DEEPFAKE EVIDENCE AND CRIMINAL TRIALS – CHALLENGES TO JUSTICE IN THE AI ERA

By: Rishita Singh

ABSTRACT

The emergence of deepfake technology, synthetic media created a profound challenge to the credibility of digital credibility in adjudications. Deepfakes have capability to fabricate content to be used for evidence, manipulate surveillance footage, or produce deceiving substantiation statements, thereby threatening the integrity of the judicial process. This paper explores the technological underpinnings of deepfakes and their implicit abuse in judicial proceedings. Though we have introduced Bharatiya Sakshya Adhiniyam 2023 this Act does not have retrospective effect and the cases filed before 1st July 2024 are still guided under the erstwhile Indian Evidence Act, 1872. This paper evaluates India's current legal framework, particularly the Indian Evidence Act, 1872 and Information Technology Act 2000, and highlights the insufficiency of these laws in addressing synthetic media contents. A relative analysis with the United States and the European Union demonstrates that India lags in both regulation and enforcement mechanisms. This paper emphasizes the need for legislative reform, judicial training, forensic invention, and public mindfulness for fair trial. In this digital era where visual and audio content can no longer be taken on their face value, securing the authenticity of electronic evidence is a challenge to secure justice.

Key Words: Deepfake, Indian Evidence Act, Credibility and Judicial Proceedings.

INTRODUCTION

The advancement of artificial intelligence (AI) has introduced transformative possibilities across different sectors, but one of its most trending is the emergence of deepfake technology. Deepfakes are largely realistic but synthetic media, generally video or audio clips, that depict people saying or doing something which they no way actually said or did. These are generated using sophisticated machine literacy models, most especially Generative inimical Networks

(GANs), where two neural networks, the ‘creator’ and the ‘discriminator’ contend in refining content until the same is nearly indistinguishable or resembles with the authentic media.

While the technology used in entertainment, education, and other sectors, its abuse poses serious threats to the criminal justice system. Deepfakes can fabricate admissions, confessions, statements, tamper with surveillance footage, or induce fake voice recordings, potentially altering the original which can be used in the court of law. A videotape of a suspect committing a crime, if synthetically generated, could lead to superfluous persuasions, eroding the public trust in judicial system.

The urgency of this trouble is particularly notable in India, where the evidence is electronic or photographs or videos. Sections 65A and 65B of the Indian Evidence Act, which govern the admissibility of electronic evidence, were framed long before the booming of deepfake technology [1]. The issue with the Indian court system is that, despite the introduction of the Bharatiya Sakshya Adhiniyam 2023, this Act lacks retrospective effect, meaning that cases filed prior to July 1, 2024, will continue to be governed by the previous Indian Evidence Act of 1872. Numerous cases remain pending that were initiated prior to July 1, 2024; hence, the Indian Evidence Act continues to be applicable in judicial procedures till their final disposal.

Although in Anvar P.V. v. P.K. Basheer and Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, [civil appl no. 20825-20826 of 2017], [2] it has been clarified on the procedural conditions of electronic evidence, they do not address authentication of AI- generated or altered media. Also, the Information Technology Act, 2000, while robust in certain areas of cyber law, remains largely silent on the issue of deepfakes evidentiary dangers.

The authorities of the United States and the European Union are taking primary way to address the problem through specialized investigation and legal reform. For instance, the EU’s AI Act 2024 make platforms accountable for the synthetic media, and U.S.A, States like California have introduced Anti-Deepfake legislation. Still, the international and decentralized nature of the laws, combined with the viral eventuality of deepfakes, limits the effectiveness of the statutes.

Through a detailed review of current literature, statutes, and relative practices, this paper aims to highlight critical gaps in the Indian Evidence Act and propose a roadmap for solution. The thing is to ensure that the integrity of digital substantiation is saved, indeed in a period when seeing even lacks credibility.

CONCEPTUAL UNDERSTANDING

The rapid growth of artificial intelligence, especially in the form of deepfakes, has urged legal scholars worldwide to study its effect in the justice delivery system. A crucial focus of this paper is to study the admission of digital evidence in ultramodern criminal justice system and to assess the need to robust practice to prevent abuse of deepfake technologies in court.

1. Digital Evidence in trials

Mekkawi. (2023). The Challenges of Digital Evidence Usage in Deepfake Crime. *Journal of Law and Emerging Technologies*.

Digital evidence refers to any information stored or transmitted in digital form that can be used in a court of law to prove or falsify a fact. Judge Mohamed Hassan Mekkawi, in his paper “The Challenges of Digital Evidence Usage in Deepfake Crime”, outlines how digital bias, ranging from smartphones and laptops to surveillance systems, leave behind vast quantities of data that can be critical in cyber examinations. He notes that ultramodern crimes are frequently committed in a digital mode, making digital evidence essential for both execution and defence. Still, the nature of this evidence is complex, largely unpredictable, and can be altered or deleted.

To be accepted in court, digital substantiation must satisfy certain principles, like data integrity, inspection, legitimacy, and proper chain of guardianship. Mekkawi highlights how encryption, steganography (caching data in images or media), and the adding use of the Internet of effects (IoT) complicate its collection. These challenges demand technical training and advanced forensic tools to ensure that digital evidence is saved and presented in its original form in the courts.

2. Legal narrative on AI- Generated substantiation

Citron, D. K., & Chesney, R. (2019). *Deep Fakes: a looming challenge for privacy, democracy, and national security*. Scholarly Commons at Boston University School of Law. https://scholarship.law.bu.edu/faculty_scholarship/640/

Legal scholars Danielle Citron and Robert Chesney, in their paper "Deep Fakes: a looming challenge for privacy, democracy, and national security", showed how deepfakes, AI-generated videos, images, or audio recordings, are creating trouble in the legal and political systems. They explain that deepfakes are created using Generative Adversarial Networks (GANs), a machine learning technology where one neural network generates content while another works on its resemblances. Citron and Chesney raise legal concerns regarding its credibility, liability, and abusiveness.

3. Deepfake Discovery and Admissibility enterprises

While technocrats have developed several tools to detect deepfakes, similar as digital watermarking and forensic analysis of facial and voice movements, the admissibility as evidence in court remains uncertain. Traditional legal systems are based on the supposition that videotape or audio recordings are admitted subjected to the certificate under section 65B of the Indian Evidence Act.

Mekkawi notes that no transnational standard presently exists for validating deepfake content as legal evidence. Countries like the United States have started to respond with state-specific laws (e.g., California anti-deepfake law for election hindrance), but global norms remain inconsistent. In India, deepfake-specific laws are nearly absent, and current rules under the Indian Evidence Act (Sections 65A and 65B) do not address synthetic media directly.

The courts in India have relied heavily on precedents like Anvar P.V. v. P.K. Basheer and Arjun Panditrao v. Kailash Kushanrao, which clarified the admissibility of electronic records. But these rulings are grounded on conventional mode of electronic documents or recordings and not related to the synthetic or AI- generated contents.

4. Gaps in the research and Indian Context

There's a clear gap in Indian legal research and justice on how to handle deepfake evidence or to shun it. While the EU and the U.S.A are exploring AI regulation. India's legal approach remains reactive rather than visionary. The lack of digital knowledge among legal professionals, and the limited authenticity of AI forensic tools in Indian courts, further escalates the threat of manipulation and abuse of evidence. This not only affects the outgrowth of individual cases but also threatens broader principles of due process and fairness in criminal justice system.

EVIDENTIARY CHALLENGES OF DEEPFAKES

1. How Courts Verify Audio- Video substantiation

The courts corroborate the facts with the evidence and statements of the witnesses. In India, sections 65A and 65B of the Indian Evidence Act govern the admissibility of electronic records. These sections emphasize the demand for an instrument attesting the integrity and origin of the digital record.

The Supreme Court in Arjun Panditrao Khotkar v. Kailash Kushanrao highlighted the obligatory nature of Section 65B [1]. Yet, these rules were designed for conventional digital substantiation, similar as CCTV footage or call records, not AI- manipulated media that can produce false admissions, fake surveillance videos, or manipulated victim statements.

In authorities like the U.S., courts apply Federal Rule of Evidence (FRE) 901 and 702 to authenticate digital media. Rule 901 demands that evidence must be authenticated that the presented evidence before the Court is what it is in actual form. In deepfake cases, this is particularly difficult because visual cues of editing may be undetectable without technical software. Consequently, the creation of deepfake information that is accepted as evidence by the court poses a significant threat to the justice delivery system.

2. Are Current Forensic Tools Enough?

Detecting deepfakes requires advance forensic analysis but sometimes it is not detectable. While some software tools can lead to, advanced deepfake contents produced by GANs (Generative inimical Networks) can bypass detection tools of any advance levels. According

to experts, deepfake discovery is presently running behind deepfake generation in a technological arms race. Professor Hany Farid, a colonist in digital forensics, warns that “we’re decades down” from dependable, universal deepfake discovery.

Some inventions like Microsoft's videotape authenticator aims to indicate inconsistencies in frame picture or search for mismatches in manipulated footage by comparing the originals contents. Still, these are either in testing stage or lack judiciary acceptance.

In India, the lack of structure and trained digital forensic experts is a major fallback. Indeed, when deepfakes are suspected, there's no clear protocol for flagging it leading to evidentiary abuse.

3. Real- Life exemplifications of Deepfake Misuse in Law and Crime

India

Several police complaints and primary reports suggest that AI- generated intimate videos and deepfake-grounded blackmail are rising. The Information Technology Act criminalizes certain forms of cyber manipulation (e.g., under Section 66E or 67), but does not address synthetic manipulation like voice cloning or face-switching.

United States

The U.S. has witnessed deepfake-related crimes. For example:

- In California and Texas, laws were passed to criminalize deepfakes used to influence the voting preference, especially if they're released within 30 days before voting. In Texas SB 751, distributing deepfake videos is an offence if circulated with the intent to harm campaigners.
- In Virginia, deepfake pornography has been criminalized.
- These above facts show that U.S. laws treat deepfakes not as standalone crimes, but as a criminal tool used to commit crime such as child pornography, damage reputation, commit fraud or influence public opinion.

European Union

In the European Union, the proposed AI Act 2024 includes provisions making the platforms and generators to label synthetic media as deepfake generated, especially in journalism and public opinion. While this does not directly address AI media, it's a pre-emptive measure to reduce the detriment effect of deepfakes entering the courtroom as evidence to influence the justice delivery system.

A notable global illustration includes the deepfake videotape of Ukrainian President Volodymyr Zelensky, where a fake videotape showing surrendering to the Russia during the 2022 war. Though not used in a court, this case underscores the implicit public security and influencing public views.

In summary, courts face significant challenges towards deepfake evidence. Traditional legal safeguards are inadequate to address the nuanced pitfalls of AI- generated media. While forensic tools are evolving, they remain limited in compass and trustability. As real-world exemplifications of deepfake abuse continue to crop, both in India and international, there's a clear need for a legal and specialized guiding frame that can directly spot, authenticate, and regulate synthetic evidence in criminal trials.

Legal Framework in India

The adding reliance on electronic substantiation in Indian felonious trials has urged a critical analysis of its legal frame, especially in the wake of complex challenges like deepfakes. Deepfakes AI- generated media that mimic real individualities' speech or conduct pose a significant threat to evidentiary integrity. Unfortunately, Indian statutory provisions and judicial precedent were not designed with this type of sophisticated manipulation in mind.

1. Indian Evidence Act Sections 65A and 65B

Section 65A of the Indian Evidence Act, 1872 deals with the admissibility of electronic records and refers to the procedures outlined in Section 65B. Section 65B provides the conditions under which an electronic record can be admitted as evidence in court.

In Anvar P.V. v. P.K. Basheer, [4] the Supreme Court held that secondary electronic evidence (like CDs, DVDs, and dispatch printouts) must be produced with Section 65B, especially the

electronic evidence. This overruled the earlier State v. Navjot Sandhu, Case number: Appeal (crl.) 373-375 of 2004 decided in 2005 ruling, which allowed admission under Section 63 and 65 for secondary evidence without section 65B. The ruling emphasized that bare production of electronic evidence without 65B is not sufficient for admissibility.

Later, in Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal (2020), the Court clarified that the instrument under Section 65B (4) is obligatory unless the original file is not traceable, in which the court can use its discretion to admit the evidence under Section 65B (1). The Court stressed that authentication is important to ensure that the relevant electronic evidence has not been tampered.

Still, neither of these rulings addresses synthetic or fabricated media generated by AI where traditional styles of instrument and source verification is either did not take place or is impossible. For illustration, if a deepfake videotape is introduced, the source may appear legible, but the content could be wholly fabricated using GANs (Generative inimical Networks) and therefore if accepted as evidence then the justice will be at risk.

2. Information Technology Act, 2000

The Information Technology Act governs cybercrimes and data protection [5]. But it lacks specific threats posed by deepfakes. While Section 66 deals with computer- related offenses and Section 67 penalizes the transmission of stag material, there are no clauses that explicitly address the manipulation or fabrication of videotape/ audio using AI. Although some offenses could be brought under the scope of deepfake fraud (e.g., Section 66D for impersonation), there's no structured frame for specifically dealing with deepfakes as evidence, neither in terms of admissibility nor trustability.

3. Are Current Laws Able of Handling Deepfakes?

At present, India’s laws are not adequately equipped to address deepfake pitfalls in the courtroom. The demand of a 65B certificate assumes that the digital content is firstly created or stored in each device or platform and that its integrity can be verified. Deepfakes, still, can

be created on open-source platforms, uploaded anonymously, or spread virally through social media without any traceable source.

There are three crucial gaps: -

1. Authentication Tools- Indian system lacks digital forensic tools specifically able of detecting deepfakes.

2. Lack of Deepfake Regulation- While the EU's AI Act and U.S. State laws (e.g., California's deepfake election law) are addressing synthetic media, India has no specific statute.

3. Judicial Awareness and Technical Capacity- Most trial courts lack specialized technical tools to determine the manipulated content. Judges frequently depend on the expert opinion and on forensic reports, which are inadequate against GAN- generated

While the precedents set in Anvar and Arjun Panditrao mark significant progress in electronic evidence, they're not fit for deepfake complications. Neither the Indian Evidence Act nor the IT Act or even the present Bharatiya Sakshya Adhiniyam, presently provides mechanisms to corroborate or rebut synthetic audio-visual content. To address this, India must urgently modernize its legal delineation, make specialized structure, and provide clear evidentiary rules for electronic evidence than may have been generated by AI to uphold the fair trial.

DEEPFAKES & RIGHT TO FAIR TRIAL

1. Article 21 and the Right to a Fair Trial

Article 21 of the Indian Constitution guarantees the right to life and liberty, which includes the right to a fair trial. This right assured that justice is neither delayed nor distorted. The use of deepfake evidence in criminal proceedings poses a direct threat to the right to life by introducing the risks of misinformation, misidentification, and factual manipulation, hence potentially embroiling innocent individuals in retaliatory lawsuits.

2. Burden of Proof and Authentication

In criminal proceedings, the burden of proof rests with the party alleging the occurrence of the incident; ultimately, it is the prosecution's responsibility to establish guilt. However, in India, there is a significant violation of this norm in instances of rape or sexual offenses against women or children. The burden of proof shifts on the accused in these offenses. When a deepfake is presented as evidence accompanied by a section 65B certificate under the Indian Evidence Act, there is no guarantee that the evidence is not artificially manufactured and is grounded in factual accuracy.

The evidentiary value under Sections 65A and 65B of the Indian Evidence Act is inadequate for establishing the authenticity of deepfake content, increasing the likelihood that the court will resolve the criminal trial solely exclusively on Section 65B concerning electronic evidence. In the absence of technological forensic software or AI specialists, Indian courts will likely encounter difficulties in determining the authenticity of electronic content or images.

3. Threat of wrongful accusation

The use of deepfake evidence in trials raises a shocking possibility that innocent individuals could be dragged into false criminal charge. For illustration, a deepfake videotape showing a person at the scene of a crime could mislead a judge or jury into believing the accused person has committed the crime. Likewise, real proof may be dismissed under the doubt of falsification.

Similar scripts produce a nipping effect on the administration of justice, where digital evidence loses credibility, and original fact become heavily reliant on precious forensic assessments, beyond the reach of utmost trial courts in India.

RECOMMENDATIONS AND CONCLUSION

The growth of deepfake technology has created a significant vulnerability within the legal system, particularly in the domain of criminal justice. This study has emphasized the inadequacy of the legal system in confronting the difficulties presented by AI-generated synthetic media. The integrity of evidence, judicial reliability, and individual rights pose significant threats that necessitate prompt attention to preserve justice in the digital era.

1. Legislative Reforms Define and Regulate Synthetic Substantiation

The Indian Evidence Act has been superseded by the Bharatiya Sakshya Adhiniyam, 2023, rendering amendments implausible. In this situation, necessary ordinance must be passed to specifically define the method for admitting synthetic evidence, including AI-generated films, photographs, and audio recordings.

This would bear adding a new provision that defines synthetic evidence as media created or altered using artificial intelligence, machine literacy, or related technologies. Clear norms for admissibility should be established, taking not just metadata or instruments under Section 65B, but also forensic authentication of content, including evidence that it has not been manipulated using GANs or analogous tools.

Such a provision would help for unlawful persuasions grounded on altered digital content and give courts a legal base to identify suspicious material unless it passes defined thresholds of authenticity.

2. Establishment of Specialized Forensic Labs

There is a critical need to produce devoted digital forensic laboratories equipped with AI and deepfake discovery tools. These labs should have the capacity to dissect visual and audio media for signs of tampering, use neural network analysis to descry inconsistencies, and issue credibility assessments permissible in court.

Presently, not many Indian institutions retain the position of specialized AI detection. Establishing forensic labs and increasing the well-equipped judges would give trial courts access to adjudicate the veracity of digital evidence and judge the credibility of its mechanisms. These labs could also support law enforcement in examinations involving synthetic media used for blackmail, fraud, or misinformation.

3. Judicial Training and Capacity Structure

The bar must be equipped to handle the specialized complications of AI-grounded evidence. The judges and lawyers are not trained to understand the AI. Regular judicial training programs, in collaboration with public law seminaries and AI experts, should be introduced.

These seminars should cover: -
• The functioning of deepfake technology (e.g., GANs).
• Modes of discovery and analysis.
• Interpretation of expert reports on synthetic media.
• Legal norms for admissibility and authentication.

This will enable judges to ask the right questions, critically estimate digital evidence, and issue rulings predicated in both legal and scientific principles.

4. Public Mindfulness and Legal Knowledge

As deepfake technologies has become more accessible, the public must be educated about their threats. Legal knowledge should be disseminated among the citizens. Educational institutions, social media platforms, and NGOs can play a vital part in creating awareness by circulating videos, posters, legal toolkits and promote responsible media.

The deepfake technology demands a visionary legal response. Without decisive reforms, India's justice system remains exposed to serious pitfalls on the admission of fake evidence, unlawful persuasions, and diminishing of the public trust. We need to close the gap between legal safeguards and technological difficulties. We need to uphold not only the integrity of our courts but the uphold the right to life and personal liberty under Article 21 of the Constitution of India. Every individual deserves a fair trial in the age of artificial era.

REFERENCES: -

[1] Laws of India. (2023, September 2). *Electronic evidence in the Indian Evidence Act: Navigating the digital frontier*. https://lawsofindia.com/2023/09/02/electronic-evidence-in-the-indian-evidence-act-navigating-the-digital-frontier.

[2] AIR 2020 SUPREME COURT 4908.

[3] Ridhi, V. N., Aggarwal, A., & KJT. (2021, June 7). *The decision in Arjun Panditrao: Admissibility of electronic evidence in India continues to face hurdles*. SCC Times. https://www.scconline.com/blog/post/2021/06/07/electronic-evidence-2/.

[4] 2014 10 SCC 473.

[5] Jha, R. (2025, January 30). *Cybersecurity laws in India: A comprehensive guide*. Lexology. https://www.lexology.com/library/detail.aspx?g=d599eba2-e69a-4121-95b4-ff84e49730c6.