International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Deepfakes and Legal Accountability: A Threat to Evidence and Reputation

Mr. Mukul Mukul

Assistant Professor, Law, Geeta Institute Of Law

Abstract

The rise of deepfake technology — hyper-realistic but artificially generated audio, video, or images — poses a significant threat to both legal systems and individual reputations. As these synthetic media become increasingly sophisticated, they blur the lines between real and fake, raising alarming concerns about the authenticity of digital evidence and the potential for reputational harm. In legal proceedings, where evidence must be credible and admissible, deepfakes introduce uncertainty, jeopardizing the fairness of trials and investigations. Beyond the courtroom, individuals can fall victim to manipulated content designed to defame, blackmail, or mislead the public, often with irreversible consequences. This paper explores the dual-edged impact of deepfakes on evidence and personal integrity, examines current legal gaps in addressing this digital menace, and underscores the urgent need for a robust legal and technological framework to ensure accountability. By balancing innovation with legal safeguards, we can begin to mitigate the risks posed by this evolving threat.

Keywords: Deepfakes, Legal Accountability, Digital Evidence, Reputational Harm, Synthetic Media, Technology and Law, Cybercrime, Misinformation, Privacy Violation, Legal Framework, Admissibility of Evidence, Media Manipulation.

1. Introduction

The digital revolution has ushered in significant advancements in artificial intelligence, one of the most controversial being the emergence of "deepfakes." These are hyper-realistic audio-visual media created using deep learning techniques, particularly Generative Adversarial Networks (GANs), which enable the manipulation of images, videos, and voices to mimic real individuals with remarkable accuracy[1]. Originally developed for creative and entertainment purposes, such as film production and language dubbing, deepfakes have evolved into a potent tool for misinformation, defamation, and cybercrime[2].

The misuse of deepfakes presents a profound threat to legal systems and societal trust. As digital evidence becomes increasingly central in both civil and criminal litigation, the authenticity of such material is critical. Deepfakes have the potential to distort reality, fabricate events, and falsely implicate individuals, thereby undermining judicial integrity and due process[3]. The Indian legal system, though equipped with

[1] Goodfellow, Ian, et al., "Generative Adversarial Networks", *Communications of the ACM*, Vol. 63, No. 11, 2020, pp. 139–144
[2] West, Darrell M., "How to Combat Fake News and Disinformation", *Brookings Institution Report*, 18 Dec. 2017.
[3] Singh, M., "Digital Evidence and its Admissibility in Indian Courts: Issues and Challenges", *NLUJ Law Review*, Vol. 8, No. 2, 2020, pp. 97–113.

the Information Technology Act, 2000 and provisions of the Indian Penal Code, 1860, lacks specific statutes to address the sophisticated challenges posed by synthetic media[4],[5].

Beyond legal proceedings, deepfakes have become a serious menace to personal reputation and mental well-being. Numerous individuals have been targeted through non-consensual deepfake pornography, political impersonations, and falsified interviews or statements circulated on social media, leading to irreparable reputational harm and emotional trauma[6]. In this evolving digital landscape, where seeing is no longer believing, the line between truth and fabrication continues to blur.

This paper aims to critically examine the implications of deepfakes on evidence integrity and individual dignity. It explores technological developments, evaluates legal responses in India and abroad, and proposes reformative strategies to ensure accountability in an era where digital deception is alarmingly accessible.

2. Understanding Deepfakes

2.1 Definition and Evolution

The term *deepfake* is derived from a combination of "deep learning" and "fake," referring to synthetic media created using artificial intelligence techniques that convincingly replicate the likeness, voice, or actions of real individuals. Deepfakes first gained public attention around 2017, when online platforms saw a surge in manipulated videos, especially non-consensual celebrity content[7] While digital manipulation is not new, the scale and sophistication brought by AI-powered deepfakes marked a new chapter in media falsification.

Initially developed for legitimate uses in the entertainment industry, such as improving visual effects in cinema or voice replication for dubbing, the technology has rapidly evolved. It is now accessible to the general public through open-source tools and mobile applications, significantly lowering the barrier for creation and misuse.[8]

2.2 Technology Behind Deepfakes (AI, GANs, etc.)

At the core of deepfake creation is a form of machine learning known as *Generative Adversarial Networks* (GANs). Introduced by Ian Goodfellow and his team in 2014, GANs work by pitting two neural networks against each other — a *generator* that creates synthetic content, and a *discriminator* that evaluates its authenticity[9]. Through multiple iterations, the generator improves its output to fool the discriminator, resulting in highly realistic, yet entirely fake, images or videos.

In addition to GANs, other technologies such as autoencoders, deep neural networks, and facial recognition software are often used to enhance accuracy and realism. The incorporation of audio

[4]Information Technology Act, 2000 (India).
[5] Indian Penal Code, 1860 (India).
[6] Agarwal, A., "Deepfakes and Indian Law: An Urgent Need for Regulation", *Indian Journal of Law and Technology*, Vol. 17, 2021, pp. 45–60.
[7] Chesney, R. and Citron, D. K., "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, Vol. 107, 2019, pp. 1753–1819.
[8] Paris, B. and Donovan, J., "Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence", *Data & Society Report*, 2019.
[9] Goodfellow, Ian, et al., "Generative Adversarial Networks", *Communications of the ACM*, Vol. 63, No. 11, 2020, pp. 139–144.

deepfakes, which clone voices using spectrogram-based models, has added a new layer of complexity, making it difficult even for trained professionals to identify manipulation without forensic tools[10].

2.3 Common Uses and Misuses

While deepfake technology holds promise in fields such as education, cinema, gaming, and language localization, its misuse has vastly outpaced its benefits. Malicious applications range from the creation of fake pornography, political misinformation, and corporate fraud, to manipulating evidence in legal cases. Several high-profile instances have shown how deepfakes can be used to impersonate politicians, forge confessions, or incite communal disharmony[11].

Moreover, the emergence of *cheap fakes*—low-effort manipulations that do not require sophisticated AI—also contributes to public confusion and undermines trust in authentic media[12]. With social media platforms enabling rapid dissemination, even debunked deepfakes can cause lasting reputational damage before they are taken down.

As this section illustrates, understanding the origin, technical structure, and misuses of deepfakes is essential to framing effective legal, ethical, and policy-based responses in the face of a rapidly evolving threat landscape.

3. Deepfakes as a Threat to Legal Evidence

As the justice system increasingly relies on digital formats for evidence—CCTV footage, audio recordings, and digital photographs—the integrity of such material becomes vital. Deepfakes, by introducing uncertainty into what was once considered indisputable, pose a direct threat to the administration of justice. The capacity to fabricate audio-visual evidence indistinguishable from the real challenges long-standing principles of proof, trust, and due process.

3.1 Manipulation of Audio-Visual Evidence

Traditionally, video and audio recordings have served as compelling forms of legal evidence. However, deepfakes undermine their reliability by enabling realistic forgeries of voice and facial expressions. For instance, a manipulated video could falsely depict a person confessing to a crime or being present at a crime scene when they were not[13]. Similarly, voice-cloning tools can produce audio clips of individuals saying things they never actually said[14].

In criminal trials, where the stakes involve imprisonment or even capital punishment, such manipulated evidence could lead to wrongful convictions or acquittals. The possibility that crucial digital content might be a sophisticated fake forces courts and investigators to second-guess even seemingly clear-cut visual proof.

3.2 Challenges in Authenticating Digital Proof

Authenticating digital evidence has always required technical scrutiny, but deepfakes raise the bar significantly. Traditional forensic methods—such as analysing metadata, compression patterns, or file

[10] Mirsky, Y. and Lee, W., "The Creation and Detection of Deepfakes: A Survey", *ACM Computing Surveys*, Vol. 54, No. 1, 2021, Article 1.
[11] Agarwal, A., "Deepfakes and Indian Law: An Urgent Need for Regulation", *Indian Journal of Law and Technology*, Vol. 17, 2021, pp. 45–60.
[12] West, Darrell M., "How to Combat Fake News and Disinformation", *Brookings Institution Report*, 18 Dec. 2017.
[13] Chesney, R. and Citron, D. K., "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, Vol. 107, 2019, pp. 1753–1819
[14] Mirsky, Y. and Lee, W., "The Creation and Detection of Deepfakes: A Survey", *ACM Computing Surveys*, Vol. 54, No. 1, 2021, Article 1.

formats—are often insufficient against well-crafted AI-generated content[15]. Moreover, many courts lack the technological infrastructure or trained personnel to distinguish authentic files from fakes.

Deepfakes also exploit gaps in current Indian evidentiary law. While the Indian Evidence Act, 1872 permits the use of electronic records, it does not specifically account for synthetic media manipulation[16]Section 65B, which governs the admissibility of electronic records, assumes the reliability of digital material under a certification process. But in the case of deepfakes, the source of the material itself may be fabricated, making the certificate unreliable if forged or obtained fraudulently[17].

The recently enacted Bhartiya Sakshya Adhiniyam, 2023, which replaces the Indian Evidence Act, continues to allow the admissibility of electronic evidence under Section 63, but like its predecessor, it still relies on procedural certification without addressing the issue of synthetic content authenticity in depth[18]. The Act has updated the language around digital records to accommodate modern formats, yet it does not specifically mandate forensic verification or AI-based authentication protocols necessary to detect deepfakes. As a result, the law remains unequipped to confront the evolving technological risks posed by deepfakes, especially in legal scenarios involving high-stakes evidence.

3.3 Impact on Investigations and Court Proceedings

The impact of deepfakes is not limited to the courtroom. Investigative agencies face considerable hurdles in verifying digital leads, especially in time-sensitive or high-profile cases. A fabricated clip released on social media can trigger mass panic, disrupt investigations, or pressure law enforcement into wrongful action[19]. Public perception is also shaped by such content, influencing media narratives and sometimes even judicial interpretations in highly mediatized cases.

Furthermore, the legal process itself can be derailed if a party claims that genuine evidence is a deepfake, a tactic known as the "liar's dividend"—where the mere existence of deepfakes casts doubt on real evidence[20]. This undermines faith in legal institutions and creates an environment where digital deception can thrive unchecked.

4. Deepfakes and Reputational Damage

While the legal system grapples with deepfakes as threats to evidence, their impact on individual dignity, public discourse, and trust in institutions is even more pervasive. Deepfakes are increasingly being used not just to deceive the courts but to intentionally harm reputations—politically, socially, and personally. With content being shared rapidly across social media platforms, victims often face immediate backlash long before any authentication can take place.

4.1 Political, Social, and Personal Implications

Deepfakes have become powerful tools for political sabotage and disinformation. Fake videos of political leaders making controversial statements, or behaving inappropriately, have been used to manipulate public

[15] Jain, A., "Admissibility of Electronic Evidence in India: Legal Provisions and Judicial Trends", *NALSAR Law Review*, Vol. 13, 2020, pp. 55–72.
[16] Indian Evidence Act, 1872 (India).
[17] Mathur, P., "Revisiting Section 65B of the Indian Evidence Act in the Era of Deepfakes", *Indian Journal of Law and Technology*, Vol. 18, 2022, pp. 101–118.
[18] Bhartiya Sakshya Adhiniyam, 2023 (India), Sec. 63.
[19] Paris, B. and Donovan, J., "Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence", *Data & Society Report*, 2019.
[20] West, Darrell M., "How to Combat Fake News and Disinformation", *Brookings Institution Report*, 18 Dec. 2017.

opinion during elections[21]. In India, several instances have surfaced where doctored videos of political figures were circulated to incite communal tensions or discredit rival parties. These fabrications, though later debunked, had already influenced public narratives and media cycles[22].

Socially, deepfakes have been used to tarnish the image of journalists, activists, and celebrities. Particularly concerning is the targeting of women through non-consensual deepfake pornography, where their faces are superimposed onto explicit material—leading to humiliation, defamation, and in some cases, withdrawal from public life[23].

4.2 Case Studies of Victims of Deepfake Harassment

One of the earliest high-profile cases involved Hollywood actresses whose faces were inserted into pornographic videos and circulated on adult websites. In India, a deepfake video of a regional actress went viral in 2020, despite her public denial and the video's proven inauthenticity. The damage, however, was already done—resulting in emotional distress, trolling, and loss of professional opportunities.

Another widely cited case involved a political party circulating a deepfake video of a national leader delivering a speech in a different language, supposedly tailored to local voters. Though it was later flagged as manipulated, the clip had already reached millions and influenced voter sentiment.

4.3 Psychological and Social Impact

Victims of deepfake harassment report severe psychological effects, including anxiety, depression, loss of confidence, and a persistent fear of being watched or misrepresented[24]. Since the internet never truly forgets, deepfake content—even if taken down—continues to exist on obscure platforms or reappears in future attacks. The social stigma attached to such content often leads to victim-blaming, especially in conservative societies.

The broader societal impact includes erosion of public trust—not only in media but also in institutions. As deepfakes become more sophisticated, people begin to doubt even real videos, creating a climate of suspicion known as the "liar's dividend," where wrongdoers dismiss authentic evidence as fake[25]. This growing scepticism threatens democratic dialogue, journalistic integrity, and the rule of law.

5. Legal Accountability and Existing Frameworks

As deepfakes evolve from isolated digital pranks to tools of serious harm, the need for legal regulation and accountability becomes increasingly urgent. While the technology itself is advancing rapidly, the legal framework—both in India and globally—has been slower to adapt. Existing laws offer some relief under broader provisions related to cybercrime, defamation, and data protection, but none are specifically tailored to address the sophisticated and deceptive nature of deepfakes.

5.1 Current Indian Laws: IT Act, IPC, and Evidence Framework

In India, the Information Technology Act, 2000 serves as the primary legislation addressing cyber offences. Section 66 and Section 66E penalise identity theft and violation of privacy, respectively, while

[21] Chesney, R. and Citron, D. K., "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, Vol. 107, 2019, pp. 1753–1819.
[22] Jain, R., "Political Deepfakes in India: Between Misinformation and Electoral Manipulation", *Economic and Political Weekly*, Vol. 55, No. 42, 2020, pp. 10–13.
[23] Agarwal, A., "Deepfakes and Indian Law: An Urgent Need for Regulation", *Indian Journal of Law and Technology*, Vol. 17, 2021, pp. 45–60.
[24] West, Darrell M., "How to Combat Fake News and Disinformation", *Brookings Institution Report*, 18 Dec. 2017.
[25] Paris, B. and Donovan, J., "Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence", *Data & Society Report*, 2019.

Section 67 targets the publication or transmission of obscene material online[1]. However, these provisions were drafted before the rise of AI-generated synthetic media and do not explicitly mention deepfakes.

The Bhartiya Nyaya Sanhita, 2023 (BNS), which replaces the Indian Penal Code, 1860, includes provisions that may be invoked in cases involving malicious use of deepfakes. Section 336 penalises acts of forgery intended to harm reputation, Section 356 addresses criminal defamation, and Section 358 deals with the circulation of false or offensive statements that could incite public disorder or enmity[26]. These sections provide a legal foundation to address reputational and social harm caused by synthetic media. However, enforcement remains a significant challenge due to the anonymity of content creators, the cross-border nature of online platforms, and the technological difficulty in tracing or verifying the origin of deepfake content. Without robust forensic and investigative mechanisms, these provisions—though legally sound—often fall short in real-world application.

Under the Indian Evidence Act, 1872, and now the Bhartiya Sakshya Adhiniyam, 2023, electronic records are admissible in court under Sections 65B and 63 respectively[27]. However, these laws still rely on digital certifications and do not require mandatory forensic testing to confirm the authenticity of the content—leaving room for manipulated evidence to slip through.

5.2 International Legal Approaches

Globally, legal systems are also struggling to keep up. The United States has enacted some state-level laws targeting deepfakes—like California's legislation banning deepfakes related to elections and non-consensual pornography[28]. The DEEPFAKES Accountability Act, introduced in the U.S. Congress, proposes watermarking deepfake content and penalising creators who intend to cause harm[29].

In the European Union, the Digital Services Act (DSA) and the General Data Protection Regulation (GDPR) provide broader protection by placing obligations on online platforms to detect and remove harmful content, including synthetic media. However, most international laws still focus more on platform accountability than on directly penalising the creators of deepfakes.

5.3 Gaps and Challenges in Regulation

Despite these frameworks, significant legal gaps remain. Indian laws are reactive rather than preventive and lack provisions for AI-generated content specifically. The absence of a unified national deepfake law creates inconsistencies in enforcement and hampers legal recourse for victims.

Moreover, regulatory bodies often lack technical capacity to detect or investigate deepfakes in a timely manner. The cross-border nature of digital content also complicates jurisdiction, especially when the content is hosted or circulated from servers outside India. Additionally, the balance between regulation and free speech—a cornerstone of democratic societies—adds further complexity.

Given these challenges, India and other jurisdictions must move toward specialised legal provisions, AI-aided detection frameworks, and international cooperation mechanisms. A proactive, well-defined law that addresses the unique threats of deepfakes—without stifling innovation—is the need of the hour.

Conclusion

The rapid advancement of deepfake technology has presented a multidimensional challenge to law, ethics, and society. Deepfakes threaten the credibility of digital evidence, compromise the reputation and dignity

[26] Bhartiya Nyaya Sanhita, 2023 (India), Ss. 336, 356, 358.
[27] Bhartiya Sakshya Adhiniyam, 2023 (India), S. 63.
[28] California Assembly Bill No. 730, 2019 (USA).
[29] Deepfakes Accountability Act, H.R. 3230, 116th Cong. (2019) (USA).

of individuals, and challenge the integrity of legal proceedings. Their misuse ranges from political manipulation and character assassination to the fabrication of legal evidence, blurring the line between truth and falsity in ways that were unimaginable just a few years ago.

While India has legal provisions under the Information Technology Act, 2000, the Bhartiya Nyaya Sanhita, 2023, and the Bhartiya Sakshya Adhiniyam, 2023, these are either outdated in scope or insufficiently tailored to address the unique threats posed by AI-generated synthetic media[30]. Furthermore, the challenges of anonymity, cross-border content dissemination, and lack of technological preparedness continue to hamper enforcement and victim protection efforts.

[30] Bhartiya Nyaya Sanhita, 2023 (India), Ss. 336, 356, 358.