Volume: 5 Issue: 6 Pages: 26-29

Deepfakes as Emerging Threats to Criminal Investigations

[*1] Swarnima Shukla

[*1] Department of Law, Prof. Rajendra Singh (Rajju Bhaiya) University, Prayagraj, Uttar Pradesh, India.

Article Info.
E-ISSN: 2583-6528
Impact Factor (QJIF): 8.4
Peer Reviewed Journal
Available online:
www.alladvancejournal.com

Received: 03/May/2026
Accepted: 06/June/2026

***Corresponding Author**
Swarnima Shukla
Department of Law, Prof. Rajendra Singh (Rajju Bhaiya) University, Prayagraj, Uttar Pradesh, India.

Abstract
The unprecedented advancement in the field of artificial intelligence has transformed the digital landscape and introduced innovative tools capable of generating hyper-realistic media. Among these developments, deepfakes have emerged as a significant challenge to contemporary criminal investigation procedures. Professors Danielle Citron and Robert Chesney define deepfake as a "hyper-realistic digital impersonation that leverages machine-learning algorithms to insert faces and voices into video and audio recordings of actual people, creating realistic impersonations out of digital whole cloth." While this technology has legitimate applications, its misuse has generated widespread concern regarding misinformation, identity theft, cybercrime, electoral manipulation and the fabrication of digital evidence. The investigative agencies are heavily dependent on electronic and digital evidence, which has amplified the risks posed by deepfake technology. Evidence such as surveillance footage, mobile phone recordings, social media content, and e-communications can be easily manipulated with deepfakes to generate convincing evidence, increasing threats to the integrity of investigative procedures and challenging traditional investigative methodology. Deepfakes can mislead investigators, falsely implicate innocent individuals and generate artificial alibis. This paper examines the concept of deepfakes, the challenges posed by deepfake-generated evidence, evaluates the adequacy of existing legal frameworks in India and explores the role of digital forensics in combating such threats. The paper states that while technological advancements have enhanced investigative capabilities, they have simultaneously created sophisticated avenues for deception that require urgent legal and technological responses.

Keywords: Deepfake, Criminal Investigation, Digital Evidence, Artificial Intelligence, Digital Forensics, Cybercrime, Evidence Law.

Introduction

Traditionally, criminal investigation involves the collection of evidence, including digital and electronic evidence, but due to the unauthorised usage of deepfakes, it becomes almost impossible to differentiate between the original and the tampered evidence. The term deepfake is derived from the combination of “deep learning” and “fake”. Yisroel Mirsky and Wenke Lee define deepfake as “AI-generated content designed to imitate authentic media with a level of realism that challenges human detection.” In other words, it can be understood as the generation of synthetic media through AI techniques capable of producing highly realistic images, videos and audio recordings.

This paper seeks to analyse deepfakes as emerging threats to criminal investigation, the challenges related to digital evidence, their impact on investigative processes and existing legal responses. It also focuses on the role of digital forensic techniques in detecting manipulated media.

Understanding Deepfakes and Their Criminal Potential

Deepfakes can be understood as the creation of videos that imitate real events, using deep learning, AI, and photo-manipulation techniques to spread misinformation. The technologies used involve GANs (Generative Adversarial Networks) or ML (Machine Learning).

Criminal uses of Deepfakes

Deepfakes are used by criminals for identity theft, impersonating prominent figures, sometimes with the intention of manipulating public opinion. Synthesisation of fabricated evidence is also a growing threat for the investigating agencies. Conventionally, during investigation, when cyber-crimes are committed, electronic evidence in the form of call detail records, photographs and SMSs are recovered as substantial pieces of evidence. Section 65-B of the Indian Evidence Act (now Section 63(4) of Bharatiya Sakshya Adhiniyam) provides safeguards to ensure the source

and authenticity of electronic records, but due to increased use of deepfake, it becomes difficult to decide as to whether to treat it as a substantial piece of evidence.

White Collar Crimes

Traditionally, white collar crimes were committed by individuals in positions of influence or professionals in powerful positions, but due to the evolution of deepfake technology, the method of committing these crimes has also changed. Fraudsters employ AI-generated identities and deepfakes to deceive the KYC (Know Your Customer) system, while bots are used to rapidly shift funds across digital assets to mask origins.

Identity Theft and Impersonation

Identity theft through deepfake involves using synthetic media to steal a person's biometric identity, which includes face, voice, and expressions, and using them to commit fraud through impersonation. This not only violates the privacy of that person but also infringes on their economic and social status. For example, artificial media maliciously generated involving a well-known person may become viral even before verifying its authenticity, distorting the person's image. It can also create non-consensual explicit photos, which leads to a breach of the person's privacy**.**

AI-generated Evidence Fabrication

The growing use of AI technology to generate fabricated evidence using deepfakes undermines public confidence in digital evidence. It also creates a dilemma for the investigating agencies involved as to whether to rely on digital evidence or not. A bench of the Supreme Court comprising Justice Rajesh Bindal and Justice Manmohan also expressed its serious concern over the use of fabricated evidence in matrimonial disputes.

Deepfakes as Threats to Digital Investigations

The widespread accessibility of deepfake technology has significantly increased the risk of fabricated digital evidence, thereby creating substantial challenges for investigative agencies. The reliability of digital evidence has plummeted over recent years.

Digital Evidence in Criminal Investigations

As we know, criminal investigation involves critical analysis of CCTV surveillance and phone call records, tracking location, tracing digital footprints, etc., but due to the free-hand access to generative AI technology, digital evidence that was earlier treated as substantive piece of evidence by the investigating authorities has now lost its credibility.

Investigative and Forensic Challenges

Deepfake technology provides synthetically altered media, which poses challenges to determining the authenticity and reliability of digital evidence.

Liar's Dividend

American legal scholars and cybersecurity experts Bobby Chesney and Danielle Citron coined the term 'liar's dividend'. The concept of the "liar's dividend" refers to the advantage gained by individuals who exploit public awareness of deepfake technology to evade accountability. As manipulated media becomes increasingly realistic, people with authentic recorded misconduct may attempt to discredit genuine audio or video evidence by claiming that it has been artificially manipulated. This issue becomes more serious as society becomes more informed about the existence and capabilities of deepfake technology. Public awareness is essential for recognising fabricated content; however, it may also encourage scepticism towards authentic recordings. As a result, even reliable digital evidence can be questioned, casting uncertainty on investigations. Therefore, the liar's dividend represents an unintended consequence of deepfake technology, where the mere possibility of manipulation allows individuals to cast doubt on truthful evidence. Cognitive biases further fuel the doubt cast upon authentic evidence, which poses a significant challenge to criminal investigations in the digital age.

False Incrimination

Deepfakes can easily clone a person's voice, expressions and videos in a manipulative way, which can make an innocent person guilty of criminal activities. This is one of the prominent reasons why courts undermine the traditional presumption of authenticity that was once placed in digital evidence.

Artificial Alibis

The original legal defence used is known as the plea of alibi, in which the accused claims that they were not present at the scene of the crime and has recorded proof of their presence elsewhere. The concept of artificial alibis refers to a fabricated or digitally manipulated defence created to deceive the investigators by faking a person's presence or activity at a specific time.

Witness Manipulation

Any inconsistency in witness and victim testimonies puts their credibility in a precarious position. In such systems, where an already traumatised victim may be dealing with difficulty in recollecting facts, false evidence generated by deepfake technology may perplex the victim further, pushing them into a state of dysphoria about the criminal justice system.

Role of Digital Forensics in Detecting Deepfakes

Digital forensics plays a vital role in identifying deepfakes and verifying the authenticity of digital evidence. With the help of advanced technologies such as artificial intelligence, machine learning and biometric analysis, forensic experts can verify the authenticity of images, videos, audio, etc.

Biometric Analysis

Forensic investigators examine videos frame by frame to identify unusual features such as extra fingers in hands, irregular body structure, eye movements, distorted facial features, etc. If any such feature is detected, then it can be concluded that the video is created to mislead them, as deepfakes are not able to comprehensively copy all features.

Metadata Examination

The National Institute of Standards and Technology (NIST) defined Metadata as "Data about data. For file systems, metadata is data that provides information about a file's contents." For detecting deepfakes, forensic experts inspect metadata, including compression patterns and digital artefacts that may reveal editing. Developing technologies such as digital watermarks, including Google's SynthID, also help to identify the authenticity of available content.

Audio and Contextual Analysis
Experts analyse speech patterns, voice characteristics and other vocal features to detect any manipulation. It is also assessed whether the background, shadows, reflection, and overall scene are natural.

Artificial Intelligence-Based Detection Tools
With the growth of AI, forensic experts have started using AI-powered detection systems for detecting manipulation. These tools are trained to identify patterns and artefacts commonly produced by deepfakes.
Digital Forensics has emerged as an essential tool in ensuring justice and preserving public trust in the legal system.

Obstacles Faced by the Legislature in Regulating Deepfake Technology
Evolving Nature of Deepfake
The pace at which technology advances surpasses the speed at which laws can be formulated and enacted. Hence, legal measures risk becoming outdated shortly after implementation.

Global Nature
Deepfake content can be created in one country and disseminated in another, which poses a jurisdictional and enforcement issue. Furthermore, identifying creators of deepfakes often requires sophisticated technical expertise and cross-border cooperation.

Lack of Technical Expertise
The lack of technical expertise among policymakers, along with the need to coordinate with technology companies and cybersecurity experts, presents a hindrance to the development of the legislation.

Balancing Regulation and Freedom of Expression
The Indian Constitution safeguards freedom of speech and expression as a Fundamental right of the citizens. Legislatures must ensure that regulations targeting harmful deepfakes do not violate constitutional or human rights protections related to speech, journalism and political commentary.

Judicial Developments
The proliferation of deepfakes has created a crisis for criminal investigations and the legal system. Due to this sudden shift in the nature of crime, the legislature is facing difficulty in taking action against all threats of digital nature. In the absence of specific laws dealing exclusively with deepfakes, the judiciary has stepped forward and interpreted the existing laws for protecting individuals from the harmful effects of manipulated digital content.

Voice Cloning and Personality Rights
The Bombay High Court in the case of Arijit Singh v. Codible Ventures LLP, through the bench of Justice Riyaz Chagla, stated that "Even though freedom of speech and expression allows for critique and commentary, it does not grant the license to exploit a celebrity's persona for commercial gain. In these circumstances, this Court is inclined to protect the Plaintiff against any wrongful exploitation of his personality rights and the right to publicity." The Judge also observed, "Making AI tools available that enable the conversion of any voice into that of a celebrity without his/her permission constitutes a violation of the celebrity's personality rights." By this, the court focused on celebrity and personality rights and how they are vulnerably targeted by these deepfakes. The Court also ordered various entities to remove such content.

John Doe Order
The Delhi High Court passed a John Doe order protecting the personality right of singer Jubin Nautiyal. Justice Tushar Rao Gedela said, "The plaintiff has a prima facie strong case and, having regard to his well-known, popular and well-accepted personality, the balance of convenience is tilted in favour of the plaintiff. In case ex parte ad-interim injunction and other directions, as sought, are not passed, the irreparable loss and injury which may be occasioned may not be compensated in monetary terms." By this, we can conclude that courts have always acted as a guardian in protecting the personality rights of citizens.

Removal of Deepfake Videos
The Bombay High Court in the case of National Stock Exchange of India Ltd. v Meta Platforms, Inc, & Ors. directed Meta Platforms for the removal of deepfake content of the MD and CEO of National Stock Exchange, in which he was depicted telling common people to join WhatsApp groups and channels to take stock-picking tips.

Legal Developments in India
The rapid advancement of deepfake technology has raised significant legal and regulatory concerns in India; consequently, certain developments are being made in the legal field.

The Information Technology Act, 2000
This Act lays the foundation of laws in the technical fields, including punishment for cheating by impersonation using computer sources. It penalises privacy violations through capturing or transmitting private images without consent. This Act also imposes stringent punishment for publishing or distributing obscene or sexually explicit material electronically.

Digital Personal Data Protection (DPDP) Act, 2023
This Act considers all digital data as personal data; thus, creating a deepfake without the explicit consent of the person involved constitutes a fundamental breach of data processing rights. The Act makes it a requirement to take consent of an individual before using his personal data. A victim of deepfake may also request the removal or erasure of their personal data from platforms or entities that unlawfully process it. Monetary penalties are imposed on organisations that fail to protect personal data.

IT Amendment Rules (2026)
To address concerns relating to fake content, privacy, personality and digital safety, the government introduced amendments under the IT Amendment Rules (2026), which imposes greater responsibilities on online platforms. The rules formally codify Synthetically Generated Information (SGI) as audio, visual, or audio-visual information created or altered using AI/algorithmic resources that convincingly appears real and portrays individuals or events falsely.
Previously, social media platforms had up to 36 hours to remove illegal content. Under the 2026 rules, platforms must remove or disable access to prohibited deepfakes within 3 hours of receiving a government or court order. In critical scenarios, the deadline drops to 2 hours.

Platforms that provide AI creation tools must embed permanent metadata, provenance markers, or unique identifiers into the outputs. For visual deepfakes, a visible "AI-Generated" label must be permanently visible; for audio, a prominently prefixed audio disclosure that can be used to identify synthetically generated audio must be present. Large social media platforms are required to prompt users to declare whether the content they are uploading is synthetic.

Recommendations

To address the threats caused by deepfakes to criminal investigations, recommendations include:

1. A dedicated legal framework should be introduced regulating the creation, distribution and misuse of deepfake content.
2. Any content created with the help of AI tools must mandatorily be labelled as AI-generated content.
3. Strict penalties must be imposed for malicious use of deepfakes.
4. Social media platforms should detect, flag, and remove harmful deepfake content of any nature.
5. The government should invest in advanced deepfake detection tools.
6. Public awareness campaigns should be conducted to educate about the risks and identification of deepfakes.
7. Specialised training should be given to police officers, public prosecutors, and judicial officers on deepfake technology, digital evidence, and investigation techniques. 8) Nations should cooperate with each other to address transnational deepfake-related offences.
8. Regulations should ensure that measures against deepfakes do not unnecessarily restrict freedom of speech and expression while adequately protecting privacy, reputation, and public interest.

Conclusion

Deepfake technology has emerged as a serious threat to criminal investigations. It easily creates fake videos, images, and audio recordings that can appear real, making it difficult for investigators and courts to distinguish between genuine and manipulated evidence. The increased misuse of deepfakes can result in false allegations, damage to reputation, misleading investigations, and challenges in delivering justice, creating a sense of crisis in the legal field.

Although existing laws can be used to deal with some deepfake-related offences, they are not sufficient to fully address the unique challenges created by this technology. To tackle this, the use of digital forensic tools, effective legal measures, and advanced detection technologies has become increasingly important in identifying fake content and protecting the integrity of investigations.

In order to effectively address this threat in the legal system, there is a need for stronger legal frameworks, improved forensic capabilities, greater public awareness, and cooperation between nations, technology companies, social media and law enforcement agencies. Such measures will help to ensure that criminal investigations remain fair, accurate, and reliable in the digital age.

References

Books

1. Kumar. Narender-Key to Criminal Courts, 7th Edition, 2024.

Case Laws

1. Arijit Singh V. Codible Ventures LLP (com IPR Suit (L)/23443/2024)
2. Jubin Nautiyal V. Jammable Ltd. & Ors. (CS(Comm) 166/2026)
3. National Stock Exchange of India Ltd. v. Meta Platforms, Inc. & Ors. (Interim Application (L) No. 21456 of 2024 In Com IPR Suit (L) No. 21111 of 2024)

Articles

1. Danielle Keats Citron & Robert Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy and National Security
2. Yisroel Mirsky & Wenke Lee, the Creation and Detection of Deepfakes: A Survey.
3. National Institute of Standards and Technology

Bare Acts

1. Bharatiya Nyay Sanhita 2023 (India)
2. Indian Evidence Act, 1872 (India)
3. Code of Criminal Procedure, 1973 (India)
4. The Information Technology Act, 2000.
5. Digital Personal Data Protection (DPDP) Act, 2023
6. IT Amendment Rules (2026)