FORENSIC ATTRIBUTION CHALLENGES IN AI-DRIVEN CYBERCRIME: IMPLICATIONS FOR CRIMINAL LIABILITY AND DIGITAL EVIDENCE UNDER INDIAN LAW

Yogalakshmi G, LL.M. (Hons), School of Excellence in Law,
The Tamilnadu Dr. Ambedkar Law University, Chennai

ABSTRACT

The adoption of artificial intelligence (AI) into cyberspace has fundamentally transformed the nature, scale, and sophistication of cybercrime. AI-driven cyber offences, ranging from deepfake fraud to autonomous malware pose unprecedented challenges to traditional legal doctrines, particularly in the areas of forensic attribution, criminal liability, and evidentiary standards. This paper critically evaluates the difficulty of identifying perpetrators in AI-mediated crimes and evaluates how Indian legal frameworks, including the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023, and the Bharatiya Sakshya Adhiniyam, 2023, respond to these challenges. It also analyses the limitations of existing laws in dealing with autonomous decision-making systems and proposes reforms based on comparative international approaches such as the EU AI Act and emerging regulatory models in the United States and China. The paper argues that without doctrinal evolution in attribution and evidentiary standards, the Indian legal system risks under-enforcement and injustice in AI-driven cybercrime cases.

Keywords: Artificial Intelligence, Forensic Attribution, AI-Driven Cyber Crime, Criminal Liability.

1. Introduction

The rapid advancement of artificial intelligence has redefined the cyber threat landscape. Unlike traditional cybercrimes, AI-driven cyber offences are characterized by automation, scalability, anonymity, and adaptive learning. These features complicate the identification of offenders and disrupt foundational principles of criminal law—particularly *actus reus* (guilty act) and *mens rea* (guilty mind).

In conventional cybercrime, attribution involves tracing digital footprints such as IP addresses, device identifiers, or user credentials. However, AI systems, especially those using machine learning algorithms can operate autonomously, generate unpredictable outputs, and obscure human involvement. This creates a **"black box problem"** in forensic analysis, where the causal link between human intent and criminal outcome becomes difficult to establish.

In India, cybercrime is primarily governed by the Information Technology Act, 2000, further supplemented by legal framework under the Bharatiya Nyaya Sanhita, 2023. However, these laws were drafted in a pre-AI era and fail to adequately address the complexities of AI-enabled offences. This paper explores the intersection of forensic attribution, criminal liability, and digital evidence in AI-driven cybercrime, focusing on Indian law while incorporating international perspectives.

2. Nature of AI-Driven Cybercrime

AI-driven cybercrime refers to offences where artificial intelligence is used as a tool, medium, or autonomous agent in the commission of crimes. These include:

Deepfake-based fraud and impersonation

Deepfake-based fraud involves using AI-generated audio, video, or images to convincingly mimic real individuals. Criminals may impersonate CEOs, government officials, or relatives to mislead victims into transferring money or sharing sensitive data. This form of cybercrime is particularly dangerous because it exploits trust in visual and auditory authenticity.

AI-powered phishing and social engineering attacks

AI-powered phishing uses machine learning to craft highly personalized and convincing scam

messages. Unlike traditional phishing, these attacks analyse user behaviour, preferences, and online activity to increase success rates. Consequently, victims are more likely to click malicious links or disclose confidential information without suspicion.

Autonomous malware and self-learning viruses

Autonomous malware refer to malicious software that can adapt and evolve without human intervention. These programs use AI techniques to detect security defences and modify their behaviour to avoid detection. Such self-learning viruses present a significant challenge to cybersecurity systems due to their unpredictability.

Algorithmic financial manipulation

Algorithmic financial manipulation involves the use of AI systems to distort markets or exploit trading systems. Criminal actors may deploy bots to create fake demand, trigger price fluctuations, or execute high-speed fraudulent trades. This undermines market integrity and can cause significant financial losses to investors and institutions.

Synthetic identity theft

Synthetic identity theft occurs when AI is designed to generate entirely new but realistic fake identities. These identities may combine real personal data with fabricated details to bypass verification systems. These are typically used to open fraudulent accounts, secure loans, or commit long-term financial fraud. Such crimes differ from traditional cyber offences in three key ways:

(a) Automation and Scale

AI systems can execute thousands of attacks simultaneously without direct human intervention.

(b) Adaptability

Machine learning models evolve based on data inputs, making them capable of bypassing security systems dynamically.

(c) Obfuscation of Responsibility

The involvement of multiple actors—developers, deployers, users, and platforms, creates

ambiguity in attributing liability. The IT Act criminalizes acts such as hacking (Section 66), identity theft (Section 66C), and impersonation (Section 66D), but it does not explicitly address AI-based offences.

3. Forensic Attribution in Cybercrime

3.1 Traditional Attribution Methods

Digital forensic attribution typically relies on:

IP address tracing

IP address tracing involves identifying the numerical label assigned to a device connected to a network. Investigators use this to track the approximate geographic location and internet service provider of a user. However, it can be unreliable due to Virtual Private Networks (VPNs), proxies, and anonymization tools that mask the real source.

Device fingerprinting

Device fingerprinting collects unique device and browser characteristics such as Operating System (OS), hardware, and settings. These combined traits allow identification of individual devices even without user login information. It is useful in linking repeated cyber activities to the same device across different sessions.

Log analysis

Log analysis involves reviewing system, application, and server logs that record digital activities. These logs help reconstruct timelines of events, including access attempts and system changes. They are necessary for detecting suspicious behaviour and tracing unauthorized actions.

Metadata examination

Metadata examination focuses on hidden data embedded within digital files such as creation date and authorship. It can reveal how, when, and by whom a file was created or modified. This helps in verifying authenticity and tracking file manipulation.

Network traffic analysis

Network traffic analysis studies data moving across networks to detect unusual communication patterns. It helps identify connections between systems involved in potential cybercrime activities. This method is widely used to detect data leaks, malware communication, and intrusion attempts.

These methods assume a direct link between the offender and the digital act.

3.2 Attribution Challenges in AI Context

AI disrupts these assumptions in several ways:

(i) Autonomous Decision-Making

AI systems can act independently, making it challenging to ascertain whether the output reflects human intent.

(ii) Layered Responsibility

Multiple actors may contribute to the system – Programmer, Data provider, End-user or Platform intermediary

(iii) Data Manipulation and Synthetic Content

AI-generated content (e.g., deepfakes) can fabricate evidence, misleading forensic investigations.

(iv) Cross-Border Operations

AI systems often operate across jurisdictions, complicating enforcement despite Section 75 of the IT Act providing extraterritorial jurisdiction.

(v) Explainability Problem

Machine learning models lack transparency, making it difficult for forensic experts to reconstruct decision pathways.

4. Criminal Liability in AI-Driven Cybercrime

4.1 Traditional Principles of Liability

Indian criminal law is based on Mens rea (intent), Actus reus (act) and Causation. However, AI systems challenge these principles because they lack legal personality and intent.

4.2 Applicable Indian Legal Framework

(a) Information Technology Act, 2000

The Indian legal framework addressing cyber offences is primarily governed by the Information Technology Act, 2000[1]. This Act contains critical provisions dealing with digital crimes, including Section 43 which penalizes unauthorized access to computer systems and damage to data, and Section 66 which covers computer-related offences involving dishonest or fraudulent intent. Further, Section 66C specifically addresses identity theft involving misuse of another person's electronic identity, while Section 66D deals with cheating by impersonation using computer resources. Together, these provisions form the regulatory framework for governing and penalizing cybercrime in India.

(b) Bharatiya Nyaya Sanhita, 2023

Addresses offences such as cheating, fraud, and criminal conspiracy[2].

(c) Bharatiya Sakshya Adhiniyam, 2023 (BSA)

Electronic evidence is primarily governed by **Sections 61, 62, and 63**. Section 61 recognises electronic records as admissible documents with the same legal effect as physical records. Section 62 defines conditions for their admissibility, focusing on authenticity and integrity. Section 63 requires a certificate to ensure the reliability and proper source of electronic evidence.

(d) The Digital Personal Data Protection Act, 2023

Although the Digital Personal Data Protection Act, 2023 (DPDPA) is primarily concerned with

[1] Information Technology Act, 2000 – Sections 43, 66, 66C, 66D
[2] Bharatiya Nyaya Sanhita, 2023 – provisions on cheating, fraud, conspiracy

personal data protection, its provisions may have consequential effects on AI-driven cybercrime regulation. The Act establishes duties for data fiduciaries to protect personal data from breaches, and its enforcement framework creates civil liability for data processors whose AI systems contribute to data theft or misuse. However, the DPDPA does not address criminal liability directly and its relationship with the IT Act and BNS frameworks requires further judicial elucidation[3].

4.3 Attribution of Liability

In AI-driven crimes, liability may extend to multiple actors depending on their role in the offence. Developers may be held responsible for negligent or biased system design, while deployers can be liable for the misuse or unsafe implementation of AI tools. End users are directly accountable when AI is intentionally used for criminal purposes, and intermediaries may incur liability for failing to exercise due diligence or maintain platform security. Additionally, concerns including discrepancies between test data and real-world performance, and also unintended learning from real-life data inputs, further complicate attribution. Nevertheless, the lack of well-defined statutory guidance in India on these AI-specific challenges creates significant legal ambiguity in fixing responsibility.

4.4 Doctrinal Approaches

Doctrinal approaches to liability in AI-driven cybercrime include several evolving principles used to address gaps in existing law. Vicarious liability may render organizations accountable for actions carried out through AI systems deployed under their control. Negligence-based liability focuses on the failure to implement adequate safeguards, security measures, or oversight mechanisms while using AI technologies. Strict liability, on the other hand, may impose responsibility regardless of intent where harm results from the use of AI systems. Additionally, the doctrine of "innocent agency" treats AI as a mere tool controlled by a human offender, thereby attributing criminal responsibility to the individual directing its use.

5. Digital Evidence and AI Challenges

5.1 Legal Framework

[3] Digital Personal Data Protection Act, 2023

Under Sections 65A and 65B of the Indian Evidence Act, 1872[4], now substantially incorporated under Sections 61 to 63 of the BSA[5], electronic records are recognised as admissible evidence, subject to compliance with prescribed conditions relating to authenticity and certification. Section 61 of the BSA provides legal validity to electronic records, whereas Sections 62 and 63 specify procedural safeguards relating to genuineness and certification requirements.

5.2 Key Judicial Decisions

In *Anvar P.V. v. P.K. Basheer* (2014)[6], the Supreme Court clarified that the procedural requirements under Section 65B of the Evidence Act must be fulfilled for electronic records to be treated as valid evidence, particularly secondary electronic evidence. This principle has now been aligned with the BSA framework requiring proper certification under Sections 62–63. Subsequently, in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal* (2020)[7], the Court once again emphasized that the Section 65B certificate is a necessary condition for admissibility and cannot be dispensed with except in limited circumstances, a position now reflected under the updated BSA provisions governing electronic evidence.

5.3 Challenges in AI Context

The deployment of artificial intelligence in cybercrime investigations has created significant challenges relating to digital evidence and forensic reliability. One major issue is the authenticity of evidence, as AI-generated content such as deepfakes can generate realistic but fabricated material. This can mislead investigators and courts. The chain of custody also becomes difficult to establish when autonomous or self-learning systems are involved. Tracing the exact origin and handling of digital evidence may be complex. Questions regarding the reliability of algorithmic outputs further complicate judicial assessment. This is particularly serious because often courts lack technical expertise to evaluate AI-generated conclusions. Additionally, high dependence on technical experts in interpreting AI systems may indirectly affect judicial independence by shifting substantial decision-making influence to specialized professionals.

[4] Indian Evidence Act, 1872 – Sections 65A, 65B
[5] Bharatiya Sakshya Adhiniyam, 2023
[6] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473
[7] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1

6. Case Law and Judicial Trends

Indian jurisprudence on AI-specific cybercrime is still in a developing stage. The courts are primarily addressing wider issues concerning digital technology and electronic evidence rather than autonomous AI systems. In *Shreya Singhal v. Union of India (2015)*[8], the Supreme Court highlighted concerns regarding vagueness and overbreadth in cyber law provisions, particularly regarding online speech and intermediary liability. In *Anvar P.V. v. P.K. Basheer (2014)*[9], the Court strengthened the evidentiary framework by mandating strict compliance with electronic evidence requirements, thereby reinforcing reliability standards for digital records. Similarly, in *Varun Kumar v. State (2018)*[10], the judiciary acknowledged the applicability of Information Technology Act provisions in prosecuting cyber offences involving electronic records and online conduct. Collectively, these decisions reflect judicial awareness of the complexities of digital evidence as well as cyber law, but the courts do not yet directly engage with the distinct challenges of AI-driven attribution, autonomy, and liability.

7. International Perspectives

Different jurisdictions across the world have started developing distinct regulatory approaches to combat the challenges created by artificial intelligence and AI-driven cybercrime. The European Union, through the EU AI Act, follows a risk-based model that imposes stricter obligations on high-risk AI systems while emphasizing transparency, accountability, and human oversight. The United States largely depends on sector-specific regulations and traditional tort-based liability principles to govern AI-related harms. China has adopted a more direct regulatory approach by introducing deep synthesis regulations and mandating the labelling of AI-generated content to prevent misuse and misinformation. A comparative analysis of these frameworks reveals a growing international trend toward clearer attribution mechanisms, greater algorithmic transparency, and recognition of shared liability among developers, deployers, and users of AI systems[11].

[8] Shreya Singhal v. Union of India, (2015) 5 SCC 1
[9] Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473
[10] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1
[11] EU AI Act; China Deep Synthesis Regulations; U.S. sector-specific AI governance framework.

8. Key Challenges Identified

8.1 Legal Gaps

Existing Indian cyber laws do not contain comprehensive provisions specifically addressing artificial intelligence and autonomous systems. This creates uncertainty in interpreting liability, intent, and accountability in AI-driven offences.

8.2 Attribution Complexity

AI systems often involve multiple stakeholders including developers, deployers, and end users, making responsibility difficult to determine. The autonomous and adaptive nature of AI further complicates identification of the actual wrongdoer.

8.3 Evidentiary Limitations

Conventional evidentiary principles were developed primarily for traditional digital records and human actions. They are often inadequate to assess the credibility and dependability of AI-generated or manipulated data.

8.4 Enforcement Issues

AI-driven cybercrimes often transcend national borders through anonymous digital networks. This creates jurisdictional conflicts and makes international investigation and enforcement more challenging.

8.5 Ethical Concerns

The accelerated growth of AI technology raises concerns regarding fairness, transparency, and improper use of automated systems. Also, legal regulation must ensure accountability without unnecessarily hindering technological innovation.

9. Recommendations and Reform Proposals

9.1 AI-Specific Legislation

India should introduce dedicated legal provisions specifically regulating artificial intelligence and autonomous systems. Recent governmental initiatives, including policy discussions and

regulatory proposals issued through the Ministry of Electronics and Information Technology, indicate that India is gradually moving toward a structured AI governance framework emphasizing transparency, accountability, and responsible AI deployment[12].

9.2 Forensic Standards

Specialized forensic protocols must be developed for the examination and preservation of AI-related digital evidence. This should include standardized methods for AI evidence analysis, deepfake detection, and verification of algorithmic outputs.

9.3 Human-in-the-Loop Supervisory Accountability Framework

Rather than relying entirely on a generalized shared liability model, a Human-in-the-Loop Supervisory Accountability Framework may provide a more effective mechanism for regulating advanced AI systems. Under this approach, a specifically identifiable human actor should remain responsible for continuously supervising, monitoring, and controlling the functioning of an AI system at a particular point in time. Such supervisory authority must possess the legal and technical ability to intervene, override automated decisions, or deactivate the system whenever necessary, especially in autonomous physical environments such as healthcare technologies, self-driving vehicles, robotics, and drones.

Nevertheless, the implementation of this framework becomes more complex in decentralized or distributed AI systems involved in cybercrime, where decision-making may occur across multiple digital networks without direct human control. Even in such circumstances, the legal system must ensure the existence of clear accountability structures to prevent diffusion of responsibility. Further, liability for AI-driven offences should not remain limited to civil or tortious consequences alone. In cases involving serious harm, financial fraud, threats to public safety, or other grave offences, criminal liability should also extend to human actors who intentionally misuse AI systems or negligently fail to exercise adequate supervision and control.

9.4 Judicial Training

Judges and legal professionals should receive regular training on emerging AI technologies

[12] Government of India, Ministry of Electronics and Information Technology (MeitY), *AI Governance and Regulatory Initiatives*, Press Information Bureau, 2025, available at: http://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/nov/doc2025115685601.pdf

and digital forensic methods. Improved technical understanding would enable courts to assess AI-related evidence more effectively and consistently.

9.5 International Cooperation

Stronger international cooperation is necessary to combat cross-border AI-driven cybercrime and digital fraud. Collaborative mechanisms for investigation, evidence sharing, and extradition can improve global cybercrime enforcement.

10. Conclusion

AI-driven cybercrime represents a paradigm shift in the form of criminal activity, challenging the foundations of forensic attribution, criminal liability, and evidentiary standards. The Indian legal system, while robust in addressing traditional cyber offences, remains inadequately equipped to handle the complexities introduced by artificial intelligence.

The absence of AI-specific provisions in the IT Act, combined with the limitations of traditional criminal law doctrines, creates significant gaps in accountability. Without reforms that incorporate technological realities, there is a high risk of under-criminalization and wrongful attribution.

A forward-looking legal framework must integrate technological understanding, doctrinal innovation, and international best practices to ensure justice in the era of intelligent machines.

References

Statutes

1. Information Technology Act, 2000 – Sections 43, 66, 66C, 66D
2. Bharatiya Nyaya Sanhita, 2023 – provisions relating to cheating, fraud, conspiracy
3. Bharatiya Sakshya Adhiniyam, 2023 – Sections 61–63
4. Digital Personal Data Protection Act, 2023

Case Laws

5. Anvar P.V. v. P.K. Basheer, (2014) 10 SCC 473
6. Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1
7. Shreya Singhal v. Union of India, (2015) 5 SCC 1

International / Comparative Sources

8. EU AI Act
9. China Deep Synthesis Regulations
10. U.S. sectoral AI governance framework

Government / Policy Sources

11. Government of India, Ministry of Electronics and Information Technology (MeitY), AI Governance and Regulatory Initiatives, Press Information Bureau, 2025, available at: http://static.pib.gov.in/WriteReadData/specificdocs/documents/2025/nov/doc2025115685601.pdf
12. NITI Aayog reports on Responsible AI