthetic identities that break down the chain of evidence and make it hard to prove intent. Cross, border hosting and foreign corporate control of platforms make it difficult to preserve and disclose evidence, even if victims suffer financial losses and damage to their reputation. Deepfakes and automated manipulation present a higher risk to public order, as the former can be spread more rapidly than the time it takes for corrective speech or enforcement to reach the affected communities. In this context, legal certainty relies on quick coordination, well, defined criteria for intervention, and evidence, gathering mechanisms that can withstand the scrutiny of a court.[26]

[25] Cybercrime Reporting and Investigation, *available at:* https://www.pib.gov.in/PressReleasePage.aspx?PRID=2226441&lang=1&reg=3 (last visited on February 20, 2026).
[26] Zhixin Pan, Prabhat Mishra, *Explainable AI for Cybersecurity* 115 (Springer, Cham, 1st edn., 2024).

Figure 1 visually represent the speed-and-scale problem using official CERT-In incident totals, where we can see a clear structural rise in the number of cyber security incidents that have been reported to and are being tracked by the national agency. The path is consistent with the legal argument that late preservation decisions could lead to the loss of evidence, which is not only irreversible but also non-compensable harm.[27]

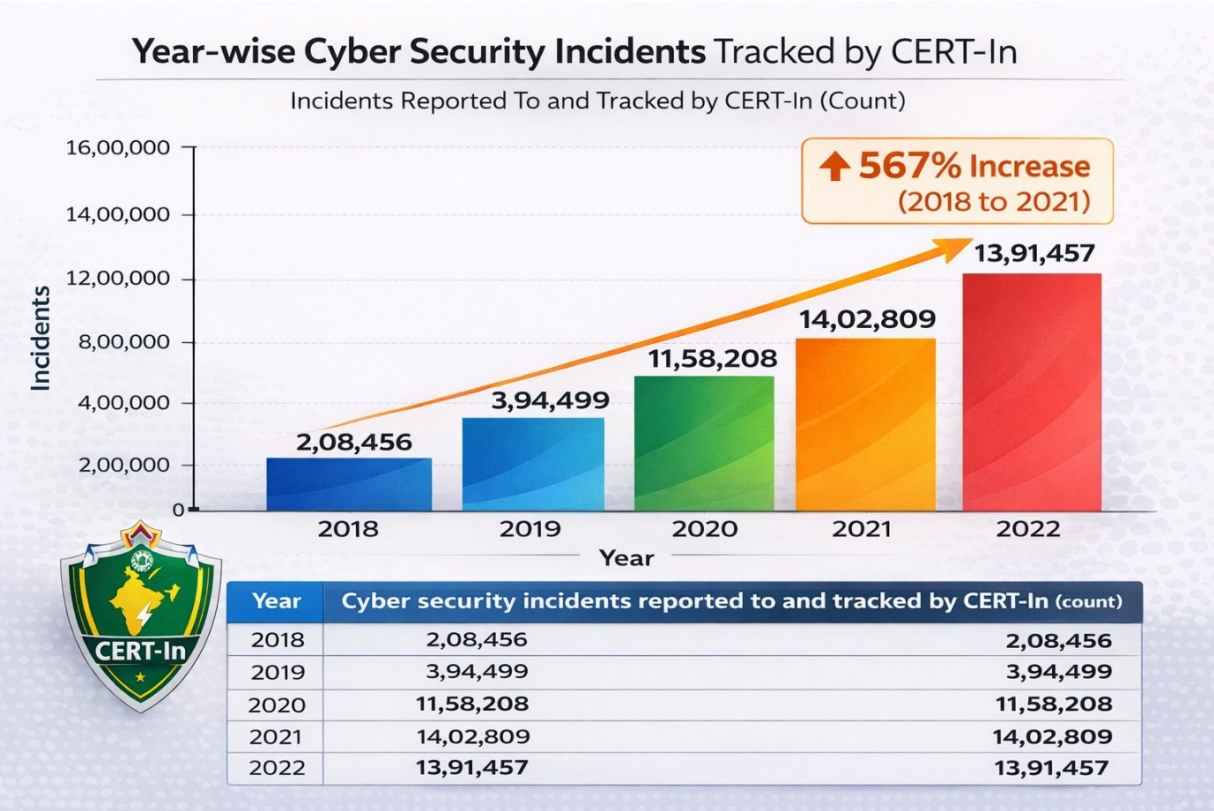


**Figure 2. Cyber security incidents "reported to and tracked" by CERT-In, year-wise totals (2018-2022).**[28]

## 1.3 INDIAN REGULATORY ARCHITECTURE

Indias legal war chest against AI-facilitated cybercrimes presently depends on the legal provisions for cyber offences, controls over intermediaries, data protection obligations, and updated criminal procedure and evidence rules. The structure is not just one bare AI law; it is a layered system where the "Information Technology Act, 2000, rules promulgated under it, and institutional mandates offer primary controls, while criminal law provides general

[27] Ishaani Priyadarshini, Rohit Sharma, *Artificial Intelligence and Cybersecurity: Advances and Innovations* 209 (CRC Press, Boca Raton, 1st edn., 2022).

[28] Safety of EV Charging Stations, *available at:* https://sansad.in/getFile/loksabhaquestions/annex/1711/AU2628.pdf?source=pqals (last visited on February 16, 2026).

categories of crimes such as fraud, intimidation, and organized conduct.[29] Whether the law provides adequate security depends on if these different layers create consistent obligations and allow law enforcement to act within set timeframes. The setup also puts to test the constitutional discipline, as the rapid introduction of measures can result in excessive nature without standards and review.[30]

### 1.3.1 Information Technology Act, 2000 and Delegated Rules

The Information Technology Act, 2000 is still the primary legal statute and the main statutory backbone of reference as it criminalizes major cyber-frauds, provides for civil liability, and defines the duties of intermediaries, etc. The Act is crafted in a technology-neutral way, thus it can be applied to any AI-enabled activities without the need for AI-specific provisions.[31] However, this neutrality is a double-edged sword: it is a strength for law enforcement but a weakness for governance, since AI not only presents new risks to model providers but also to the platforms where the tools are integrated. Delegated rules, advisory directions and compliance frameworks address the gaps in due diligence, grievance mechanisms and content governance expectations, respectively. Legal certainty is achieved through effective coordination of these policy instruments with the aim that enforcement becomes evidence-led and rights-consistent rather than merely reactive.[32]

#### *1.3.1.1 Substantive Cyber Offences and Penalties*

When AI is deployed to illegally access, damage, impersonate, or violate privacy through computer resources, substantive cyber-crimes provide the main route for charging. The key civil wrong of unauthorized access and interference is Section 43 of the Information Technology Act, 2000. The language of Section 43 reads without the permission of the owner or any other person who is in charge of a computer, computer system or computer network and this wording is in line with the charges against AI-based programs that scrape, brute-force, or probe systems on a large scale.[33] Later, criminal provisions come into play if theft of identity,

[29] Anumodan Tiwari, "AI-Powered Cybercrime Investigations under BNS", 7 *International Journal of Law Management & Humanities* 1479 (2024).
[30] Harshit Bidhuri, "Regulating Artificial Intelligence in India: Bridging the Legal Vacuum", 11 *International Journal of Law* 112 (2025).
[31] D. P. Mittal, *Law Relating to Information Technology, E-Commerce, E-Governance & Cyber Crimes* 204 (Commercial Law Publishers (India) Pvt. Ltd., Delhi, 1st edn., 2024).
[32] Shikher Deep Aggarwal, Kush Kalra, *Commentary on the Information Technology Act* 137 (Whitesmann, Delhi, 2nd edn., 2024).
[33] The Information Technology Act, 2000 (Act No. 21 of 2000), s. 43.

cheating by personation, invasion of privacy, obscenity, or cyber terrorism are supported by the facts of the case.[34]

#### *1.3.1.2 Intermediary Safe Harbour and Due Diligence*

The concept of safe harbor is vital for legal certainty, especially since AI-related damages are mostly disseminated through platforms and not directly produced by platform actors. Section 79 of the Information Technology Act, 2000 acts as a conditional immunity that allows the platform to continue its operations while imposing a certain level of diligence. Section 79 of the Act says that an intermediary shall not be liable for any third party information, data, or communication link made available or hosted by him.[35] The immunity granted will disappear if the platform operator is aware of the unauthorized content and does not take the necessary steps to remove it; clause (b) of sub-section (3) states fails to expeditiously remove or disable access to that material. AI-driven content distribution adds to the difficulty in assessing the platforms awareness, the issuing of a formal notice, and the taking of reasonable proactive measures.[36]

#### *1.3.1.3 Intermediary Guidelines 2021 and Compliance Design*

The Intermediary Guidelines framework seeks to operationalize statutory diligence by setting out measures, inter alia, for user grievance redressal, content takedown and transparency practices. Use of AI complicates the compliance landscape as online platforms can be hosts of both user-generated content and AI-generated content thereby making the distinction between mere hosting and actively shaping the content less clear.[37] Therefore, compliance frameworks need to consider how to incorporate mechanisms for identifying, labelling, and tracing synthetic content without imposing heavy burdens that would discourage lawful speech activities. Legal certainty gets a boost from compliance measures that are auditable, uniformly implemented, and can be traced back to documented decision rules. On the contrary, it is

[34] Sarvesh Raizada, "Constitutionality of Section 66 of the Information Technology Act, 2000", 15 *Pena Claims* 1 (2021).
[35] The Information Technology Act, 2000 (Act No. 21 of 2000), s. 79.
[36] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, *available at:* https://prsindia.org/billtrack/prs-products/the-information-technology-intermediary-guidelines-and-digital-media-ethics-code-rules-2021 (last visited on February 12, 2026).
[37] Puneet Bhasin, *Law Relating to Social Media Crimes, Intermediaries, Digital Media and OTT Platforms* 176 (OakBridge Publishing, Gurgaon, 1st edn., 2022).

undermined by measures that are arbitrary, non-transparent, or incapables in maintaining pre-takedown and account suspension evidence.[38]

#### *1.3.1.4 Amendments on Takedown Timelines and AI Labelling*

Short takedown windows are often justified by arguments for harm-minimization, nevertheless, they can weaken legal security if they compel platforms to make mistake-prone decisions due to the threat of liability exposure. In the case of deepfake and scam scenarios, a delay in takedown can change the situation of a single victim into one of mass victimization, thus, turning it into a matter of public order.[39] Using a label and disclosure for AI-generated or materially changed content is basically aimed at reducing the level of deception by users being able to identify authenticity through various cues. There is a risk of non-compliance if appending a label is unclear, if detection is technically flawed, or if attackers purposely remove watermarks and other signals. Legal security entails being clear about the categories that trigger the measures, having reasonable timelines, and providing for disputing channels when erroneous removals occur.[40]

### 1.3.2 Digital Personal Data Protection Act, 2023 and DPDP Rules, 2025

The Digital Personal Data Protection Act, 2023 reorients governance of AI deployment towards a method in which the collection, processing and breach response related to personal digital data used in training, profiling and targeting shall be regulated. It is indispensable to have consent since AI systems mostly obtain behavioral data and deduced preferences. Section 6 of the Digital Personal Data Protection Act, 2023 stipulates The consent given by the Data Principal shall be free, specific, informed, unconditional and unambiguous.[41] Security obligations are very significant from the point of operation in the AI incident response; Section 8 goes on to state If there is a personal data breach, the Data Fiduciary shall, within a reasonable time, inform the Board and each affected Data Principal, of such breach.[42] The Digital Personal Data Protection Rules, 2025 set the framework for meeting requirements for notices, consent management, and breach handling through government-issued instruments.

---

[38] G. V. S. Jagannadha Rao, *Ethics Code for Social Media, OTT and Digital Media: Commentary on Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021* 119 (Asia Law House, Hyderabad, 1st edn., 2021).

[39] Robert Gorwa, Michael Veale, "Moderating Model Marketplaces: Platform Governance Puzzles for AI Intermediaries", 16 *Law, Innovation and Technology* 341 (2024).

[40] Matthew B. Kugler, Carly Pace, "Deepfake Privacy: Attitudes and Regulation", 116 *Northwestern University Law Review* 611 (2021).

[41] The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023), s. 6.

[42] *Id.*, s. 8.

### 1.3.3 New Criminal Law Framework and Evidence Modernization

AI-enabled crimes reveal the biggest flaws more in the ... flows of investigations than in the definitions of crimes themselves. Updated criminal procedure frameworks are important because rapid harms demand quick preservation, seizure, and cross-platform coordination with legality constraints all coming into play. Modernizations of evidence is important because courts require trustworthy methods to separate authentic artefacts from synthetic fabrications, and also because defence strategies are increasingly focusing on disputing provenance and integrity instead of just the substance.[43] Legal security is enhanced when the steps of the investigation are documented, time-stamped, and after that, the judicial scrutiny is possible. It is decreased when preservation is delayed, when devices are imaged without a ... protocol, or when platform disclosures ... are not routed through legally valid channels capable of supporting admissibility.[44]

#### *1.3.3.1 Bharatiya Nyaya Sanhita, 2023*

Major offences still play a significant role in AI-assisted crimes as the damage is most of the time categorized into classic crimes: cheating, forgery, impersonation, criminal intimidation, extortion, and organized scam conduct. The Bharatiya Nyaya Sanhita, 2023, gives us the words for the intent-based crime where AI is just the tool used and not the crime itself.[45] AI-assisted impersonation, such as deepfake voice scams, can be considered cheating and personation if there is inducement and intention to deceive. AI-generated documents might be considered forgery and use-of-forged-document if material alteration and knowledge are proved. Legal requirement is that prosecuting options should correspond to evidence of intention and causation rather than simply considering the use of AI as guilt.[46]

#### *1.3.3.2 Bharatiya Sakshya Adhiniyam, 2023*

Electronic proof rules govern the extent to which prosecutions against deepfakes, synthetic logs, and AI-generated artefacts can be successful. The Bharatiya Sakshya Adhiniyam, 2023 is significant because it lays down the framework for the proving of electronic records, the conditions under which secondary electronic outputs may be admissible, and the certification

[43] Sarkar, The Bharatiya Sakshya Adhiniyam, *2023 (In 2 Volumes)* 96 (Kamal Law House, Kolkata, 1st edn., 2024).
[44] Y. P. Bhagat, Kumar Keshav, *Commentary on the Bharatiya Nyaya Sanhita, 2023* 188 (Whitesmann, Delhi, 1st edn., 2025).
[45] Arushi Bajpai, Akash Gupta, et.al., "An 'Indigenous' Criminal Code for India? The Bharatiya Nyaya Sanhita, 2023 and Its Resonances With the Past", 45 *Statute Law Review* 112 (2024).
[46] Kush V Trivedi, "Bhartiya Nyaya Sanhita: India's New Transformative Criminal Law", 11 *International Journal of Law* 1 (2025).

or reliability conditions that are expected.[47] AI creates a challenge in the scenario of this Act since the outputs may be generated without a stable "original" , and the logs may be fabricated to look like system records. Legal security is enhanced when investigators keep the source devices, record the metadata, and have verifiable hashes for chain-of-custody. On the other hand, it is compromised when the evidence is copied via uncontrolled channels or when the authenticity is claimed without having verifiable provenance documentation.[48]

### 1.3.4 Cybersecurity Institutions and Operational Mandates

Institutional coordination is extremely crucial for enforcement as AI incidents tend to cover a wide range of areas such as critical infrastructure, consumer fraud, platform governance and cross-border data flows. CERT-In is the agency entrusted with being the national incident response node, while the protection of critical infrastructure is done through specialized coordination frameworks. The Indian Cyber Crime Coordination Centre (I4C) aids in the integration of policing and reporting, and sector regulators influence compliance incentives through licensing and supervisory expectations.[49] The law enforcement aspect of security relies on coordination mechanisms which ensure that evidence is properly preserved, duplication of efforts is minimized, and platforms are not given conflicting directions. Investigations lose valuable time if there are no clear handoffs and shared standards; furthermore, victims end up getting mixed signals and platforms are burdened with fragmented legal demands which in turn hinder quick and lawful response.[50]

## 1.4 LEGAL SECURITY GAPS IN ENFORCEMENT AND FORENSICS

Legal security gaps emerge where the ability to enforce laws, as well as procedural safeguards, fail to keep up with the AI-enabled scale and synthetic deception. They are not an abstract concept; these gaps operate at the levels of attribution, evidence integrity, response speed, jurisdiction, and remedies. AI raises the level of doubt in the question of who acted, what was genuine, and when the damage was irreversible.[51] The result is a situation where the enforcement system is capable of over-removal and under-enforcement simultaneously,

[47] The Bhartiya Sakshya Bill, 2023, *available at:* https://prsindia.org/billtrack/prs-products/prs-bill-summary-4276 (last visited on February 7, 2026).
[48] The Bharatiya Sakshya Bill, 2023, *available at:* https://prsindia.org/billtrack/the-bharatiya-sakshya-bill-2023 (last visited on February 8, 2026).
[49] Harish Chander, Gagandeep Kaur, *Cyber Laws and IT Protection* 212 (PHI Learning Pvt. Ltd., Delhi, 2nd edn., 2022).
[50] Nilakshi Jain, Ramesh Menon, *Cyber Security and Cyber Laws* 151 (Wiley India Pvt. Ltd., New Delhi, 1st edn., 2021).
[51] Prasanna S, Lavanya P, "Digital Forensics and Cybercrime Investigation", 1 *ILE Journal of Law and Forensic Science* 1 (2023).

victims have to bear heavy burdens, and, at the same time, offenders exploit delays. The filling of these gaps will require doctrinal clarity, evidence-ready processes, and remedies tailored to the needs of digitals harms that are fast-moving as opposed to the slow civil timelines.[52]

### 1.4.1 Attribution and Intent in AI-Assisted Offences

Proof of mens rea becomes very difficult when AI is used to automate major steps, distribute decision-making across different tools, and at the same time allow offenders to claim that the system acted independently Additionally, one must infer intent from configuration decisions, prompt histories, planning documents, and transaction patterns, rather than a direct manual action.[53] Foreseeability becomes paramount: liability analysis depends on whether the harmful outputs were reasonably predictable from the system capabilities, safeguards, and user behavior. Legal security needs investigative methods that record the human "control plane", such as account recovery trails, API keys, payment links, and device associations. Missing these artefacts, the cases turn to speculation about automation, thus diminishing the fairness of both the prosecution and the defence.[54]

### 1.4.2 Evidence Integrity Against Synthetic Artefacts

Synthetic artefacts are a serious threat to chain-of-custody as they may be fabricated, altered, or generated independently of any stable source. Original recordings can be "deepfaked", synthetic documents can trick an official style, and tampered logs can fabricate a reasonable story of intrusion or consent.[55] The challenge of the forensic expert is not merely to detect; it is to demonstrate in court the integrity of the evidence, including the times and ways a file was created, stored, transmitted, and edited. Legal security should be ensured by standardized imaging, hashing, metadata capturing, and saving platform-side records. In cases where evidence is shared by means of messaging apps or screenshot recordings, authenticity commitments are risked greatly and trials become games of unsupported accusations.[56]

---

[52] Dr. Ram Prakash Chaubey, "Cybercrime Investigation in India: An Analysis of Digital Evidence and Its Role in Proving Cybercrimes", 7 *International Journal of Law, Policy and Social Review* 25 (2025).

[53] Artificial Intelligence, Criminal Infringement and Intent – Is the Criminal Law Ready?, *available at:* https://link.springer.com/article/10.1007/s40319-025-01592-7 (last visited on February 5, 2026).

[54] A Survey of Cyber Threat Attribution: Challenges, Techniques and Applications, *available at:* https://www.sciencedirect.com/science/article/pii/S0167404825002950 (last visited on February 6, 2026).

[55] Christian Rathgeb, Ruben Tolosana, et.al., *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks* 221 (Springer, Cham, 1st edn., 2022).

[56] Eoghan Casey, *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet* 164 (Academic Press, Boston, 3rd edn., 2011).

### 1.4.3 Platform Governance and Speed of Harm

These notice-based takedown policies have a hard time when the damage is done through virus-like spreading and when the copies get spread over several accounts and platforms within just a few minutes. Deepfake takedown after reputational damage does not restore privacy or dignity; the main harm is the exposure itself and the ongoing risk of the content being found again. Scam campaigns are also outpacing complaint channels by using rapid account churn and cheap throws of short-lived URLs.[57] Thus, a legal security issue can be seen as a matter of rights related to time-to-response because a delay in taking action can result in a harm that cannot be repaired. However, an ultra-fast removal carried out without due safeguards might risk arbitrary censorship and loss of evidence, which calls for the establishment of procedures that would make it possible to keep the evidence safe and at the same time allow for quick containment.[58]

### 1.4.4 Cross-border Investigation Constraints

Cross-border infrastructure creates jurisdictional and procedural friction, among other things, when data is stored internationally, platforms have their headquarters outside India, and intermediaries follow foreign disclosure standards. Mutual legal assistance methods may be lagging behind the speed of AI-driven campaigns, and preservation requests may not be uniformly honored. Data localization requirements may be at odds with platform architecture, while encryption and privacy policies make getting access to content and metadata more difficult.[59] Legal certainty demands that preservation and disclosure take place through channels that are both domestically legal and consistent with the norms of international co-operation. Otherwise, in the absence of this, works will merely rely on partial records, perpetrators will take advantage of overlaps between jurisdictions, and victims will be in a state of prolonged uncertainty with little chance of getting their losses reimbursed.[60]

### 1.4.5 Victim Protection and Remedies

Victim remedies should be in line with the time duration of AI harms which, in many cases, require assistance at the earliest time rather than after the damages. Quick court orders,

[57] Tarleton Gillespie, "Content Moderation, AI, and the Question of Scale", 7 *Big Data & Society* 1 (2020).
[58] Robert Gorwa, Reuben Binns, et.al., "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance", 7 *Big Data & Society* 1 (2020).
[59] Cross Border Data-Sharing and India, *available at:* https://cis-india.org/internet-governance/files/mlat-report (last visited on February 3, 2026).
[60] Electronic Evidence Country Fiche: India, *available at:* https://www.unodc.org/cld/uploads/pdf/El%20Evidence%20Hub/SOUTH_ASIA/eEvidence_Country_Fiche_INDIA.pdf (last visited on February 4, 2026).

extremely rapid injunctions, and coordinated takedown requests are still the main actions when deepfakes and impersonation put the threat of loss of livelihood and safety.[61] Compensation rule should provide for the covering of direct financial loss, consequential damages due to reputational harm, and expenses of remediation such as recovering an account and restoring one's identity. Restorative avenues such as helplines and platform support channels are important as victims usually require guidance for evidence preservation and banking coordination. Legal certainty is enhanced when remedy channels are easy to use, uniform across countries, and backed by predictable documentation requirements.[62]

## 1.5 DOCTRINAL CASE LAW ANALYSIS AND REGULATORY IMPLICATIONS

Judicial doctrine can secure the law by delineating the limits of the constitution, enumerating the duties of the intermediaries, and regulating electronic evidence. By their rulings, the courts also guide how a return can be made effectively, especially in cases of urgency when one identity has been misused, and media has been synthesized. In the field of AI, the case law is like a regulating mechanism: it limits the powers of the administration, requires the observance of due process, and offers the criteria for the actions of the platforms.[63] Moreover, it points out the areas where the wording of the legislation and the executions of the rules have to be very clear to be able to withstand the test of the constitutionality. The following cases demonstrate how the Indian judiciary is currently playing a role in creating the framework for the rapid use of AI and the encounters with cybercrime.[64]

### 1.5.1 Doctrinal Lenses for AI Cybercrime Regulation

A systematic method involves five interconnected perspectives: privacy, speech, due process, platform responsibility, and evidentiary reliability. Privacy laws set the boundaries for data-driven profiling, surveillance methods, and overly aggressive response tactics, and at the same time, they help to understand the damage caused by the composition of deepfake and impersonation. Speech laws restrict unnecessarily broad content controls and require that prohibited categories be defined with great care.[65] Due process laws point to the need for

[61] R. Sriniwas, *Law Relating to Crime Victims Compensation* 198 (Orient Publishing Company, New Delhi, 2nd edn., 2024).
[62] Debarati Halder, K. Jaishankar, *Cyber Crimes Against Women in India* 142 (SAGE Publications India, New Delhi, 1st edn., 2016).
[63] Abhishek Gotety, "Analysing Regulatory Regimes for AI-Based Legal Technology: An Indian Perspective", 14 *NUJS Law Review* 3 (2021).
[64] David C. Vladeck, "Machines Without Principals: Liability Rules and Artificial Intelligence", 89 *Washington Law Review* 117 (2014).
[65] OECD AI Principles Overview, *available at:* https://oecd.ai/en/ai-principles (last visited on February 11, 2026).

communication of motives in orders, the possibility of review, and fair procedures for removal and blocking. Platform responsibility laws refer to the distribution of risk among intermediaries and service providers. Evidentiary reliability laws guarantee that criminal prosecutions and court interventions are based on real evidence rather than on synthetic or tampered artefacts.[66]

### 1.5.2 Landmark and Recent Indian Case Laws Shaping AI-Era Cyber Regulation

The doctrinal reference of landmark constitutional cases as well as the recent injunction practice is very central to AI-era legal security, since the regulation must be both effective and bounded. Constitutional judgments place limits on executive and delegated instruments that could have otherwise been used to impose vague speech controls or unreviewable surveillance. Evidence decisions affect whether AI-manipulated records can be demonstrated with integrity.[67] Recent personality and publicity rights orders show how courts use the power of urgent relief against synthetic misuse of likeness and voice, most of the time through platform directions and John Doe-style approaches. Such rulings determine the governance of rapid AI deployment: by outlining permissible state action, private compliance duties, and proof thresholds.[68]

#### *1.5.2.1 Shreya Singhal v. Union of India, (2015) 5 SCC 1*

In the case of *Shreya Singhal v. Union of India*[69], Supreme Court of India considered the case of prosecuting persons under Section 66A of the Information Technology Act, 2000 for posting on social media and their consequent arrest on the ground of vague offense language. The Court, therefore, declared that Section 66A is unconstitutional.[70] The decision is significant for AI regulation in that broadly worded restrictions on false or offensive content may lead to a speech-chilling effect when applied to synthetic media. The rule of law requires small, tightly defined categories and court-monitored removal and blocking actions.

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

In the case of *Justice K.S. Puttaswamy (Retd.) v. Union of India*[71], the Supreme Court of India dealt with the constitutional challenges against State actions that touched personal autonomy

[66] AI Risk Management Framework, *available at:* https://www.nist.gov/itl/ai-risk-management-framework (last visited on February 12, 2026).
[67] Shikher Deep Aggarwal, Kush Kalra, *Commentary on the Information Technology Act, 2000* 119 (Whitesmann Publishing Co., Delhi, 2nd edn., 2024).
[68] Pavan Duggal, *Cyber Law: An Exhaustive Section Wise Commentary on the Information Technology Act, 2000* 173 (Universal Law Publishing, Delhi, 2nd edn., 2017).
[69] (2015) 5 SCC 1.
[70] The Information Technology Act, 2000 (Act No. 21 of 2000), s. 66A.
[71] (2017) 10 SCC 1.

and control of personal data. The Court recorded privacy as a fundamental right. The decision is crucial in the context of AI deployment since large-scale profiling, behavioral targeting, and surveillance tooling are dependent on personal data inference and aggregation. In case of AI-enabled cyber harm, the doctrine identifies dignity and autonomy as recognized legal interests, thus it supports the restricted collection of data, the limiting of purposes, and the accountability of data misuse, and at the same time, it requires the legality and proportionality of investigative intrusions.

#### *1.5.2.3 Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, 2020 SCC Online SC 571*

In the case of *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal*[72], the Supreme Court was dealing with cases relating to elections where the attached electronic records were used as evidence without being properly certified, thus creating the issue of their admissibility and proof. The Court ruled that "the certificate required under Section 65B(4) is mandatory". This requirement is very important for situations involving AI because appropriately deepfakes and synthetic logs are the main issues of authenticity. It is beneficial for the law to stipulate insistence on strict proof conditions by the courts that thereby are securing the main legal effect, that being the fixation of source integrity by investigators and litigants, the documentation of device provenance, and the avoidance of the use of screenshots and forwarded media as an alternative to the submission of electronic proof that is admissible.

#### *1.5.2.4 Manohar Lal Sharma v. Union of India, 2021 SCC Online SC 985*

In the case of *Manohar Lal Sharma v. Union of India*[73], the Supreme Court of India addressed the allegations raised about the use of spyware to target individuals and the States reliance on national security as a reason to deny disclosure. The Court directed an institutional inquiry saying we constitute a Committee to investigate the allegations. The point for AI-enabled surveillance is that the mere usage of security claims cannot be a reason to escape judicial review, and hence the intrusion of digital tools need to be checked by appropriate oversight mechanisms. Legal security for speedy deployment must consist of reviewable authorization, the least possible collection, and the establishment of adequate safeguards that will protect both the legitimacy of security purposes and the enjoyment of fundamental rights.

[72] 2020 SCC Online SC 571.
[73] 2021 SCC Online SC 985.

#### *1.5.2.5 Amitabh Bachchan v. Rajat Nagi, 2022 SCC Online Del 4110*

In the case of *Amitabh Bachchan v. Rajat Nagi*[74], the Delhi High Court came down on the unauthorized use of the plaintiff's name, image, and persona through online media channels, thus exposing the plaintiff to risks of deception as well as commercial misappropriation. The Court issued a prohibitory order, stating "an ad interim injunction is granted." The ruling is pertinent to AI-related crimes since deepfake celebrity endorsements and voice cloning frauds use the victim's persona to get money from them. Besides that, legal protection becomes more effective when courts perceive the misuse of a public figure's persona to be a harm of the first degree, hence allowing a quick removal and restraint without any loss of evidence and at the same time giving chance to platform-directed compliance.

#### *1.5.2.6 Anil Kapoor v. Simply Life India & Ors, 2023 SCC Online Del 6914*

In the case of *Anil Kapoor v. Simply Life India & Ors*[75], the Delhi High Court dealt with the issue of plaintiff's image and character being used without permission in ways which could deceive people and harm the reputation. The Court provided protective relief by stating that "an injunction is granted" against the misuse and unauthorized exploitation. There is a direct connection to AI deployment: generative synthetic likeness and voice cloning technologies make it more frequent and believable that someone's persona will be misused on different platforms. From a legal point of view, the courts play a crucial role in the timely and effective issuing of orders, which are the basis for ensuring takedown, stopping further distribution, and holding platforms accountable for providing access to their records to pinpoint the source of the violation.

### 1.5.3 Liability and Accountability Map for AI Actors

As control is shared, the accountability of AI-enabled cybercrime can transition among developers, deployers, users, and intermediaries. Developers determine the default safeguards, logging, and misuse resistance; deployers decide the access control, rate limiting, and identity verification; users provide the intent and operational direction; intermediaries are the ones to mediate the dissemination and monetization.[76] Legal security should be a mapping exercise that avoids the two extremes: strict liability for mere capability on one side and immunity that disregards foreseeable harm on the other. Accountability should follow control, knowledge,

[74] 2022 SCC Online Del 4110.
[75] 2023 SCC Online Del 6914.
[76] Athina Sachoulidou, "AI Systems and Criminal Liability: A Call for Action", 11 *Oslo Law Review* 1 (2024).

and reasonable preventive capacity. In situations where platforms not only host content but also generate it, the role of the intermediary combines with that of the active participant, thus, the demand for governance, transparency, and the preservation of evidence is even greater.[77]

#### *1.5.3.1 Criminal Liability Triggers*

Criminal exposure rises when AI is exploited as a means to perform traditional illegal acts on a large scale, particularly where victims are fraudulently tricked into transfers, coerced, or sexually exploited. Determination of guilt usually depends on evidence of dishonest intention, knowledge, inducement, and participation in coordinated conduct like mule networks. The use of AI may constitute an aggravating factor if it leads to an increased victim vulnerability, targets children, or facilitates organized criminal operations through automation.[78] Legal clarity demands a strict distinction between the mere availability of a tool and the culpable use thereof, which should be corroborated by evidence such as planning, prompts, transaction trails, and coordination. Extensively criminalizing AI experimentation in the absence of victim harm is highly likely to lead to arbitrary enforcement and such a situation is detrimental to the overall legitimacy of the law.[79]

#### *1.5.3.2 Civil Remedies and Urgent Injunction Practice*

Civil remedies are a quick way to stop things going out of control if the criminal procedures are so slow that the damage cannot be reversed anymore. Courts increasingly base their decisions on interim ex parte measures, including John Doe order such as where the defendants are unknown and the distribution is still going on. In deepfake and impersonation cases, court orders are used as preventive relief measures directed at the restraint, removal, and prohibition of access through accounts and mirror links.[80] The logic of damages is often relegated to the background of immediate protection, yet compensation is still a matter of concern for the proven loss and the cost of the remedy. The law depends on the orders of the court being very particular, capable of being carried out, lasting for a limited time where this is necessary, and

[77] Gyandeep Chaudhary, "Artificial Intelligence: The Liability Paradox", 1 *ILI Law Review* 144 (2020).
[78] Combating Misinformation and Deepfakes: Directive for Intermediaries' Compliance of IT Rules, *available at:* https://www.argus-p.com/updates/updates/combating-misinformation-and-deepfakes-meitys-directive-for-intermediaries-compliance-of-it-rules/ (last visited on February 5, 2026).
[79] Advisory to All Intermediaries to Comply With IT Rules on Deepfakes, *available at:* https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=1990542 (last visited on February 6, 2026).
[80] K. M. Sharma, Stay Orders, *Temporary Injunctions & Interlocutory Orders* 211 (Kamal Publishers, New Delhi, 3rd edn., 2022).

accompanied by preservation directions which prevent the loss of evidence during the removal.[81]

#### *1.5.3.3 Regulatory Compliance and Audit Expectations*

Regulatory expectations are moving toward demonstrable governance: incident reporting, risk documentation, and verifiable controls rather than policy statements. For AI deployment, legal security is improved when model-risk documentation covers training data provenance, access control, monitoring for misuse, and escalation pathways for harm reports.[82] User verification becomes relevant where services are attractive to scammers, particularly for agentic tools with transaction capability. Compliance also requires harmonisation with data protection duties, so that logging and monitoring remain purpose-limited and secure. Where audit trails are absent or unverifiable, enforcement becomes speculative and platforms face heightened risk of adverse inference in both civil and regulatory proceedings.[83]

## 1.6 CONCLUSION

AI-enabled cybercrime creates a need for an effective legal security framework that is capable of matching the rapid pace at which digital harm occurs, without losing the rule of law. India's legal system can respond to this challenge by seeing AI primarily as a factor that escalates existing cyber and other crimes, and at the same time, it should focus on the areas where AI concentrates risks such as model access, tool integration, and platform amplification, to clarify the duties. The Information Technology Act, 2000 continues to be the main source of offences and the role of intermediaries, with the conditional structure of Section 79 allowing for a calibrated platform responsibility instead of absolute immunity or absolute liability.[84] The data protection law provides a parallel mechanism for handling matters such as profiling, breach response, and lawful processing, thereby making prevention compatible with privacy.

Judicial doctrine sets the controlling framework: principles limiting speech vagueness, e. g. , speech limitation, restrict synthetic media regulation; principles of privacy govern the extent of surveillance and profiling; principles of the use of electronic evidence in court require strict discipline of authentication and certification. The examples of cases at hand show how courts are capable of granting fast relief, especially in situations of personal misuse, and at the same

[81] C. K. Takwani, *Civil Procedure* 187 (Eastern Book Company, Lucknow, 8th edn., 2017).
[82] Margot E. Kaminski, "Algorithmic Impact Assessments Under the GDPR", 11 *International Data Privacy Law* 125 (2021).
[83] Emily P. Goodman, "Algorithmic Auditing: Chasing AI Accountability", 39 *Santa Clara High Technology Law Journal* 297 (2023).
[84] The Information Technology Act, 2000 (Act No. 21 of 2000), s. 79.

time, their standards ensuring the quality of the case review and fairness are preserved. Legal security gets a boost when courts and regulators demand evidence-preserving takedowns, reasoned directions, and tightly defined prohibitions. On the other hand, it is weakened when the need for speed results in the creation of vague categories, the issuance of unclear takedown orders, or the use of investigatory shortcuts that lead to the collapse of trial admissibility.[85]

Alignment across laws, platform rules, and institutional mandates can be achieved by three integrations. First, rapid response should be combined with mandatory evidence preservation and recording of decision trails to avoid the AI era's synthetic evidence problem that might help the perpetrators of violations to get away from enforcement. Second, data protection compliance should be incorporated into AI governance in such a way that consent, purpose limitation, and breach notification duties serve as preventive controls rather than post-harm formalities. Third, through institutional coordination via incident response and cybercrime coordination nodes, there must be agreement on the respective thresholds and handoffs across sectors This integrated approach enables fast AI deployment under enforceable duties, judicially limited discretion, and dependable proof standards.[86]

## 1.7 SUGGESTIONS

In line with the article's focus on AI-accelerated cyber harms and the legal security needed to govern rapid deployment, the following measures strengthen enforceability without diluting constitutional safeguards.

1. Create a statutory AI deployment duty for entities that provide model access, tool-integrations, or synthetic-media generation features, mapped to their degree of control and foreseeability of misuse. Implement it through tiered due-diligence schedules for intermediaries and through mandatory security clauses for enterprise model providers. A predetermined Compliance Officer will be responsible for signing quarterly risk statements covering topics such as access controls, abuse monitoring, and incident response. This will decrease enforcement overreliance on end-users, while also making upstream negligence legally apparent.
2. Mandate that an evidence-preservation "hold" first be implemented when a platform receives a court order, a reasoned government intimation, or a credible victim grievance

[85] The National AI Portal of India, *available at:* https://indiaai.gov.in/indiaaiportal (last visited on February 4, 2026).
[86] National Cyber Crime Reporting Portal, *available at:* https://cybercrime.gov.in/Accept.aspx (last visited on February 3, 2026).

about deepfakes, NCII, or impersonation. The hold should entail hashing the content, saving the original upload file, and recording the first uploader account, device identifiers, and distribution links before deletion. Set a standard retention period (e.g. 180 days) and limit access only to legitimate requests with unalterable audit trails. This keeps the evidences which can be used in court while also fulfilling the requirement of disabling within three hours in the amended takedown timeline.

3. Implement a standard "SGI origin bundle" that platforms should fix with each tagged synthetic item, which should entail persistent metadata, unique IDs, and a cryptographic signature of the generation service. The bundle should be required to remain as a whole when the content is downloaded, re-uploaded, or forwarded, with an indicator showing tampering if the stripping happens. The bundle should be made available in a secure portal to investigators and courts so that authenticity can be checked without a huge disclosure of user content. This makes the labeling and metadata functions operationally related to synthetically generated information that has been introduced.
4. Implement risk-based access controls for agentic AI systems that may perform transactions, gather data, or communicate with third-party tools. Capabilities with a high level of risk should be accompanied by more robust identity verification, step-up authentication, and hard rate limits, whereas low-risk creative functions may still be kept friction-light. Platforms must use least-privilege tool permissions and default-deny outbound operations unless the user especially authorizes every step. Such measures not only minimize the effects of prompt injection but also make automated laundering significantly more difficult to scale.
5. Aligning the platform logging with the DPDP framework can be achieved by establishing a minimum set of security logs which are purpose-limited, encrypted, and retained only for defined investigation windows. Any expansion of monitoring or profiling to detect scams or synthetic media should be subject to privacy impact assessments. In instances where logs contain personal data, there should be a requirement for role-based access controls, breach notification playbooks, and deletion schedules that are in line with the DPDP Rules, 2025. This helps in providing attribution while at the same time, lowers the risk of surveillance overreach.
6. Create a unified incident-reporting pipeline that marries CERT-In's six-hour reporting mandate and the DPDP breach notification requirements, along with platform takedown mechanisms. Employ one single machine-readable incident report that records technical indicators, impacted users, preservation measures, as well as financial fraud

methods. Set up automated routing such that consumer-scale fraud is directed towards cybercrime coordination channels whereas sector-specific attacks are routed to incident-response teams, along with clear handoff timestamps. This not only helps to reduce the duplication of efforts but also enhances the time-to-containment in multi-platform events.

7. Quickly implement restitution by mandating banks, UPI apps, and wallet providers to facilitate a "golden hour" freeze protocol for suspected AI-assisted payment fraud. The protocol must enable temporary holds on transactions that have been identified as red-flag triggers, with fast supervisory checks to prevent arbitrary blocking. Standardize communication between providers to allow for the identification of mule accounts to be flagged across banks while still keeping evidence for the prosecution. It is a measure to reduce non-recoverable losses which have been running ahead of both the criminal process and civil remedies.
8. Release national forensic SOPs for deepfakes and synthetic documents that clarify how to clone devices, record metadata, and create hashes, and require accredited labs for disputed cases. Judges and police officers should be taught to consider screenshots and circulated videos only as leads for further investigation, and not as direct evidence. Promote the practice of courts requiring Section 65B certification and support documentation at the earliest stages, also in emergency relief situations if possible. This will significantly reduce the number of authenticity issues and strengthen the prosecution cases.
9. Strengthening cross-border preservation through the establishment of a 24/7 'rapid preservation' communication mechanism between major platforms, model providers, and cloud hosts, alongside the use of standardized request templates. Employ time-bound preservation orders which may be subsequently transformed into formal mutual legal assistance requests, thus lessening evidence loss as a result of jurisdictional delays. Make it obligatory for platforms to issue standardized transparency reports on the number of preservation requests that were received, complied with, or rejected, with the use of standard reason codes. Such measures turn jurisdictional frictions into procedural challenges rather than obstacles that end up undermining the case.
10. Ensure speed does not undermine the due process by making it obligatory that every removal instruction of three hours or less be justified, signed electronically, and recorded in a register that is searchable and supports later judicial review. Users should be provided with a quick appeal channel and a separate channel for victims to receive

the evidence copy so that in the rush neither the right to free speech nor the remedies are sacrificed. Publish quarterly metrics on false positives, reinstatements, and time-to-action for SGI and scam categories, and use them to refine thresholds. This renders the quick controls accountable and brings them in line with constitutional speech and privacy discipline as interpreted by the leading Supreme Court doctrine.

## BIBLIOGRAPHY

### PRIMARY SOURCES

#### 1. Statutes


- The Bharatiya Nagarik Suraksha Sanhita, 2023 (Act No. 46 of 2023)
- The Bharatiya Nyaya Sanhita, 2023 (Act No. 45 of 2023)
- The Bharatiya Sakshya Adhiniyam, 2023 (Act No. 47 of 2023)
- The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023)
- The Information Technology Act, 2000 (Act No. 21 of 2000)


### SECONDARY SOURCES

#### 2. Books


- Aggarwal, S. D., Kalra, K., *Commentary on the Information Technology Act* (Whitesmann, Delhi, 2nd edn., 2024).
- Aggarwal, S. D., Kalra, K., *Commentary on the Information Technology Act, 2000* (Whitesmann Publishing Co, Delhi, 2nd edn., 2024).
- Bhagat, Y. P., Keshav, K., *Commentary on the Bharatiya Nyaya Sanhita, 2023* (Whitesmann, Delhi, 1st edn., 2025).
- Bhasin, P., *Law Relating to Social Media Crimes, Intermediaries, Digital Media and OTT Platforms* (OakBridge Publishing, Gurgaon, 1st edn., 2022).
- Casey, E., *Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet* (Academic Press, Boston, 3rd edn., 2011).
- Chander, H., Kaur, G., *Cyber Laws and IT Protection* (PHI Learning Pvt. Ltd, Delhi, 2nd edn., 2022).
- Chandrasekhar, B. K., *Civil Procedure* (Eastern Book Company, Lucknow, 8th edn., 2017).
- Chio, C., Freeman, D., *Machine Learning and Security: Protecting Systems with Data and Algorithms* (O'Reilly Media, Sebastopol, 1st edn., 2018).
- Duggal, P., *Cyber Law: An Exhaustive Section Wise Commentary on the Information Technology Act, 2000* (Universal Law Publishing, Delhi, 2nd edn., 2017).
- Hadnagy, C., *Social Engineering: The Science of Human Hacking* (John Wiley & Sons, Hoboken, 2nd edn., 2018).

- Halder, D., Jaishankar, K., *Cyber Crimes Against Women in India* (SAGE Publications India, New Delhi, 1st edn., 2016).
- Jagannadha Rao, G. V. S., *Ethics Code for Social Media, OTT and Digital Media: Commentary on Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021* (Asia Law House, Hyderabad, 1st edn., 2021).
- Jain, N., Menon, R., *Cyber Security and Cyber Laws* (Wiley India Pvt. Ltd, New Delhi, 1st edn., 2021).
- Jakobsson, M., Myers, S., *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft* (John Wiley & Sons, Hoboken, 1st edn., 2007).
- Pan, Z., Mishra, P., *Explainable AI for Cybersecurity* (Springer, Cham, 1st edn., 2024).
- Phoenix, J., Taylor, M., *Prompt Engineering for Generative AI* (O'Reilly Media, Sebastopol, 1st edn., 2024).
- Priyadarshini, I., Sharma, R., *Artificial Intelligence and Cybersecurity: Advances and Innovations* (CRC Press, Boca Raton, 1st edn., 2022).
- Rathgeb, C., Tolosana, R., et al., *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks* (Springer, Cham, 1st edn., 2022).
- Sarkar, The Bharatiya Sakshya Adhiniyam, *(In 2 Volumes)* (Kamal Law House, Kolkata, 1st edn., 2024).
- Sharma, K. M., Stay Orders, *Temporary Injunctions & Interlocutory Orders* (Kamal Publishers, New Delhi, 3rd edn., 2022).
- Sikos, L. F., *AI in Cybersecurity* (Springer, Cham, 1st edn., 2019).
- Sikos, L. F., *AI-Enabled Cybersecurity* (Springer Nature Switzerland, Cham, 1st edn., 2021).
- Sriniwas, R., *Law Relating to Crime Victims Compensation* (Orient Publishing Company, New Delhi, 2nd edn., 2024).
- Vorobeychik, Y., Kantarcioglu, M., *Adversarial Machine Learning* (Morgan & Claypool Publishers, San Rafael, 1st edn., 2018).
- Wilson, S., *The Developer's Playbook for Large Language Model Security* (O'Reilly Media, Sebastopol, 1st edn., 2024).

## 3. Articles


- Avgerinos, T., Cha, S. K., et al., "Automatic Exploit Generation", 57 *Communications of the ACM* 74 (2014).
- Bajpai, A., Gupta, A., et al., "An 'Indigenous' Criminal Code for India? The Bharatiya Nyaya Sanhita, 2023 and Its Resonances With the Past", 45 *Statute Law Review* 112 (2024).
- Bidhuri, H., "Regulating Artificial Intelligence in India: Bridging the Legal Vacuum", 11 *International Journal of Law* 112 (2025).
- Chaubey, R. P., "Cybercrime Investigation in India: An Analysis of Digital Evidence and Its Role in Proving Cybercrimes", 7 *International Journal of Law, Policy and Social Review* 25 (2025).
- Chaudhary, G., "Artificial Intelligence: The Liability Paradox", 1 *ILI Law Review* 144 (2020).
- Chopra, A., Shukla, A., "Exploring the Misuse of Deepfake Technology in India: Implications for Society", 16 *Global Media Journal-Indian Edition* 1 (2024).
- Dhaka, N. K., Kumar, N., "Cybercrime in India: Law, Policy, & Enforcement in AI Era", 15 *Indian Streams Research Journal* 1 (2025).
- Doan, T. M., Khoa, N., et al., "AAGAN: Android Malware Generation System Based on the Adversarial Autoencoder and GAN", 11 *Vietnam Journal of Computer Science* 275 (2024).
- Gillespie, T., "Content Moderation, AI, and the Question of Scale", 7 *Big Data & Society* 1 (2020).
- Goodman, E. P., "Algorithmic Auditing: Chasing AI Accountability", 39 *Santa Clara High Technology Law Journal* 297 (2023).
- Gorwa, R., Binns, R., et al., "Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance", 7 *Big Data & Society* 1 (2020).
- Gorwa, R., Veale, M., "Moderating Model Marketplaces: Platform Governance Puzzles for AI Intermediaries", 16 *Law, Innovation and Technology* 341 (2024).
- Gotety, A., "Analysing Regulatory Regimes for AI-Based Legal Technology: An Indian Perspective", 14 *NUJS Law Review* 3 (2021).
- Kaminski, M. E., "Algorithmic Impact Assessments Under the GDPR", 11 *International Data Privacy Law* 125 (2021).

- Kugler, M. B., Pace, C., "Deepfake Privacy: Attitudes and Regulation", 116 *Northwestern University Law Review* 611 (2021).
- Madhumitha, P. S., "A Study on the Role of Artificial Intelligence in Preventing Cyber Crimes - Indian and International Perspective", 4 *Indian Journal of Integrated Research in Law* 836 (2024).
- Maramreddy, Y. R., Muppavaram, K., "Detecting and Mitigating Data Poisoning Attacks in Machine Learning: A Weighted Average Approach", 14 *Engineering, Technology & Applied Science Research* 15505 (2024).
- Raizada, S., "Constitutionality of Section 66 of the Information Technology Act, 2000", 15 *Pena Claims* 1 (2021).
- Sachoulidou, A., "AI Systems and Criminal Liability: A Call for Action", 11 *Oslo Law Review* 1 (2024).
- Schmitt, M., "Digital Deception: Generative AI in Social Engineering Attacks", 57 *Artificial Intelligence Review* 324 (2024).
- Tian, Z., Chen, L., et al., "A Comprehensive Survey on Poisoning Attacks and Countermeasures in Machine Learning", 55 *ACM Computing Surveys* 166 (2022).
- Tiwari, A., "AI-Powered Cybercrime Investigations under BNS", 7 *International Journal of Law Management & Humanities* 1479 (2024).
- Trivedi, K. V., "Bhartiya Nyaya Sanhita: India's New Transformative Criminal Law", 11 *International Journal of Law* 1 (2025).
- Vladeck, D. C., "Machines Without Principals: Liability Rules and Artificial Intelligence", 89 *Washington Law Review* 117 (2014).


## 4. Websites


- https://cis-india.org
- https://cybercrime.gov.in
- https://genai.owasp.org
- https://indiaai.gov.in
- https://link.springer.com
- https://oecd.ai
- https://owasp.org
- https://prsindia.org
- https://sansad.in

- https://www.argus-p.com
- https://www.europol.europa.eu
- https://www.interpol.int
- https://www.meity.gov.in
- https://www.nist.gov
- https://www.npci.org.in
- https://www.pib.gov.in
- https://www.rbi.org.in
- https://www.sciencedirect.com
- https://www.unodc.org