ARTIFICIAL INTELLIGENCE-ENABLED CYBERCRIME: RETHINKING CRIMINAL LIABILITY, DIGITAL EVIDENCE AND CYBER LAW IN INDIA

Eshan Singhal, Ph.D., Research Scholar, Faculty of Law,
Dr. Bhimrao Ambedkar University, Agra

Dr. Shiti Kanth Dubey, Professor, Faculty of Law,
Dr. Bhimrao Ambedkar University, Agra

ABSTRACT

Artificial intelligence has fundamentally transformed the technological environment in which cybercrime is conceived, executed and investigated. Generative artificial intelligence, large language models, deepfake systems, automated vulnerability discovery, synthetic identities and AI-assisted social engineering can substantially increase the speed, scale and sophistication of criminal activity. At the same time, the increasing autonomy of artificial intelligence creates difficult questions concerning attribution, mens rea, causation, criminal participation and evidentiary reliability. This paper examines the emerging phenomenon of artificial intelligence-enabled cybercrime from an Indian doctrinal and comparative legal perspective. It evaluates the extent to which the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023, the Bharatiya Nagarik Suraksha Sanhita, 2023 and the Bharatiya Sakshya Adhiniyam, 2023 can address offences committed or substantially facilitated through artificial intelligence. The paper distinguishes between AI-assisted, AI-enhanced, AI-automated and increasingly autonomous cybercrime and argues that the degree of technological involvement has direct implications for attribution and criminal responsibility. It further examines emerging challenges involving deepfake fraud, automated phishing, synthetic identities, AI-generated malware, financial deception and manipulation of digital evidence. A comparative analysis of the European Union's risk-based artificial-intelligence regulatory framework, the United Kingdom's approach and selected United States developments is undertaken to identify regulatory lessons for India. The paper argues that Indian cyber law does not presently require an entirely separate offence of "AI crime"; instead, it requires an adaptive framework addressing attribution, developer and deployer responsibility, evidence preservation, platform accountability and international cooperation. The study proposes an AI-Cybercrime Liability Framework based upon technological involvement, foreseeability, control, intent and resulting harm.

Keywords: Artificial Intelligence; Cybercrime; Generative AI; Digital Evidence; Cyber Law;

I. INTRODUCTION

Artificial intelligence ("AI") has rapidly evolved from a specialised technological field into a general-purpose technology influencing commerce, communication, finance, healthcare, education, national security and criminal justice[1]. Generative AI systems can produce text, images, audio-video and computer code with increasing sophistication. Large language models can generate human-like communications, while other AI systems can analyse enormous quantities of information, identify patterns and automate tasks previously requiring substantial human intervention. The same capabilities that make AI economically and socially valuable can also be exploited for criminal purposes. Cybercriminals can use AI to produce highly personalised phishing communications, impersonate individuals through synthetic audio and video, automate social-engineering operations, generate malicious code, identify potential vulnerabilities, create synthetic identities and manipulate digital information. AI therefore does not merely constitute another technology through which crime can be committed[2]. In certain circumstances, it can alter the speed, scale, personalisation and degree of automation of criminal conduct.

This transformation creates a fundamental legal problem. Traditional criminal law generally assumes that a human actor performs a prohibited act with the requisite mental state. AI-enabled cybercrime complicates this assumption because several actors may contribute to the eventual offence. A developer may create a general-purpose AI model; a company may deploy the model; an intermediary may provide access to it; a user may intentionally exploit it for criminal purposes; and the AI system itself may perform substantial operational steps. The central legal question is therefore not whether artificial intelligence can "commit" a crime in the philosophical sense. The more immediate legal question is:

When an AI system substantially facilitates or performs conduct resulting in a cyber-offence, upon whom should criminal responsibility be imposed?

The answer cannot simply be that the developer of the AI system is responsible. A general-purpose technology may have thousands or millions of legitimate uses. Imposing criminal liability merely because an AI system was capable of being misused would create an extremely

[1] Adib Bin Rashid and Md Ashfakul Karim Kaushik, *"AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications"* 7 HYBRID ADVANCES 1, 2 (2024).
[2] Marc Schmitt and Ivan Flechais, *"Digital Deception: Generative artificial intelligence in social engineering and phising"* 57 AI REVIEW 1, 10 (2024).

broad form of liability inconsistent with principles of legality, mens rea and culpability. Conversely, treating AI as merely a neutral tool may also become inadequate where developers, deployers or users intentionally design or configure systems for criminal purposes, deliberately ignore foreseeable risks, or exercise substantial control over an AI system that causes unlawful harm. The legal challenge is consequently one of attribution.

India's existing cybercrime framework was developed before the emergence of contemporary generative AI. The IT Act, 2000[3] remains the principal specialised legislation governing cyber offences. Its provisions address computer-related offences, identity theft, cheating by personation, privacy violations and cyber terrorism, among other matters. The statutory framework is technologically capable of applying to some AI-enabled conduct because the relevant offences focus on unlawful conduct involving computer resources or communication devices rather than necessarily prescribing a particular software technology. Nevertheless, AI-enabled cybercrime introduces circumstances that were not central to the original legislative design. The use of deepfakes, synthetic identities, autonomous agents and AI-generated malware raises questions concerning the relationship between the traditional concepts of actus reus, mens rea and technological agency.

The problem has international significance. The European Union has adopted a comprehensive risk-based approach to artificial intelligence through the EU AI Act[4]. The emerging international regulatory environment increasingly recognises that AI systems can create risks extending beyond privacy and discrimination into cybersecurity and broader societal harms. Contemporary scholarship also increasingly examines how AI regulation should conceptualise and quantify risk, demonstrating that the regulatory debate has moved beyond simple questions of technological innovation toward risk governance and accountability. The issue is equally relevant to India because the country is simultaneously expanding its digital economy, deploying AI systems and confronting increasingly sophisticated cybercrime. Recent Indian policy and technology developments demonstrate that AI is already being considered within cybersecurity and law-enforcement environments[5]. For example, Indian policing institutions are exploring AI-assisted crime detection, data analytics and cyber investigations.

[3] The Information Technology Act, 2000, India Code (2000).
[4] Gustavo Gil Gasiola, *"Rebuilding the pyramid: The AI Act's risk-based approach using a binary decision diagram"* 58 COMP. LAW & SECURITY REV. 1, 7 (2025)
[5] Tanvi Saxena, *AI Policing and Surveillance in India,* SRPF (June 1, 2026), AI Policing and Surveillance in India – SPRF

The legal system must therefore address a dual reality: AI can be used both to commit cybercrime and to prevent and investigate it. This paper argues that India does not necessarily require an entirely separate criminal code for AI-enabled offences. Instead, it requires an AI-sensitive interpretation and regulatory architecture capable of applying existing criminal-law principles while establishing clear rules for attribution, control, foreseeability, evidence, developer responsibility and international cooperation. The paper develops this argument through doctrinal and comparative legal research.

The central research problem arises from the mismatch between the architecture of conventional criminal liability and the increasingly distributed nature of AI-enabled cybercrime. A conventional cybercrime may involve:

Human offender → computer system → victim

whereas, an AI-enabled cybercrime may involve:

Developer → AI model → platform/deployer → user → AI-generated output → digital infrastructure → victim.

The chain becomes even more complicated where an AI agent independently selects targets, generates communications, adapts its strategy and executes multiple operations[6]. This creates at least six legal questions:

1. Who possesses the requisite mens rea?

2. Who controls the AI system?

3. Was the harmful outcome foreseeable?

4. What degree of human intervention is required?

5. Can existing cybercrime provisions adequately capture AI-mediated conduct?

6. How should courts authenticate AI-generated evidence?

[6] Ranjan Sapkota et al., *"AI agents vs. Agentic AI: A conceptual taxonomy, applications and challenges"* 126 INFO. FUSION 1, 28 (2026).

These questions demonstrate that AI-enabled cybercrime is not merely a technological issue. It is fundamentally a question of criminal-law attribution and regulatory design. This paper has five objectives:

1. To examine the nature and classification of AI-enabled cybercrime.

2. To analyse whether existing Indian criminal and cyber laws can adequately address AI-assisted and AI-enabled cyber offences.

3. To examine the attribution of criminal responsibility among AI users, developers, deployers and intermediaries.

4. To analyse the emerging evidentiary challenges associated with AI-generated and AI-manipulated digital evidence.

5. To develop a comparative and rights-compatible framework for regulating AI-enabled cybercrime in India.

II. CONCEPTUALISING AI-ENABLED CYBERCRIME

A significant conceptual difficulty arises because "AI-enabled cybercrime" can encompass conduct ranging from a human using AI merely to draft a phishing email to highly automated systems capable of performing multiple stages of an attack[7]. A useful taxonomy is therefore necessary.

A. Level I: AI-Assisted Cybercrime: At the first level, AI functions primarily as a tool. For example, a criminal may ask a generative AI system to: improve the grammar of a phishing email; translate a fraudulent message; create a fake website; generate social-engineering scripts; or analyse publicly available information about a victim. The human actor retains substantial control over the criminal conduct. The legal analysis in such cases is relatively straightforward. AI is analogous to other tools used by an offender. The user remains the principal actor and can ordinarily be assessed under existing criminal-law principles.

B. Level II: AI-Enhanced Cybercrime: At the second level, AI significantly increases the effectiveness of criminal conduct. Examples include: personalised phishing campaigns;

[7] Rosario Girasa, *Artifical Intelligence as a disruptive technology* (Palgrave Macmillan 2nd edn., 2025).

automated victim profiling; deepfake impersonation; voice cloning; automated vulnerability identification; and large-scale social engineering. The AI system may perform substantial analytical or generative functions, but the human actor retains strategic control.

C. Level III: AI-Automated Cybercrime: At the third level, AI performs substantial operational functions with limited continuous human intervention. An AI system may: identify potential victims; generate communications; modify messages according to responses; select successful strategies; initiate transactions; and report results to the human operator. At this stage, the question of human control becomes significantly more complicated.

D. Level IV: Semi-Autonomous or Autonomous Cybercrime: The fourth level involves AI systems capable of making operational decisions with minimal or no real-time human direction. The system may identify a vulnerability, determine an exploitation strategy and execute an attack. This represents the most difficult category for conventional criminal law because the temporal connection between human intention and individual criminal acts becomes attenuated. However, the existence of autonomy should not automatically eliminate human responsibility. The legal inquiry should instead examine:

- Who designed the system?
- Who deployed it?
- Who controlled it?
- Who benefited from it?
- What risks were foreseeable?
- What safeguards were deliberately omitted?

These questions provide a more useful foundation for legal attribution than the abstract question of whether an AI system itself possesses criminal responsibility.

III. NOVEL AI-ENABLED CYBER-OFFENCES

A. PHISHING AND SOCIAL ENGINEERING

Phishing has traditionally relied upon deceptive communications designed to induce victims to

disclose credentials, transfer money or reveal confidential information. Generative AI substantially enhances this technique[8]. AI can produce communications that: contain grammatically sophisticated language; imitate the communication style of particular organisations; operate across multiple languages; adapt messages to individual victims; generate large numbers of personalised communications; and respond dynamically to victims.

The legal significance is that AI reduces the cost of producing convincing deception. A criminal organisation that previously required individuals with language and social-engineering skills may increasingly automate those functions.

The underlying offence may nevertheless remain conventional cheating, fraud, personation or identity theft. The novel issue is therefore not necessarily the creation of a new offence but the scale and automation of existing criminal conduct. This distinction supports the argument that Indian law should avoid creating an unnecessarily broad offence called "AI misuse." Instead, AI should operate as an aggravating or attributional factor where it materially increases the sophistication, scale or harm of an existing offence.

B. DEEPFAKES AND SYNTHETIC IDENTITY CRIME

Deepfakes represent one of the clearest examples of AI-enabled cybercrime. Synthetic audio and video can be used to impersonate: family members; business executives; government officials; police officers; judges; celebrities; and financial professionals[9]. The resulting conduct may constitute cheating, personation, identity theft, extortion, defamation, privacy violation or other offences depending on the facts. The Indian IT Act already criminalises cheating by personation through a computer resource or communication device[10]. It also contains provisions addressing identity theft and privacy violations. The difficulty is therefore not necessarily the absence of applicable offences. The deeper issue is whether existing provisions can accommodate synthetic identity.

Traditional identity theft assumes that an offender misappropriates an existing identifier—such as a password or electronic signature. Synthetic identity fraud can operate differently. An

[8] Daniela Popescul and Laura Diana Radu, *"AI in phishing detection: a bibliometric review"* FRONTEIRS 1, 15 (2025).
[9] Bart van der Sloot and Yvette Wagensveld, *"Deepfakes: regulatory challenges for the synthetic society"* 46 COMP. LAW & SECURITY REV. 1, 18 (2022).
[10] The Information Technology Act, 2000, India Code (2000), § 66D.

offender may construct a fictional identity by combining authentic and fabricated information. Similarly, a deepfake may imitate a person without accessing that person's actual account or credentials. This creates an important doctrinal distinction: identity theft is not necessarily equivalent to identity simulation. [11]Indian legislation should therefore be interpreted in a manner capable of addressing fraudulent synthetic representation without artificially limiting offences to stolen credentials.

C. AI-GENERATED MALWARE AND AUTOMATED CYBER ATTACKS

AI may also assist offenders in generating malicious code or adapting existing malware. The legal consequences may involve: unauthorised access; computer-related offences; data interference; system interference; cyber extortion; financial crime; and potentially national-security offences[12]. The IT Act already contains technology-neutral provisions concerning unauthorised acts involving computer resources. The use of AI does not necessarily transform the underlying conduct into a new offence. However, AI may affect the analysis of mens rea and participation. Suppose an individual deliberately instructs an AI system to create malicious code and deploy it against a particular target. The human actor has clearly exercised substantial control. By contrast, suppose an organisation deploys an AI security agent capable of automatically interacting with external systems and the system unexpectedly causes damage to a third-party computer. The question becomes more difficult. Criminal liability should not automatically follow from mere deployment of an AI system. The law should examine whether the organisation:

- knew of the relevant risk;
- had the ability to control the system;
- failed to implement reasonable safeguards;
- deliberately permitted the relevant capability; or
- intended or consciously disregarded the resulting harm.

[11] Huei-Wen Teng et al., *"Digital assets: risk, regulations, mitigation"* 12 FINANCIAL INNOVATION 1, 25 (2026).
[12] Nadia Khadam et.al, *How to punish cyber criminals: A study to investigate the target and consequence based punishments for malware attacks in UK, USA, China, Ethiopia & Pakistan"* 9 HELIYON 1, 22 (2023).

This is where conventional concepts of recklessness, knowledge, negligence and causation become important.

IV. THE PROBLEM OF CRIMINAL ATTRIBUTION

Mens Rea in AI-Enabled Cybercrime: Mens rea represents one of the most important doctrinal challenges. Traditional criminal offences may require: intention; knowledge; dishonesty; fraud; recklessness; or another specified mental element[13]. An AI system itself does not necessarily possess legally recognised intention or knowledge. Consequently, the law must identify the relevant human mental state. For an AI-enabled offence, the inquiry should consider:

1. Who initiated the AI operation?

2. What did that person know?

3. What outcome did the person intend?

4. What level of control did the person exercise?

5. What risks were reasonably foreseeable?

6. Did the person deliberately disregard known safeguards?

This approach avoids the conceptual error of treating AI's computational processes as equivalent to human mens rea. AI may generate an output, but criminal law must ordinarily identify a legally responsible person or entity whose conduct satisfies the requirements of the offence. Hence, criminal law generally rests upon the principle that punishment should correspond to culpable conduct[14] and AI complicates attribution because the causal chain may involve multiple actors[15]. Consider four hypothetical scenarios.

[13] James Edward and Tarek Yusai, *"Internal morality of criminal law"* 43 OXF J. LEG. STUD. 475, 476 (2023).
[14] Antony Nicolas Allot, General Principles of Criminal Law July 22 2026 Crime - Punishment, Liability, Offenses | Britannica
[15] Mark Coeckelbergh, *"Artifical Intelligence, Responsibility Attribution, and a Relational Justification of Explainability"* 26 SCI ENG ETHICS. 2051, 2065 (2019).

Scenario 1: Direct criminal use

A person intentionally uses an AI tool to create a deepfake video and defraud a victim.

Primary responsibility: user.

Scenario 2: Criminal deployment

A company deliberately configures an AI agent to identify and exploit vulnerable computer systems.

Primary responsibility: deployer/operator.

Scenario 3: Reckless deployment

A company deploys an autonomous system despite knowing that it has repeatedly generated harmful outputs and fails to implement available safeguards.

Potential responsibility: organisation and responsible decision-makers, subject to the applicable statutory offence and mental-state requirements.

Scenario 4: Unforeseeable misuse

A general-purpose AI system is lawfully developed and a third party independently uses it for a sophisticated cyber-attack in a manner that was not reasonably foreseeable and over which the developer exercised no relevant control.

Criminal responsibility of developer: substantially more difficult to establish.

This classification demonstrates why a blanket rule imposing liability upon developers would be inappropriate.

V. LIABILITY OF VARIOUS ACTORS

A. DEVLOPER LIABILITY

The possibility of holding AI developers criminally responsible deserves careful consideration.

A developer creates a general-purpose AI system that can be used for both lawful and unlawful purposes[16]. The mere possibility of misuse should not generate criminal responsibility. Otherwise, developers of: programming languages; encryption technologies; operating systems; browsers; cloud services; and communication platforms could potentially face criminal liability whenever criminals misuse their products. Such an approach would be incompatible with technological innovation and principles of culpability. Developer liability should therefore require additional circumstances, such as:

- deliberate design for criminal purposes;
- intentional facilitation of a specific offence;
- substantial knowledge of a criminal deployment;
- deliberate circumvention of legal safeguards; or
- another legally recognised basis of participation or culpability.

The distinction between general-purpose development and purposeful criminal facilitation should therefore form an essential part of future AI-cybercrime law.

B. DEPLOYER AND USER LIABILITY

The strongest basis for criminal attribution ordinarily exists where the individual or organisation controls the deployment of AI for a criminal purpose. The deployer may:

- select the target;
- provide the instructions;
- configure the system;
- determine the operational parameters;
- receive the benefits; and

[16] Athina Sachoulidou, *"AI systems and criminal liability"* 11 OSLO LAW REV. 1, 9 (2024).

- intentionally initiate the criminal process.

The law should therefore focus substantially upon control and purpose. This suggests a principle:

The greater the person's control over the AI system and the greater the person's knowledge of its intended criminal use, the stronger the basis for criminal attribution. Conversely, where control and knowledge diminish, liability should require additional proof.

C. CORPORATE LIABILITY AND AI-ENABLED CYBERCRIME

AI-enabled cybercrime may also occur within corporate environments. An organisation may deploy AI systems for-cybersecurity; fraud detection; automated customer communications; financial transactions; network management; or decision-making[17]. If such a system causes unlawful harm, the legal question becomes whether responsibility should attach to-the organisation; senior management; the system operator; the developer; or no individual. Corporate criminal liability should not become automatic merely because an AI system operated within an organisation. Instead, courts should examine:

1. organisational knowledge & control;

2. policies & risk assessments;

3. safeguards & supervision;

4. compliance mechanisms; and

5. causal connection between organisational conduct and the offence.

This approach would align AI governance with established principles of corporate responsibility while recognising the distinctive characteristics of automated systems.

VI. AI AND DIGITAL EVIDENCE

AI does not only create new forms of cybercrime. It also challenges the reliability of evidence

[17] Ramanpreet Kaur, *"Artifical Intelligence for cybersecurity: Literature review and future research directions"* 97 INFO. FUSION 1, 26 (2023).

used in criminal proceedings. This is particularly evident with: deepfake videos; cloned voices; synthetic photographs; AI-generated documents; manipulated metadata; generated chat conversations; and altered digital records. The BSA, 2023 provides a statutory framework for electronic and digital records[18]. The challenge is that formal admissibility does not necessarily establish authenticity. A court may receive an electronic record and still need to determine:

- Was it generated by the person alleged?
- Has it been altered?
- Was the underlying recording authentic?
- Was the AI system itself manipulated?
- Can the chain of custody be established?

AI therefore introduces a new dimension to electronic evidence:

Traditional question:

Is the digital record authentic?

Emerging question:

Is the digital record itself synthetic or AI-manipulated?

This distinction will become increasingly important.

DEEPFAKES AND THE BURDEN OF PROOF: They create a particularly serious evidentiary problem because visual and audio evidence has traditionally been psychologically persuasive to human observers. A video may appear authentic while being entirely synthetic. Consequently, courts should increasingly rely upon: metadata analysis; source verification; forensic examination; cryptographic provenance; device records; server logs; corroborating communications; witness testimony; and expert evidence[19]. The development of AI-generated

[18] The Bharatiya Sakshya Adhiniyam, 2023, India Code (2023).
[19] Hitesh Bhatt, *"Industry 4.0 crime scene investigation: A review and zero-trust digital forensic architecture for evidence integrity"* 14 FOREIGN SC. INT. REPORTS 1, 18 (2026).

evidence also strengthens the case for standardised digital forensic procedures. India's evidentiary framework should therefore evolve beyond the simple distinction between "electronic" and "non-electronic" evidence. The relevant distinction increasingly becomes: *authentic digital evidence versus synthetic or manipulated digital evidence.*

VII. COMPARTIVE STUDY AND COMPARATIVE LESSONS

A. THE EUROPEAN UNION APPROACH

The European Union provides an important comparative model because its AI regulatory framework adopts a risk-based approach rather than treating all AI systems identically. The EU AI Act establishes different regulatory obligations according to the level and nature of risk associated with AI systems. The importance of this approach for cybercrime regulation lies in the principle of graduated responsibility. A general-purpose low-risk system should not necessarily be regulated in the same manner as an AI system capable of causing substantial harm. This principle can be adapted to cybercrime. AI systems with significant capabilities for: autonomous cyber operations; vulnerability discovery; identity manipulation; large-scale social engineering; or automated exploitation, could attract enhanced governance requirements.

The EU's regulatory experience also demonstrates the increasing importance of risk assessment, accountability and post-deployment monitoring. Current scholarship examining the EU AI Act's risk-based model highlights the difficulty of defining, quantifying and governing AI risk while balancing fundamental rights against legitimate technological uses. For India, the important lesson is not to replicate the EU AI Act mechanically[20]. Instead, India could adopt a risk-sensitive cybercrime framework that distinguishes AI systems according to their capacity to facilitate criminal conduct.

B. UNITED KINGDOM AND UNITED STATES

The United Kingdom has generally approached AI governance through a combination of existing legal frameworks, sector-specific regulation and institutional guidance rather than immediately creating a single comprehensive AI criminal code. This approach provides an important lesson for India: emerging technologies do not always require entirely new criminal

[20] Sakti Prasad and Srichandan, *"What India can learn from EU's AI reset"* Express Opinion (Aug. 24, 2026)

offences. Where existing offences adequately describe the prohibited conduct, the law can apply existing principles.

The United States provides another instructive example because its legal framework is distributed across federal criminal statutes, sectoral legislation, state law and regulatory enforcement. The American experience demonstrates both the flexibility and fragmentation that can accompany technology regulation. India can therefore draw a middle path: retain technology-neutral criminal offences while establishing AI-specific governance obligations for high-risk systems.

C. INDIAN LEGAL FRAMEWORK AND ITS LIMITATIONS

India's existing framework provides several legal tools capable of addressing AI-enabled cybercrime. The IT Act contains provisions concerning identity theft, cheating by personation, privacy violations and computer-related offences. The BNS provides the general criminal-law framework for offences such as cheating, forgery, criminal intimidation and other conduct that may be facilitated through AI. The BNSS provides the procedural architecture for investigation and trial[21]. The BSA provides the evidentiary framework for electronic and digital records. Consequently, the argument that India has no legal framework for AI-enabled cybercrime would be inaccurate. The more precise conclusion is that India lacks a specific integrated attribution framework for AI-mediated criminal conduct. The current law can generally punish the human offender who intentionally uses AI as a tool. Greater uncertainty emerges where:

- AI performs substantial autonomous functions;
- multiple actors contribute to the system;
- harm is caused through unexpected AI behaviour;
- developers possess varying degrees of knowledge;
- organisations fail to supervise AI systems; or
- synthetic evidence is presented before courts.

[21] The Bharatiya Nagarik Suraksha Sanhita, 2023, India Code (2023).

D. INTERNATIONAL CYBERCRIME FRAMEWORK

AI-enabled cybercrime also reinforces the importance of international cybercrime cooperation.

An AI-generated phishing operation may involve: an operator in India; an AI service hosted elsewhere; cloud infrastructure in another jurisdiction; victims in multiple countries; and financial transfers through international payment systems. The Budapest Convention provides a useful international framework because it combines substantive cybercrime provisions, procedural mechanisms concerning electronic evidence and international cooperation[22]. Its relevance becomes even greater in an AI environment because digital evidence may be distributed across several jurisdictions. International cooperation should therefore address:

- rapid preservation of data;
- cross-border access to electronic evidence;
- identification of AI service users;
- attribution of infrastructure;
- cryptocurrency tracing;
- extradition; and
- coordinated investigation.

VIII. PROPOSED FRAMEWORK AND REFORMS FOR INDIA

I. AI-CYBERCRIME LIABILITY FRAMEWORK

This paper proposes a five-factor AI-Cybercrime Liability Framework**.**

Factor 1: Human Control

The first question should be: Who controlled the AI system at the time of the relevant conduct?

[22] Filippo Spiezia, *"International cooperation and protection of victims in cyberspace: welcoming Protocol II to the Budapest Convention on Cybercrime"* 23 ERA FORUM 101, 106 (2022).

Greater control should strengthen the basis for attribution.

Factor 2: Criminal Purpose

The second question should be: Was the system intentionally deployed for criminal purposes? Intentional deployment should substantially strengthen criminal responsibility.

Factor 3: Knowledge and Foreseeability

The third question should be: Did the relevant actor know, or under the applicable offence ought the relevant risk to have been recognised, that the system was likely to produce unlawful consequences? This factor is particularly important for corporate and developer responsibility.

Factor 4: Technological Contribution

The fourth question should be: How substantially did AI contribute to the commission of the offence? The law should distinguish between:

- AI merely generating text;
- AI selecting victims;
- AI designing the attack;
- AI executing the attack; and
- AI autonomously adapting the attack.

Factor 5: Causal Connection

Finally: Can the defendant's conduct be sufficiently connected to the resulting criminal harm? The existence of an AI system should not break causation automatically.

The five factors can be represented as: Control + Purpose + Knowledge/Foreseeability + AI Contribution + Causation = Basis for Attribution

This framework avoids two extremes.

Extreme 1: AI exceptionalism- Creating a completely separate criminal regime for every AI-enabled offence.

Extreme 2: AI neutrality- Treating AI as legally irrelevant regardless of the degree of autonomy or risk.

The proposed framework instead incorporates AI into conventional principles of criminal responsibility while recognising the technological characteristics that make attribution more difficult.

II. REGULATORY REFORMS

A. AI-Specific Cybercrime Guidelines

India should develop specialised guidelines addressing the investigation and prosecution of AI-enabled cybercrime. These guidelines should address: AI-generated fraud; deepfakes; synthetic identities; AI-generated malware; autonomous cyber-attacks; AI evidence; and attribution.

B. Mandatory Risk Assessment for High-Risk AI Systems

Developers and deployers of AI systems possessing substantial cyber capabilities should undertake documented risk assessments. Such assessments should consider: misuse potential; access controls; logging; monitoring; incident response; human oversight; and evidence preservation.

C. Audit Trails

High-risk AI systems should maintain appropriate logs capable of establishing:

- who accessed the system;
- when it was accessed;
- what instructions were provided;
- what outputs were generated; and
- what actions were subsequently taken.

Such records could become critical digital evidence.

D. Provenance of Synthetic Media

India should encourage technological standards for identifying AI-generated content. This may involve: cryptographic provenance; watermarking; content credentials; metadata; platform-level disclosure; and forensic verification.

E. Specialised AI-Cybercrime Forensics

Cyber forensic laboratories should develop dedicated capabilities for: deepfake detection; AI-generated text analysis; synthetic audio detection; model-output attribution; prompt and log analysis; and AI-assisted malware investigation.

F. Judicial Training

Judges and prosecutors should receive specialised training concerning: AI systems; synthetic media; probabilistic outputs; deepfake detection; algorithmic evidence; and AI forensic methodologies.

Along with these reforms, the legislation should not overlook these important considerations:

A. CONSTITUTIONAL AND HUMAN-RIGHTS

Effective regulation of AI-enabled cybercrime must remain consistent with constitutional rights. India's constitutional framework places particular importance upon: equality; freedom of speech and expression; privacy; personal liberty; and due process. The Supreme Court's recognition of privacy as a constitutionally protected right in *Justice K.S. Puttaswamy case*[23] is particularly relevant to AI systems involving personal data, profiling and surveillance. Similarly, *Shreya Singhal case*[24] demonstrates that cyber regulation cannot be separated from freedom of speech and expression. AI regulation must therefore avoid creating excessively broad powers that could be used to suppress legitimate expression. For example, regulation concerning deepfakes should distinguish between: malicious impersonation; fraud; non-consensual sexual content; legitimate parody; satire; artistic expression; and political

[23] *Justice K.S. Puttaswamy (Retd.) v. Union of India*, AIR 2017 SC 4161
[24] *Shreya Singhal v. Union of India,* AIR 2015 SC 1523

commentary. The legal objective should be to regulate harmful conduct, rather than technology or expression as such.

B. THE BALANCE BETWEEN INNOVATION AND SECURITY

A major policy risk is overregulation. AI systems have legitimate cybersecurity applications.

They can assist with: threat detection; malware analysis; vulnerability identification; fraud prevention; incident response; digital forensics; and law-enforcement intelligence.

Indeed, Indian policing institutions are already exploring AI and data analytics for crime investigation and cybercrime response. Consequently, a regulatory framework that treats AI primarily as a threat would be counterproductive. The objective should instead be secure innovation. The same technology may be: *a cyber-weapon in one context and a cybersecurity tool in another.* Legal regulation should therefore focus upon use, intent, control and risk rather than technological capability alone.

XXVII. CONCLUSION

Artificial intelligence is transforming cybercrime by reducing the cost of sophisticated attacks, increasing the speed and scale of criminal operations and enabling new forms of deception. Generative AI, deepfakes, synthetic identities, automated social engineering and AI-assisted malware demonstrate that cybercrime is entering a period in which technological capability may increasingly be separated from continuous human intervention. India already possesses several legal provisions capable of addressing many AI-enabled offences. The IT Act, 2000 criminalises conduct including identity theft and cheating by personation through computer resources, while the general criminal-law framework can apply to fraud, forgery, intimidation and other offences committed through digital means. The central weakness is therefore not simply the absence of an AI-specific offence. The more fundamental problem concerns attribution.

When an AI system contributes substantially to a criminal offence, criminal responsibility should not automatically shift to the developer merely because the developer created a system capable of misuse. Nor should the existence of AI autonomy automatically eliminate human responsibility. The appropriate approach is to examine control, criminal purpose, knowledge or foreseeability, technological contribution and causation. The paper therefore proposes an

AI-Cybercrime Liability Framework based upon these five principles. The framework recognises four stages of AI-enabled offending: AI-assisted, AI-enhanced, AI-automated and AI-autonomous or semi-autonomous cybercrime. This classification allows the legal system to distinguish a human offender who simply uses AI as a tool from an organisation that deliberately deploys an autonomous system for criminal purposes.

Comparative analysis demonstrates that India can also learn from risk-based approaches to AI governance. The European Union's regulatory model demonstrates the value of graduated obligations according to risk, while the experience of other jurisdictions illustrates the continuing importance of technology-neutral criminal law. India should consequently resist the temptation to create a separate criminal offence for every new AI capability. Such an approach would rapidly become obsolete. Instead, India should develop an adaptive regulatory architecture comprising: *risk assessment; human oversight; auditability; AI-system logging; digital provenance; specialised forensic capacity; judicial training; platform responsibility; and international cooperation.*

At the evidentiary level, the emergence of deepfakes and synthetic media requires courts to move beyond the traditional question of whether an electronic record exists. The increasingly important question will be whether the record is authentic, manipulated or entirely synthetic.

At the international level, AI-enabled cybercrime reinforces the necessity of rapid cross-border cooperation because AI systems, cloud infrastructure, offenders, victims and digital evidence may exist in different jurisdictions. Ultimately, AI should not be conceptualised exclusively as a threat to cybersecurity. It is simultaneously a source of cyber risk and a potential instrument for cyber defence. A legally sustainable framework must therefore balance innovation with accountability. The appropriate regulatory principle is consequently not *"regulate AI because it is dangerous",* but rather: *regulate the use, control, risk and consequences of AI in proportion to the harm it creates.* For Indian cyber law, this represents the more durable path toward addressing the next generation of cybercrime while preserving technological innovation, constitutional rights and the fundamental principles of criminal responsibility.

REFERENCES

1) Adib Bin Rashid and Md Ashfakul Karim Kaushik, "AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications" 7 HYBRID ADVANCES 1, 2 (2024).

2) Antony Nicolas Allot, General Principles of Criminal Law July 22 2026 Crime - Punishment, Liability, Offenses | Britannica

3) Athina Sachoulidou, *"AI systems and criminal liability"* 11 OSLO LAW REV. 1, 9 (2024).

4) Bart van der Sloot and Yvette Wagensveld, *"Deepfakes: regulatory challenges for the synthetic society"* 46 COMP. LAW & SECURITY REV. 1, 18 (2022).

5) Daniela Popescul and Laura Diana Radu, *"AI in phishing detection: a bibliometric review"* FRONTEIRS 1, 15 (2025).

6) Filippo Spiezia, *"International cooperation and protection of victims in cyberspace: welcoming Protocol II to the Budapest Convention on Cybercrime"* 23 ERA FORUM 101, 106 (2022).

7) Gustavo Gil Gasiola, *"Rebuilding the pyramid: The AI Act's risk-based approach using a binary decision diagram"* 58 COMP. LAW & SECURITY REV. 1, 7 (2025)

8) Hitesh Bhatt, *"Industry 4.0 crime scene investigation: A review and zero-trust digital forensic architecture for evidence integrity"* 14 FOREIGN SC. INT. REPORTS 1, 18 (2026).

9) Huei-Wen Teng et al., *"Digital assets: risk, regulations, mitigation"* 12 FINANCIAL INNOVATION 1, 25 (2026).

10) James Edward and Tarek Yusai, *"Internal morality of criminal law"* 43 OXF J. LEG. STUD. 475, 476 (2023).

11) *Justice K.S. Puttaswamy (Retd.) v. Union of India,* AIR 2017 SC 4161

12) Marc Schmitt and Ivan Flechais, *"Digital Deception: Generative artificial intelligence in social engineering and phising"* 57 AI REVIEW 1, 10 (2024).

13) Mark Coeckelbergh, *"Artifical Intelligence, Responsibility Attribution, and a Relational Justification of Explainability"* 26 SCI ENG ETHICS. 2051, 2065 (2019).

14) Nadia Khadam et.al, *How to punish cyber criminals: A study to investigate the target and consequence based punishments for malware attacks in UK, USA, China, Ethiopia & Pakistan"* 9 HELIYON 1, 22 (2023).

15) Ramanpreet Kaur, *"Artifical Intelligence for cybersecurity: Literature review and future research directions"* 97 INFO. FUSION 1, 26 (2023).

16) Ranjan Sapkota et al., *"AI agents vs. Agentic AI: A conceptual taxonomy, applications and challenges"* 126 INFO. FUSION 1, 28 (2026).

17) Rosario Girasa, *Artifical Intelligence as a disruptive technology* (Palgrave Macmillan 2nd edn., 2025).

18) Sakti Prasad and Srichandan, *"What India can learn from EU's AI reset"* Express Opinion (Aug. 24, 2026)

19) *Shreya Singhal v. Union of India,* AIR 2015 SC 1523

20) Tanvi Saxena, *AI Policing and Surveillance in India,* SRPF (June 1, 2026), AI Policing and Surveillance in India – SPRF

21) The Bharatiya Nagarik Suraksha Sanhita, 2023, India Code (2023).

22) The Bharatiya Sakshya Adhiniyam, 2023, India Code (2023).

23) The Information Technology Act, 2000, India Code (2000), § 66D.

24) The Information Technology Act, 2000, India Code (2000).