FROM DEEPFAKES TO DIGITAL DECEPTION: RETHINKING CRIMINAL LIABILITY, EVIDENTIARY RELIABILITY AND THE REGULATION OF AI-GENERATED SYNTHETIC MEDIA IN INDIA

Dr. Dhananjay Madheshia, Deen Dayal Upadhyaya Gorakhpur University Gorakhpur / Buddha Law College

Diwakar Prasad Dwivedi Deen Dayal Upadhyaya Gorakhpur University Gorakhpur / Buddha Law College

ABSTRACT

The rapid development of generative artificial intelligence has transformed the production and dissemination of digital content. Deepfakes and other forms of AI-generated synthetic media can convincingly manipulate images, videos, voices and textual communications, thereby creating serious challenges for criminal law, evidence and regulatory governance. In India, the legal response to synthetic media remains distributed across the Information Technology Act, 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, the Bharatiya Nyaya Sanhita, 2023, and the Bharatiya Sakshya Adhiniyam, 2023. The problem is not merely that synthetic media may facilitate traditional offences such as cheating, impersonation, forgery, obscenity, defamation and identity theft; it also threatens the reliability of digital evidence itself. The Bharatiya Sakshya Adhiniyam expressly recognises electronic and digital records, but the increasing sophistication of synthetic media requires courts to adopt stronger approaches to authentication, provenance, metadata, expert examination and chain of custody. This paper examines the adequacy of India's existing framework and argues for a technology-neutral but rights-sensitive regulatory model based on provenance, mandatory disclosure, platform accountability, specialised forensic capacity and calibrated criminal liability. It concludes that regulation should distinguish between lawful synthetic expression and malicious digital deception rather than treating AI-generated content as inherently unlawful.

Keywords: Deepfakes, Artificial Intelligence, Synthetic Media, Criminal Liability, Digital Evidence, Bharatiya Sakshya Adhiniyam, Information Technology Act, Cybercrime, Intermediary Liability.

INTRODUCTION

Artificial intelligence has fundamentally altered the manner in which digital information is created, reproduced and consumed. Generative AI systems can now produce realistic photographs, videos, voices and other audiovisual material with limited technical expertise. Deepfakes represent one particularly significant manifestation of this development. They employ machine-learning techniques to replace, imitate or synthetically generate the identity, appearance or voice of an individual. Although the technology has legitimate applications in entertainment, education, accessibility and artistic expression, its malicious use creates substantial legal and social risks.

The danger lies not simply in the falsity of manipulated content but in its capacity to appear authentic. A fabricated video depicting a public figure making a statement, a synthetic voice authorising a financial transaction, or an altered intimate image can cause reputational, financial and psychological harm before its falsity is established. In electoral and public-policy contexts, synthetic media can also undermine democratic discourse by making it increasingly difficult to distinguish genuine communications from fabricated ones.

India's legal system does not presently possess a single comprehensive statute exclusively regulating deepfakes. Instead, liability may arise under several existing laws depending upon the nature, intention and consequences of the conduct. The Information Technology Act, 2000 contains provisions concerning identity theft, cheating by personation, privacy violations and sexually explicit electronic material.[1] The Information Technology Rules additionally impose obligations upon intermediaries concerning unlawful information and synthetic content. The regulatory framework has also evolved through governmental advisories addressing deepfakes and AI-generated misinformation.[2]

The evidentiary dimension presents an equally important problem. The Bharatiya Sakshya Adhiniyam, 2023 recognises electronic and digital records and provides a statutory framework for their admissibility.Yet the same technological development that expands the availability of

[1] The Information Technology Act, 2000, ss. 66C, 66D, 66E, 67, 67A, 67B, 67C and 79, Government of India, India Code

[2] Ministry of Electronics and Information Technology, Government of India, *Advisory on Deepfakes/Synthetic Content and Due Diligence by Intermediaries*, 26 December 2023. The advisory required intermediaries to comply with applicable provisions of the Information Technology Act and Rules and addressed synthetic creation, generation and modification of conten

digital evidence also makes visual and audio material increasingly susceptible to manipulation. Consequently, the traditional assumption that a recording is a reliable representation of an event can no longer be accepted without adequate authentication.

DEEPFAKES AND THE TRANSFORMATION OF DIGITAL DECEPTION

Deepfakes represent one of the most consequential applications of artificial intelligence in the contemporary digital environment. They form part of the broader category of synthetic media, encompassing photographs, videos, audio recordings and other digital materials that are generated or substantially modified through artificial intelligence and machine-learning technologies. The principal legal difficulty does not arise from the existence or use of artificial intelligence itself, but from its capacity to create highly convincing representations of persons, events and statements that may never have existed in reality. Consequently, the law must increasingly distinguish between legitimate technological creativity and the deliberate use of synthetic media to deceive, manipulate or cause harm.

Deepfakes are particularly problematic because of three interconnected characteristics. First, they possess an unprecedented ability to reproduce human appearance, facial expressions and voice with considerable realism. A synthetic video may therefore depict an individual apparently making a statement or engaging in conduct that never occurred. Secondly, deepfakes can be generated and disseminated rapidly and inexpensively, enabling a single fabricated recording to reach millions of users within a short period. Thirdly, once uploaded to social-media platforms, removal of the original material may not prevent its continued circulation through downloads, screenshots, reposts, edited versions and other derivative content.

Traditional forgery generally involved the alteration or fabrication of physical documents, signatures, photographs or identifiable electronic files. Synthetic media fundamentally complicates this model because falsification can occur at the level of human perception. An individual may see a video that appears authentic, hear a familiar voice or observe apparently genuine facial expressions without having any immediate reason to suspect manipulation. Thus, deepfakes challenge the conventional evidentiary assumption that audiovisual material provides a relatively direct representation of reality.

This phenomenon may appropriately be described as “digital deception by simulated

presence." The victim is not simply presented with inaccurate information. Instead, artificial intelligence creates a simulated presence of a real person or event, giving the fabricated material the persuasive authority normally associated with direct observation. This is particularly dangerous where recipients already trust the person being impersonated. A synthetic video of a political leader, for example, may influence public opinion; an artificial voice resembling a business executive may be used to induce an employee to transfer money; and an altered intimate image may cause severe reputational and psychological injury.

CRIMINAL LIABILITY UNDER INDIAN LAW

Indian law can address several forms of malicious deepfake activity through existing criminal and cyber-law provisions, although significant conceptual and practical limitations remain. The Information Technology Act, 2000 contains provisions that may become applicable where deepfake technology is employed for identity theft, personation, privacy violations or unlawful publication of sexually explicit material. Section 66C addresses identity theft, while Section 66D deals with cheating by personation through a computer resource or communication device.[3] These provisions are particularly relevant where an offender uses another person's identity or identifying characteristics to deceive a victim for an unlawful purpose.

A deepfake may therefore become the technological instrument through which a conventional offence is committed. For instance, where an AI-generated voice impersonates an individual and induces another person to transfer money, the conduct may potentially fall within the framework of cheating by personation. Similarly, the unauthorised creation or dissemination of sexually explicit synthetic material may attract liability under the applicable provisions of the Information Technology Act.[4]

The Bharatiya Nyaya Sanhita, 2023 (BNS) also provides a broader criminal-law framework relevant to deepfake-enabled conduct. Depending upon the facts, offences concerning cheating, forgery, criminal intimidation, defamation and other forms of deception may become applicable.[3] The significance of the BNS is therefore not that it establishes a completely separate category of "deepfake crime," but that existing offences can potentially be applied when artificial intelligence is used as a means of committing unlawful conduct.

[3] The Information Technology Act, 2000, ss. 66C–66D, Government of India, India Code
[4] The Information Technology Act, 2000, ss. 66E, 67 and 67A, Government of India, India Code

Nevertheless, reliance exclusively upon traditional offences presents difficulties. Existing criminal provisions are primarily designed around particular forms of conduct, intention and harm. A deepfake can produce substantial injury even when the conventional ingredients of cheating or forgery cannot easily be established. A fabricated video may seriously damage a person's reputation without causing measurable financial loss. Similarly, an AI-generated political statement may distort democratic discourse without necessarily satisfying the elements of a conventional criminal offence.

Accordingly, criminal liability should be based upon the nature and purpose of the synthetic content rather than merely upon the fact that AI was used. Relevant considerations should include intention, knowledge, deception, material falsity, absence of consent, foreseeable harm and the legal interest affected. This approach would prevent legitimate creative uses of AI from being criminalised while ensuring that malicious synthetic deception is appropriately punished.[5]

INTERMEDIARY RESPONSIBILITY AND PLATFORM REGULATION

Social-media intermediaries occupy a central position in the dissemination of deepfakes. Their algorithms can amplify content through recommendations, trends and automated distribution, thereby transforming an isolated fabrication into a mass-distributed digital phenomenon. Consequently, legal responsibility cannot be examined solely from the perspective of the original creator. It must also consider the obligations of platforms that host, distribute, recommend or monetise synthetic media.

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 establish due-diligence obligations for intermediaries. India's regulatory approach has progressively responded to the growing threat posed by AI-generated and manipulated content. In December 2023, the Ministry of Electronics and Information Technology issued an advisory directing intermediaries to comply with existing legal obligations concerning misinformation and deepfakes and to take appropriate measures regarding synthetically generated or manipulated information.[6]

[5] Ministry of Electronics and Information Technology, Government of India, *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021*, as amended, provisions relating to synthetically generated information.
[6] The Bharatiya Nyaya Sanhita, 2023, Act No. 45 of 2023, Government of India, India Code

The subsequent development of the regulatory framework has increasingly focused on synthetically generated information (SGI). The amended IT Rules place greater emphasis on identifying and regulating synthetically generated or modified information, thereby moving beyond a purely reactive model based upon post-publication removal.[7] This approach recognises that the architecture through which synthetic content is created and distributed can itself contribute to the scale of digital deception.

However, intermediary regulation must remain constitutionally balanced. Article 19(1)(a) of the Constitution of India protects freedom of speech and expression, subject to reasonable restrictions under Article 19(2).[8] Not every AI-generated image, video or audio recording is harmful. Satire, parody, artistic expression, political criticism, education and entertainment may legitimately employ synthetic media. A blanket prohibition on AI-generated content could therefore have a chilling effect on lawful expression.

A more appropriate approach is risk-based regulation. Synthetic content impersonating real persons, facilitating fraud, depicting individuals in non-consensual sexual contexts or creating significant risks to public safety should attract stronger regulatory obligations. Conversely, clearly fictional, artistic or satirical content should generally remain permissible, particularly where appropriate disclosure or labelling makes its synthetic character apparent.

The ultimate objective should therefore be to regulate deceptive and harmful uses of synthetic media rather than artificial intelligence as such. Indian law must preserve technological innovation while ensuring accountability for manipulation, impersonation and fraud. Effective regulation should combine criminal liability, intermediary due diligence, transparency, provenance mechanisms, rapid grievance redressal and digital literacy. Such a balanced framework can address the transformative threat of deepfakes without unnecessarily restricting legitimate technological and expressive freedom.

EVIDENTIARY RELIABILITY AND THE BHARATIYA SAKSHYA ADHINIYAM

The evidentiary implications of deepfakes may ultimately be more significant than the question of substantive criminal liability. The Bharatiya Sakshya Adhiniyam, 2023 expressly recognises

[7] Ministry of Electronics and Information Technology, Government of India, *Advisory on Deepfakes/Synthetic Content and Due Diligence by Intermediaries*, 26 December 2023.
[8] INDIA CONST. art. 19(1)(a) & art. 19(2).

electronic or digital records. Section 61 provides that an electronic or digital record cannot be denied legal effect merely because it is electronic or digital, subject to the statutory requirements. Sections 62 and 63 provide the framework concerning proof and admissibility of electronic records.[9]

This legal recognition is important because modern criminal investigations increasingly depend upon CCTV recordings, mobile-phone data, photographs, recordings, social-media communications and cloud-based information.

However, admissibility is not equivalent to authenticity or reliability. A digital file may technically satisfy the requirements for admission while the substantive content represented by that file may nevertheless be manipulated. The emergence of deepfakes therefore requires courts to distinguish between:

1. whether a digital record exists;
2. whether the record has been properly obtained;
3. whether the record has remained unaltered;
4. whether the device or system generating it can be identified;
5. whether the content accurately represents the event it purports to depict; and
6. whether AI manipulation has occurred.

The traditional forensic concept of chain of custody consequently assumes greater importance. Investigators should preserve original files, metadata, hash values, device information and acquisition records wherever possible.

Expert evidence will also become increasingly important. The court may need assistance from forensic specialists capable of examining compression patterns, metadata, audio-wave characteristics, facial inconsistencies, file provenance and other indicators of manipulation. Yet AI-detection tools themselves are not infallible. A detector trained against one generation of

[9] The Bharatiya Sakshya Adhiniyam, 2023, ss. 61–63. Section 63 specifically establishes conditions concerning the admissibility of information contained in electronic records

synthetic media may perform poorly against newer models.[10]

Accordingly, the court should avoid treating an AI detector's output as conclusive proof. Detection should instead be considered one component of a broader evidentiary assessment.

THE PROBLEM OF THE "LIAR'S DIVIDEND"

An especially serious consequence of deepfake technology is the possibility that genuine evidence may be dismissed as fabricated. Once society becomes aware that realistic synthetic media can be generated, an accused person may claim that authentic recordings are deepfakes.

This phenomenon is sometimes described as the "liar's dividend." The existence of deepfakes therefore creates a double danger: false evidence may be accepted as genuine, while genuine evidence may be rejected as false.

Indian courts must therefore develop a sophisticated authenticity methodology. Instead of relying exclusively upon visual appearance, courts should evaluate corroborating evidence, source provenance, metadata, witness testimony, device records, contemporaneous communications and forensic examination.

The Bharatiya Sakshya Adhiniyam's recognition of electronic records provides the statutory foundation, but judicial practice must evolve alongside technology. The objective should not be to make courts technically dependent upon AI-detection software; rather, courts should develop a multidisciplinary approach combining law, digital forensics and technological expertise.

NEED FOR A COMPREHENSIVE REGULATORY FRAMEWORK

India requires a coherent framework specifically addressing synthetic media while avoiding excessive criminalisation. Such a framework should contain at least five components.

First, provenance and labelling: AI-generated or materially AI-modified content should, wherever technically feasible, contain reliable provenance information or machine-readable indicators. India's regulatory approach has already moved in this direction.[9]

[10] The evidentiary challenges created by generative AI under the Bharatiya Sakshya Adhiniyam, 2023.

Secondly, differentiated criminal liability: Mere creation of synthetic content should not constitute an offence. Criminal liability should arise where synthetic media is intentionally or knowingly used to commit fraud, impersonation, sexual exploitation, intimidation, reputational harm or other legally recognised wrongdoing.

Thirdly, platform accountability: Intermediaries should maintain effective reporting, verification and rapid-response mechanisms for high-risk synthetic content while preserving procedural safeguards against arbitrary removal.[11]

Fourthly, forensic capacity: India needs specialised digital-forensic laboratories and trained investigators capable of examining AI-generated media. Law-enforcement personnel, prosecutors and judges should receive continuing technological training.

Fifthly, protection of individual identity: Indian law should strengthen protections concerning a person's image, voice, likeness and biometric characteristics, particularly where AI is used without consent for commercial exploitation, fraud or sexualised manipulation.

CONCLUSION

Deepfakes require Indian criminal law and evidence law to reconsider the relationship between technology, truth and legal responsibility. The central difficulty is not that existing law is entirely incapable of addressing synthetic media. Rather, existing legal provisions were largely designed for a technological environment in which digital records were assumed to be relatively stable and manipulation was comparatively difficult.

The Information Technology Act, Bharatiya Nyaya Sanhita and Bharatiya Sakshya Adhiniyam collectively provide important foundations for responding to AI-enabled deception. The statutory recognition of electronic evidence and the developing intermediary framework demonstrate that Indian law is capable of adapting to technological change. Nevertheless, deepfakes expose gaps concerning authenticity, attribution, provenance, consent and platform responsibility.

The future regulatory model should therefore avoid both technological exceptionalism and excessive criminalisation. AI-generated media should not be treated as inherently unlawful.

[11] Chesney, R., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), 1753-1819.

Instead, law should focus on deception, intent, harm, consent and accountability. Courts should adopt stronger standards for authenticating audiovisual evidence, while platforms should implement meaningful provenance and disclosure mechanisms. At the same time, legitimate artistic, satirical, educational and creative uses of generative AI must remain protected.

Ultimately, the challenge posed by deepfakes is not simply a challenge to cyber law. It is a challenge to the legal system's ability to determine what is real, who is responsible and what evidence can be trusted. India's response must therefore combine substantive criminal law, procedural safeguards, digital forensics, technological regulation and constitutional rights. Only such an integrated approach can ensure that technological innovation does not become a tool for digital deception without accountability.

BIBLIOGRAPHY

1. Agarwal, S., & Singh, P. (2022). Artificial Intelligence and the Law: Addressing the Challenges of Synthetic Media. International Journal of Law and Information Technology, 30(1), 1-25.

2. Chesney, R., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107(6), 1753-1819.

3. Choudhury, S., & Das, S. (2023). Regulatory Approaches to Deepfake Technology in India: A Comparative Study. Indian Journal of Law and Policy, 9(2), 89-104.

4. Dutta, S. (2021). Regulation of AI-Generated Media in Emerging Economies: Focus on India. Technology and Law Review, 12(3), 78-94.

5. Farid, H. (2019). Photo Forensics: The Science of Detecting Digital Forgeries. IEEE Signal Processing Magazine, 26(2), 16-25.

6. Franks, B. (2021). The Ethical Implications of Synthetic Media. AI & Society, 36, 543-556.

7. Khandelwal, S., & Kumar, R. (2021). Legal Challenges and Regulatory Framework for Deepfake Technology in India. Indian Journal of Law and Technology, 17, 45-67.

8. Kietzmann, J., & Canhoto, A. (2013). Bittersweet! Understanding and Managing Electronic Word of Mouth. Journal of Public Affairs, 13(2), 146-159.

9. Kshetri, N. (2020). The Rise of Deepfake Technology and Its Impact on Society. IEEE Computer, 53(2), 88-92.

10. Kumar, P., & Sharma, V. (2023). Rethinking Legal Frameworks for Synthetic Media in India. Asian Journal of Comparative Law, 18(1), 67-85.

11. Ng, V. (2020). Deepfakes and the Law: New Challenges for Evidence and Liability. Cybersecurity and Law Journal, 4(3), 112-125.

12. Rini, R. (2020). Deepfakes and the Philosophy of Trust. Philosophy & Technology, 33, 547–560.

13. Singh, R., & Gupta, A. (2022). Criminal Liability for Deepfakes in India: Legal Perspectives. Indian Journal of Criminal Law, 15(1), 23-45.

14. Vincent, B., & Wiles, J. (2020). Evidence Reliability in the Age of Deepfakes. Journal of Forensic Sciences, 65(4), 1123-1131.

15. Zhang, Y., & Chen, H. (2021). Detecting Deepfakes: Techniques and Challenges. IEEE Transactions on Information Forensics and Security, 16, 2454-2468.