DEEPFAKE TECHNOLOGY AND CRIMINAL LIABILITY: AN ANALYSIS UNDER THE BHARATIYA NYAYA SANHITA, 2023

Dr. Komila Aggarwal, Assistant Professor and Programme In-charge,
Dogra Law College, Jammu.

ABSTRACT

The rapid advancement of artificial intelligence has led to the emergence of deepfake technology, enabling the creation of highly realistic yet fabricated audio, video, and image content that poses significant challenges to criminal justice systems worldwide. In India, deepfakes have become powerful tools for identity theft, financial fraud, cyber extortion, electoral manipulation, defamation, and the dissemination of non-consensual intimate content, thereby threatening individual dignity, public trust, and democratic governance. Although the Bharatiya Nyaya Sanhita, 2023 (BNS) represents a significant reform in India's criminal law framework, it does not expressly define or criminalise the malicious creation and dissemination of synthetic media, resulting in notable legislative and enforcement gaps. This paper critically examines the applicability of existing provisions of the Bharatiya Nyaya Sanhita, 2023, to deepfake-related offences, particularly those concerning forgery, cheating, personation, defamation, and other technology-enabled crimes. It further identifies doctrinal ambiguities, evidentiary concerns, and practical challenges faced by investigative agencies and courts in addressing offences involving AI-generated content. The study argues that the absence of a dedicated legal framework limits effective victim protection, creates uncertainty in prosecution, and undermines the deterrent value of criminal law. It advocates the introduction of a comprehensive statutory framework incorporating a clear definition of deepfake offences, graded penalties based on the nature and severity of harm, standardized digital forensic protocols, and expedited procedural safeguards for victim redressal. Adopting a doctrinal and analytical approach, the paper concludes that a technology-responsive legal framework under the Bharatiya Nyaya Sanhita, 2023, is essential to address emerging forms of digital deception, ensure accountability, strengthen the criminal justice system, and uphold the rule of law in India's evolving digital landscape.

Keywords: Deepfake Technology, Synthetic Media; Bharatiya Nyaya Sanhita, 2023, Artificial Intelligence, Criminal Law Reform, Digital Forensics, Cybercrime; Digital Evidence.

1. Introduction

The rapid advancement of artificial intelligence (AI) has transformed digital communication by enabling the creation of sophisticated synthetic media capable of replicating human speech, facial expressions, and mannerisms with remarkable accuracy. Among these developments, deepfake technology has emerged as one of the most significant challenges to contemporary criminal law. Using deep learning algorithms, particularly Generative Adversarial Networks (GANs), deepfakes produce highly realistic but fabricated audio, video, and image content that is often indistinguishable from authentic recordings. Although the technology has legitimate applications in entertainment, education, healthcare, and digital innovation, its malicious use has created serious legal and societal concerns. In recent years, deepfake technology has increasingly been employed for criminal purposes, including identity theft, financial fraud, cyber extortion, impersonation, election-related misinformation, defamation, and the creation of non-consensual intimate images. Such misuse not only causes substantial economic and reputational harm but also undermines public confidence in digital evidence and online communication. The widespread use of social media and digital platforms in India has further accelerated the dissemination of manipulated content, making the detection and prosecution of deepfake-related offences increasingly complex.

The enactment of the Bharatiya Nyaya Sanhita, 2023 (BNS), replacing the Indian Penal Code, 1860, represents a significant reform in India's substantive criminal law. The legislation seeks to modernize criminal law by addressing emerging forms of crime and strengthening the administration of justice. However, despite the growing prevalence of AI-enabled offences, the BNS does not expressly define or criminalize deepfake technology or synthetic media. Consequently, offences involving deepfakes must presently be addressed through existing provisions relating to forgery, cheating, personation, defamation, fabrication of electronic records, and other technology-enabled offences. The absence of a dedicated statutory framework gives rise to interpretational uncertainties regarding the scope of criminal liability, evidentiary standards, and the adequacy of existing punishments. The increasing use of deepfakes also presents significant challenges to criminal investigations and judicial proceedings. Digital audio and video recordings have traditionally been regarded as reliable forms of evidence. However, advances in AI-generated synthetic media have made it increasingly difficult to distinguish genuine recordings from manipulated content. This raises concerns regarding the admissibility, authenticity, and evidentiary value of electronic evidence,

while simultaneously enabling offenders to evade liability by questioning the authenticity of genuine digital evidence. These developments place additional responsibilities on investigative agencies, forensic experts, and courts in evaluating AI-generated content[1].

The misuse of deepfake technology also implicates fundamental constitutional rights, particularly the rights to privacy, dignity, reputation, and personal liberty guaranteed under Article 21 of the Constitution of India. Victims of malicious deepfakes often suffer irreversible reputational damage, psychological trauma, financial loss, and social ostracism. The existing legal framework therefore requires careful examination to determine whether it adequately protects victims and effectively addresses emerging forms of AI-enabled criminality. Against this backdrop, this paper critically analyses the concept of deepfake technology and examines the extent to which the provisions of the Bharatiya Nyaya Sanhita, 2023 can be applied to offences involving AI-generated synthetic media. It evaluates the adequacy of the existing statutory framework in addressing criminal liability, identifies legislative and enforcement gaps, and proposes legal reforms to ensure effective prosecution, victim protection, and technological accountability. The study argues that the evolving nature of AI-generated crimes necessitates a clear and comprehensive legal framework capable of responding to the challenges posed by deepfake technology while preserving the integrity of India's criminal justice system.

2. Understanding Deepfake Technology: Origin, Mechanism, and Criminal Misuse

2.1. Origin and Evolution of Deepfake Technology

Deepfake technology represents one of the most significant advancements in artificial intelligence (AI), enabling the creation of highly realistic yet fabricated digital content. The term *deepfake* is derived from the words "deep learning" and "fake," referring to synthetic media generated using advanced machine learning algorithms. Unlike conventional photo or video editing, which relies primarily on manual manipulation of digital content, deepfake technology employs AI models capable of learning and replicating facial expressions, voice patterns, body movements, and other human characteristics with remarkable precision. Consequently, the resulting audio, video, or images often appear authentic, making it

[1] Shubham Pandey and Gaurav Jadhav, 'Emerging Technologies and Law: Legal Status of Tackling Crimes Relating to Deepfakes in India' *SCC Times* (17 March 2023) https://www.scconline.com/blog/post/2023/03/17/emerging-technologies-and-law-legal-status-of-tackling-crimes-relating-to-deepfakes-in-india/ accessed 18 July 2026.

increasingly difficult to distinguish fabricated content from genuine recordings[2]. Initially developed as a research application within the field of artificial intelligence, deepfake technology has evolved rapidly due to advances in computational power, machine learning techniques, and the widespread availability of large datasets. The release of open-source software and commercially available AI applications has significantly reduced the technical expertise required to create synthetic media, allowing individuals with limited technological knowledge to generate convincing deepfakes.

2.2 Technological Mechanism of Deepfakes

Deepfake technology primarily relies on deep neural networks, a subset of machine learning designed to imitate the functioning of the human brain by processing large volumes of data and identifying complex patterns. Among the most widely used techniques are Generative Adversarial Networks (GANs), introduced by Goodfellow et al. (2014). A GAN consists of two interconnected neural networks: the *generator*, which creates synthetic content, and the *discriminator*, which evaluates whether the generated output resembles authentic data. Through repeated training and continuous feedback, the generator gradually improves its ability to produce realistic images, videos, or audio recordings capable of deceiving both human observers and automated verification systems. Apart from GANs, modern deepfake systems also employ autoencoders, facial mapping, face-swapping, voice cloning, and lip synchronisation technologies. Autoencoders analyse facial characteristics and reconstruct digital images, while face-swapping algorithms replace one person's face with another in existing videos. Voice-cloning software replicates an individual's tone, accent, pronunciation, and speaking style using only a limited audio sample. These technological developments have significantly enhanced the realism and accessibility of synthetic media.

2.3 Accessibility and Democratization of Deepfake Technology

One of the most concerning developments in recent years has been the increasing accessibility of deepfake technology. Earlier, creating AI-generated synthetic media required specialised knowledge of artificial intelligence, advanced programming skills, and expensive computing infrastructure. Today, however, open-source applications such as DeepFaceLab, FaceSwap, and several consumer-grade mobile applications have simplified the creation of deepfakes.

[2] Pavel Korshunov and Sébastien Marcel, 'DeepFakes: A New Threat to Face Recognition? Assessment and Detection' (2018) *arXiv* https://arxiv.org/abs/1812.08685 accessed 17 July 2026.

Cloud computing services and user-friendly AI platforms have further lowered technical barriers, enabling almost anyone with a smartphone or personal computer to generate manipulated digital content[3]. The democratization of deepfake technology has transformed it from a specialised research tool into a widely available technology with significant legal implications. While increased accessibility promotes innovation and creativity, it also expands opportunities for misuse by cybercriminals, fraudsters, and other malicious actors.

2.4 Criminal Misuse of Deepfake Technology

The misuse of deepfake technology has emerged as a serious challenge for criminal justice systems worldwide. Among the most prevalent forms of abuse is the creation of non-consensual intimate imagery, where a person's face is digitally superimposed onto explicit content without consent. Such synthetic images and videos frequently target women, causing severe psychological trauma, reputational harm, and violations of privacy and dignity. The permanent nature of online dissemination further intensifies the injury, as complete removal of such content from digital platforms is often impossible.

Deepfakes are also increasingly employed to facilitate financial fraud and identity-based deception. Artificial intelligence enables offenders to clone voices, fabricate video calls, and impersonate corporate executives, family members, or public officials to obtain money or confidential information. Internationally reported incidents involving AI-generated voice cloning and fraudulent video conferences demonstrate the growing sophistication of such offences. Similar threats are particularly relevant in India due to the rapid expansion of digital banking, Unified Payments Interface (UPI) transactions, and online financial services.

In addition to financial crimes, deepfake technology has been used to spread misinformation, manipulate electoral discourse, damage reputations through fabricated videos, and facilitate cyber extortion. The speed with which manipulated content can be disseminated through social media platforms significantly amplifies its impact, often causing irreversible harm before its authenticity can be challenged. These developments illustrate that deepfakes are no longer merely technological innovations but have evolved into powerful instruments capable of

[3] Joel Stehouwer, Hung Dang, Feng Liu, Xiaoming Liu and Anil K Jain, 'On the Detection of Digital Face Manipulation' (2019) *arXiv* https://arxiv.org/abs/1910.01717 accessed 17 July 2026.

facilitating diverse forms of criminal activity[4].

Understanding the origin, technological mechanism, and criminal misuse of deepfake technology is essential for evaluating the adequacy of existing criminal laws. The increasing sophistication of AI-generated synthetic media underscores the need to examine whether the Bharatiya Nyaya Sanhita, 2023 provides an effective legal framework to address emerging forms of technology-enabled crime and impose appropriate criminal liability on offenders. The increasing use of deepfake technology in the political sphere has raised significant concerns regarding electoral integrity and democratic discourse. During the period preceding the 2024 General Elections, manipulated videos depicting Congress leader Rahul Gandhi were circulated widely on social media platforms. These AI-generated or digitally altered videos falsely portrayed him making inflammatory remarks and expressing views that differed from his actual public statements, thereby creating the potential to mislead voters and influence public opinion. Independent fact-checking organisations, including Alt News and BOOM Live, examined the circulated content and identified several videos as manipulated or deceptively edited. Through digital forensic analysis, including frame-by-frame comparison, audio examination, and verification against original recordings, these organisations demonstrated that the fabricated content had altered facial expressions, speech patterns, and contextual elements to create a misleading narrative[5]. The rapid dissemination of such manipulated media highlighted the growing sophistication of synthetic content and the challenges associated with distinguishing authentic political communication from fabricated material[6].

3. Legal Framework Governing Deepfake-Related Offences under the Bharatiya Nyaya Sanhita, 2023

The Bharatiya Nyaya Sanhita, 2023 (BNS) marks a significant shift in India's substantive criminal law by replacing the Indian Penal Code, 1860. Although the legislation seeks to modernize criminal law and address emerging forms of criminality, it does not expressly define or criminalize deepfake technology or AI-generated synthetic media. Consequently, the legal

[4] Abhimanyu Ghoshal, 'Twitter, Pornhub and Other Platforms Ban AI-Generated Celebrity Porn' *The Next Web* (7 February 2018) https://thenextweb.com/news/twitter-pornhub-and-other-platforms-ban-ai-generated-celebrity-porn accessed 17 July 2026.

[5] Amrita Khalid, 'Deepfake Videos Are a Far, Far Bigger Problem for Women' *Quartz* (7 October 2019) https://qz.com/1723476/deepfake-videos-feature-mostly-porn-according-to-new-study-from-deeptrace-labs accessed 17 July 2026.

[6] Toby Antony, 'Case against Kerala Man for Creating Fake Video on Rahul Gandhi' *The New Indian Express* (4 May 2024) https://www.newindianexpress.com/states/kerala/2024/May/04/case-against-kerala-man-for-creating-fake-video-on-rahul-gandhi accessed 17 July 2026.

treatment of deepfake-related offences depends upon the interpretation and application of existing provisions of the BNS[7]. This legislative omission creates uncertainty regarding the extent to which traditional criminal offences can effectively encompass AI-enabled manipulation and digital deception.

3.1 Forgery and False Electronic Records

Deepfake-related conduct may fall within the ambit of Section 336 of the Bharatiya Nyaya Sanhita, 2023, which criminalizes the making of a false document or false electronic record with the intention of causing injury, committing fraud, or inducing another person to part with property. Since deepfakes involve the creation of manipulated audio, video, or image content capable of deceiving viewers, such synthetic media may be treated as false electronic records under the statutory framework. Nevertheless, the concept of forgery under the BNS was primarily designed to address falsified documents and electronic records rather than entirely AI-generated representations. As a result, the application of Section 336 to deepfake technology may require broader judicial interpretation to accommodate emerging forms of digital manipulation. In *National Stock Exchange of India Ltd. v. Meta Platforms Inc. & Ors.*[8], The Bombay High Court directed the removal of deepfake videos falsely portraying the National Stock Exchange and misleading investors through fabricated stock recommendations. The Court highlighted the responsibility of digital intermediaries in curbing the dissemination of deceptive AI-generated content (2024).

3.2 Financial Fraud and Cybercrime

The misuse of deepfake technology has significantly expanded the scope and sophistication of financial fraud and cybercrime. Advances in artificial intelligence have enabled offenders to generate highly convincing cloned voices, fabricated video messages, and synthetic digital identities capable of impersonating corporate executives, public officials, financial institution representatives, or even family members. Such AI-generated content is frequently used to deceive victims into transferring funds, authorising financial transactions, disclosing confidential information, or granting unauthorised access to sensitive systems. Unlike conventional cyber fraud, which typically relies on forged documents, phishing emails, or

[7] The Bharatiya Nyaya Sanhita, 2023 (Act 45 of 2023), Gazette of India, Extraordinary, Part II, Section 1 (25 December 2023) https://www.mha.gov.in/sites/default/files/250883_english_01042024.pdf accessed 17 July 2026.
[8] Bombay High Court (2024).

misleading communications, deepfake-enabled fraud exploits the credibility of realistic audiovisual content. The ability of artificial intelligence to produce convincing voice recordings and video interactions substantially enhances the effectiveness of deceptive schemes by creating a false impression of authenticity and authority. Consequently, victims are more likely to rely on such communications without questioning their legitimacy, resulting in significant financial losses and identity-related offences. In *National Stock Exchange of India Ltd. v. Meta Platforms Inc. & Ors.*,[9] The National Stock Exchange (NSE) instituted proceedings before the Bombay High Court seeking the removal of AI-generated deepfake videos circulated on various social media platforms. The impugned videos falsely depicted the Managing Director and Chief Executive Officer of the NSE recommending stock investments and inviting investors to join fraudulent stock-picking groups. Recognising the deceptive nature of the content and the potential harm to investors, the Court directed the intermediary platforms to remove the fabricated videos and comply with their due diligence obligations under Rule 3(1) of the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. The Court further observed that intermediaries are required to act expeditiously upon receiving complaints concerning misleading or unlawful synthetic content and ordered the removal of the impugned content within the prescribed timeframe, thereby reinforcing the responsibility of online platforms in addressing AI-generated misinformation and deepfake-related fraud.

3.3 Cheating and Personation

Deepfakes are increasingly employed to facilitate impersonation, financial fraud, and identity-based deception. AI-generated voice cloning and fabricated video communications enable offenders to impersonate individuals for unlawful financial gain or other fraudulent purposes. Such conduct may attract liability under Section 319 of the Bharatiya Nyaya Sanhita, 2023, which deals with cheating by personation. However, conventional provisions on cheating and personation were enacted to address traditional forms of deception and do not specifically contemplate sophisticated AI-generated impersonation capable of deceiving numerous victims simultaneously. The technological complexity and widespread impact of deepfake-enabled fraud therefore expose the limitations of relying solely on general provisions governing cheating and personation. The protection of personality rights was further reinforced in *Suniel*

[9] Interim Application (L) No 21456 of 2024 in Commercial IP Suit (L) No 21111 of 2024 (Bom HC, 16 July 2024)

Shetty v. John Doe[10], where the Bombay High Court granted an injunction restraining the unauthorised digital exploitation of the actor's identity, image, and persona through artificial intelligence and deepfake technology (2025)

3.4 Defamation and Injury to Reputation

The circulation of deepfake videos or audio recordings falsely depicting an individual engaging in unlawful, immoral, or objectionable conduct may give rise to criminal liability for defamation. Furthermore, Section 336(4) of the Bharatiya Nyaya Sanhita, 2023, specifically prescribes enhanced punishment where forgery of a document or electronic record is committed with the intention of harming the reputation of another person. In *Karan Johar v. Ashok Kumar & Ors.*[11] The Delhi High Court restrained the misuse of filmmaker Karan Johar's name, voice, image, and personality through AI-generated and deepfake content, reaffirming that personality rights form an integral part of an individual's identity and deserve legal protection against technological misuse (2025). Deepfake content possesses a higher degree of persuasive value because it appears visually and audibly authentic, thereby causing more severe reputational harm than conventional defamatory statements. The rapid dissemination of manipulated content through digital platforms further aggravates the injury by making its removal difficult and allowing continuous circulation across multiple online platforms. In *Gaurav Bhatia v. Naveen Kumar & Ors.*[12] The Delhi High Court restrained the circulation of defamatory deepfake videos and recognised that the unauthorised dissemination of manipulated digital content infringes an individual's reputation and personality rights, thereby warranting judicial protection (2024).

Deepfake technology has the potential to facilitate the dissemination of defamatory and hate-based content by fabricating audio, video, or images that falsely depict individuals making offensive or inflammatory statements. Such manipulated content may seriously damage an individual's reputation or be used to incite communal disharmony and social unrest. In such situations, the existing provisions of the Bharatiya Nyaya Sanhita, 2023, relating to defamation and hate speech may provide a basis for criminal liability. In particular, where deepfake content is created or circulated with the intention of promoting hostility between different religious,

[10] Commercial IP Suit (L) No 27824 of 2025, Interim Application (L) No 27826 of 2025 (Bombay High Court, decided in 2025)
[11] CS(COMM) 400/2025 (Del HC).
[12] 2024 SCC OnLine Del 2704

linguistic, racial, caste, or regional groups, Section 196 of the Bharatiya Nyaya Sanhita, 2023 (formerly Section 153A of the Indian Penal Code, 1860) may be attracted. The provision seeks to preserve public order by penalising acts that promote enmity or disturb communal harmony. Nevertheless, the application of these provisions to AI-generated synthetic media also raises broader constitutional concerns regarding the balance between protecting freedom of speech and expression under Article 19(1)(a) of the Constitution and preventing the misuse of digital technologies to spread misinformation, hatred, and reputational harm.

3.5. Obscenity and Non-Consensual Intimate Deepfakes

A particularly serious manifestation of deepfake misuse involves the creation and dissemination of non-consensual intimate images and videos through artificial intelligence. Such content infringes the victim's privacy, dignity, and reputation while causing severe emotional and psychological harm. Although offences relating to obscenity and sexually explicit material under the existing criminal law framework may be invoked depending upon the facts of each case, the BNS does not expressly recognize synthetic intimate imagery as a distinct criminal offence. The absence of specific statutory recognition may limit effective prosecution and restrict the availability of victim-centric remedies in cases involving AI-generated sexual exploitation.

3.6 Deepfake Pornography and Gender-Based Harm

One of the most alarming manifestations of deepfake technology is the creation and dissemination of non-consensual intimate content, commonly referred to as deepfake pornography. This involves the use of artificial intelligence to digitally superimpose an individual's face onto sexually explicit images or videos without their knowledge or consent, thereby producing fabricated content that appears authentic. Women are disproportionately targeted by this form of abuse, making it a serious concern from the perspectives of gender justice, privacy, and the protection of human dignity. The consequences of deepfake pornography extend well beyond the digital environment. Victims frequently suffer severe emotional distress, psychological trauma, social stigma, reputational harm, and violations of their privacy and personal autonomy. The fact that such content is artificially generated does not lessen its impact, as viewers often perceive the manipulated material as genuine. Once disseminated through digital platforms, the content can be replicated and shared extensively, making its complete removal virtually impossible and prolonging the victim's suffering. The

misuse of artificial intelligence in generating sexually explicit synthetic media exposes the limitations of traditional legal frameworks, which were not specifically designed to address AI-enabled forms of exploitation. The increasing prevalence of deepfake pornography therefore highlights the need for a robust legal response capable of providing effective victim protection, ensuring accountability, and addressing the unique harms arising from technology-facilitated gender-based violence. In *Bollea v Gawker Media*[13], The publication of Hulk Hogan's private sex tape without consent was held to violate privacy rights, resulting in a substantial damages award. The principles of privacy, dignity, and consent are directly applicable to deepfake pornography, where fabricated intimate images similarly infringe personal autonomy.

3.7 Applicability of the Information Technology Act, 2000

The Information Technology Act, 2000 constitutes the primary legislative framework governing cyber activities and electronic transactions in India. Although the Act does not expressly regulate deepfake technology or AI-generated synthetic media, several of its provisions may be invoked to address offences arising from the misuse of such technology. For instance, Section 66D penalises cheating by personation through the use of computer resources and may be applied where deepfakes are used to facilitate identity theft, financial fraud, or digital impersonation. Similarly, Section 66, which deals with computer-related offences, may be relevant in cases involving unauthorised access or manipulation of digital systems. While Section 66A formerly addressed the communication of offensive electronic messages, it was declared unconstitutional by the Supreme Court in *Shreya Singhal v Union of India*[14] and is no longer in force. Further, the Act imposes due diligence obligations on intermediaries under Section 79, read with the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, requiring online platforms to exercise reasonable care in preventing and removing unlawful digital content. These provisions indirectly contribute to regulating the dissemination of malicious deepfake material. Nevertheless, the Information Technology Act was enacted before the emergence of sophisticated artificial intelligence technologies and does not specifically recognise or criminalise the creation and circulation of deepfakes. Consequently, the existing legal framework remains inadequate to address the complex challenges posed by AI-generated synthetic media, thereby underscoring the need for

[13] *LLC*, No 12012447CI-011 (Fla Cir Ct, Pinellas County, 18 March 2016).
[14] (2015) 5 SCC 1.

comprehensive legislative reforms specifically targeting deepfake-related offences[15].

4. Comparative Perspective: Regulatory Approaches to Deepfake Technology

The rapid advancement of deepfake technology has prompted several jurisdictions to introduce legislative and regulatory measures aimed at addressing the risks posed by AI-generated synthetic media. Although the nature and extent of regulation vary across countries, these approaches collectively demonstrate an increasing recognition of the need to balance technological innovation with the protection of individual rights, public confidence, electoral integrity, and digital security. A comparative analysis of these regulatory models provides valuable guidance for India in developing an effective legal framework to address deepfake-related offences.

4.1 United States

The United States has not enacted a comprehensive federal law specifically regulating deepfake technology. Instead, regulation has developed through a combination of state legislation and sector-specific federal initiatives targeting particular forms of misuse. Several states have enacted laws prohibiting the creation and dissemination of non-consensual intimate deepfakes and manipulated media intended to interfere with elections. More recently, the Take It Down Act, 2025 requires online platforms to remove non-consensual intimate imagery, including AI-generated content, within forty-eight hours of receiving a valid complaint, while providing enforcement through the Federal Trade Commission. In addition, the proposed No FAKES Act seeks to establish federal protection against the unauthorised digital replication of an individual's voice, image, and likeness[16]. Despite these regulatory developments, deepfake regulation in the United States remains subject to the constitutional guarantees of freedom of speech under the First Amendment. Consequently, courts continue to balance the regulation of harmful synthetic media against legitimate forms of artistic expression, parody, satire, and political speech. One of the most widely discussed instances of manipulated digital media involved a doctored video of former United States House Speaker Nancy Pelosi, which gained widespread circulation across social media platforms. The video was deliberately altered by reducing its playback speed and modifying the audio to create the misleading impression that

[15] Information Technology Act 2000 (Act 21 of 2000) https://www.indiacode.nic.in/bitstream/123456789/13116/1/it_act_2000_updated.pdf accessed 17 July 2026.
[16] Danielle Keats Citron and Robert Chesney, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 *California Law Review* 1753.

she was slurring her speech and exhibiting signs of mental impairment. Although the video was not created using sophisticated deepfake algorithms, it demonstrated how digitally manipulated audiovisual content could significantly distort public perception and rapidly spread misinformation[17].

4.2 European Union

The European Union has adopted one of the most comprehensive regulatory frameworks for artificial intelligence through the Artificial Intelligence Act (AI Act). Article 50(4) of the AI Act, together with the European Commission's 2026 Guidelines, introduces transparency obligations requiring providers of AI systems to ensure that AI-generated or manipulated content is clearly disclosed. The framework also encourages the use of machine-readable watermarking and other technical measures to identify synthetic media. Unlike many jurisdictions, the European Union adopts a risk-based approach that distinguishes harmful deepfakes from legitimate uses of artificial intelligence. Accordingly, lighter regulatory obligations apply to content created solely for artistic, fictional, or satirical purposes, provided that appropriate disclosures are made[18]. By emphasising transparency, accountability, and responsible innovation, the European Union seeks to reduce deception while safeguarding freedom of expression and technological development.

4.3 United Kingdom

The United Kingdom has addressed the misuse of deepfake technology through amendments to the Online Safety Act, 2023[19], which criminalise the creation and dissemination of non-consensual deepfake intimate images. The legislation imposes duties upon digital platforms to address harmful online content while strengthening protection for victims of technology-facilitated abuse. At the same time, existing principles of defamation law and statutory exceptions relating to parody and fair dealing continue to safeguard legitimate forms of expression. The UK's regulatory approach therefore combines criminal sanctions for harmful conduct with protections for lawful speech and creative expression.

[17] Kate Klonick, 'The Nancy Pelosi "Deepfake" and the Future of Truth Online' *The New York Times* (26 May 2019) https://www.nytimes.com/2019/05/26/opinion/nancy-pelosi-video-facebook.html accessed 17 July 2026.
[18] Robert Chesney and Danielle Keats Citron, 'Deep Fakes and the New Disinformation War' (2019) 107 *Foreign Affairs* 147.
[19] Online Safety Act 2023 (UK) https://www.legislation.gov.uk/ukpga/2023/50/contents accessed 17 July 2026.

4.4 China

China has adopted one of the most stringent regulatory regimes governing AI-generated synthetic media through the Deep Synthesis Provisions, which came into force in January 2023. The regulations require providers of deep synthesis services to verify user identities, implement content monitoring systems, and clearly label AI-generated or manipulated content. They further prohibit the creation or dissemination of synthetic media that may mislead the public, spread false information, or endanger national security and public interests. Although this framework provides extensive regulatory oversight and platform accountability, it offers comparatively limited protection for freedom of expression, particularly in relation to politically sensitive or satirical content.

5. Challenges in Regulating Deepfake Technology

5.1 Lack of a Specific Legal Framework

One of the foremost challenges in addressing deepfake-related offences in India is the absence of a dedicated statutory framework that expressly regulates the creation, dissemination, and misuse of AI-generated synthetic media. Although malicious deepfakes may be prosecuted under provisions relating to forgery, cheating, personation, defamation, obscenity, and offences involving electronic records, the Bharatiya Nyaya Sanhita, 2023 does not specifically recognise deepfakes as a distinct criminal offence. Consequently, courts are often required to adopt an expansive interpretation of existing penal provisions to determine criminal liability. The unauthorised creation and circulation of deepfakes also implicate the fundamental right to privacy guaranteed under Article 21 of the Constitution of India. As recognised in *Justice K.S. Puttaswamy (Retd.) v. Union of India*[20], Informational privacy includes an individual's right to control the collection, use, and dissemination of personal information. The unauthorised use of a person's image, voice, or other identifying characteristics for generating synthetic media therefore constitutes a serious infringement of privacy and personal autonomy. However, the existing legal framework does not specifically provide remedies for such AI-enabled violations, resulting in uncertainty regarding the scope of criminal liability and victim protection[21].

[20] (2017) 10 SCC 1.

[21] Zahra Khanjani, Gabrielle Watson and Vandana P Janeja, 'How Deep Are the Fakes? Focusing on Audio

5.2 Enforcement Challenges

The investigation and prosecution of deepfake-related offences present considerable practical difficulties. The identification of manipulated digital content requires advanced forensic techniques capable of distinguishing authentic electronic records from AI-generated synthetic media. As deepfake technology continues to evolve, the sophistication of manipulation methods has made detection increasingly complex, thereby reducing the effectiveness of conventional digital forensic tools. Effective enforcement further depends upon specialised technical expertise, adequate investigative infrastructure, and continuous training of law enforcement personnel[22]. However, limited access to advanced forensic technologies, insufficient institutional capacity, and relatively low levels of cybercrime awareness among investigating agencies continue to impede the effective detection and prosecution of deepfake-enabled offences. These technological and institutional limitations significantly affect the administration of criminal justice in cases involving AI-generated content.

5.3 Jurisdictional Challenges

The transnational nature of cyberspace presents another significant challenge in regulating deepfake-related crimes. AI-generated synthetic content may be created in one country, hosted on servers located in another jurisdiction, and disseminated globally within a matter of seconds. Although the Bharatiya Nyaya Sanhita, 2023, provides for the prosecution of offences committed within India, the application of criminal jurisdiction becomes uncertain where the creator, platform provider, server location, and victim are situated in different countries. The absence of clear legal standards governing cross-border liability further complicates investigation and prosecution[23]. Questions frequently arise regarding whether criminal responsibility should be attributed to the creator of the deepfake, the developer of the underlying AI model, or the intermediary hosting the content. Differences in national legal frameworks relating to intermediary liability, data protection, and criminal jurisdiction often delay enforcement and necessitate greater international cooperation for the preservation of electronic evidence and the prosecution of offenders.

Deepfake: A Survey' (2021) *arXiv* https://arxiv.org/abs/2111.14203 accessed 17 July 2026.

[22] Ignas Kalpokas and Julija Kalpokienė, *Deepfakes: A Realistic Assessment of Potentials, Risks, and Policy Regulation* (Springer 2022).

[23] Mohamed Hassan Mekkawi, 'The Challenges of Digital Evidence Usage in Deepfake Crimes Era' (2023) 3(2) *Journal of Law and Emerging Technologies* 176.

5.4 Privacy and Consent

Deepfake technology also raises complex issues relating to data protection, privacy, and informed consent. The Digital Personal Data Protection Act, 2023 requires data fiduciaries to obtain free, informed, specific, and unambiguous consent before processing personal data. Nevertheless, concerns remain regarding the use of publicly available photographs, videos, and voice recordings for training artificial intelligence models and generating synthetic media. Individuals may unknowingly consent to broad data processing terms without fully understanding the potential use of their personal information for AI-based applications. The importance of meaningful consent was recognised by the Delhi High Court in *Facebook India Online Services Pvt. Ltd. v. Competition Commission of India*[24], where the Court examined the fairness of WhatsApp's privacy policy requiring users to share data with Facebook as a condition for continuing to use the service. The decision highlighted concerns regarding unequal bargaining power and the validity of consent obtained through mandatory acceptance of privacy policies. Similar concerns arise in the context of deepfake technology, where personal data may be processed for AI training or synthetic media generation without the individual's genuine awareness or meaningful consent. These challenges demonstrate the need for stronger legal safeguards governing the use of personal data for artificial intelligence and deepfake technologies.

6. Recommendations for Strengthening the Legal Framework on Deepfake Technology in India

The rapid advancement of deepfake technology presents significant challenges to cybersecurity, privacy, criminal justice, and the integrity of digital communications in India. The ability of artificial intelligence to generate highly realistic but fabricated audio, video, and image content has facilitated offences such as financial fraud, identity theft, impersonation, electoral misinformation, and reputational harm. Addressing these emerging threats requires a comprehensive strategy that combines legislative reform, technological innovation, institutional capacity building, and effective regulatory oversight. The following recommendations may strengthen India's legal response to deepfake-related offences.

[24] 2022 SCC OnLine Del 3148.

6.1 Enactment of a Dedicated Legal Framework

One of the foremost priorities is the enactment of comprehensive legislation specifically regulating deepfake technology. At present, offences involving deepfakes are addressed indirectly through the Bharatiya Nyaya Sanhita, 2023, the Information Technology Act, 2000, and the Digital Personal Data Protection Act, 2023. While these enactments provide remedies for certain forms of cybercrime, they do not expressly define or criminalise the malicious creation and dissemination of AI-generated synthetic media[25]. Consequently, Parliament should consider enacting a dedicated statute, such as a Deepfake Prevention and Regulation Act, which clearly defines deepfakes, distinguishes legitimate uses of artificial intelligence from malicious manipulation, and prescribes proportionate criminal penalties for the unauthorised creation, publication, and circulation of deceptive synthetic content. In addition, the Bharatiya Nyaya Sanhita, 2023 and the Information Technology Act, 2000, should be amended to expressly recognise deepfake-related offences involving identity theft, financial fraud, electoral interference, non-consensual intimate imagery, and digital impersonation. Such statutory recognition would provide greater legal certainty and facilitate more effective investigation and prosecution[26].

6.2 Strengthening Digital Forensic and Law Enforcement Capabilities

Effective enforcement of laws governing deepfake technology requires specialised technical expertise and modern digital forensic infrastructure. The increasing sophistication of AI-generated synthetic media has made the detection and authentication of manipulated content considerably more challenging. Therefore, specialised Deepfake Forensic Laboratories should be established under the supervision of the Indian Computer Emergency Response Team (CERT-In) and other competent authorities to assist in the identification, preservation, and forensic examination of synthetic media. Further, dedicated cybercrime units within the Central Bureau of Investigation, the Enforcement Directorate, and State Cyber Crime Cells should receive specialised training in investigating AI-enabled offences. Capacity-building programmes focusing on digital evidence collection, forensic analysis, and emerging artificial intelligence technologies would significantly enhance the effectiveness of criminal

[25] Henry Ajder, Giorgio Patrini, Francesco Cavalli and Laurence Cullen, *The State of Deepfakes: Landscape, Threats, and Impact* (Deeptrace Labs 2019).
[26] Ronald J Allen, Eleanor Swift, Richard B Kuhns, Michael S Pardo and David A Sonenshein, *Evidence: Text, Problems, and Cases* (6th edn, Aspen Publishing 2022).

investigations. Expedited forensic examination and timely legal action would also assist in preventing the rapid dissemination of harmful deepfake content, particularly in cases involving financial fraud, electoral manipulation, and organised cybercrime.

6.3 Deployment of AI-Based Deepfake Detection Technologies

Given the increasing sophistication of deepfake technology, legal regulation alone is insufficient to address the challenges posed by AI-generated synthetic media. Technological solutions must complement legislative measures to ensure the timely identification and mitigation of manipulated digital content. As deepfake generation tools continue to evolve, the development of advanced detection mechanisms has become essential for preserving the authenticity of digital communications and electronic evidence[27].

To strengthen technological resilience against deepfakes, the Ministry of Electronics and Information Technology (MeitY) should promote the development of a National Deepfake Detection System (NDDS) capable of identifying manipulated audio, video, and image content in real time. Such a system may employ artificial intelligence and machine learning techniques to analyse digital content, detect signs of manipulation, and assist law enforcement agencies in forensic investigations[28]. The integration of AI-based detection tools across social media platforms, news organisations, and digital content providers would significantly improve the early identification and removal of deceptive synthetic media.

6.4 Strengthening Platform Accountability and Content Governance

Social media platforms and digital intermediaries play a critical role in the dissemination of deepfake content. Consequently, effective regulation of synthetic media requires greater accountability on the part of technology companies responsible for hosting and distributing user-generated content. While the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 impose certain due diligence obligations upon intermediaries, additional measures may be necessary to address the unique challenges posed by deepfake technology[29]. Additionally, a robust notice-and-takedown mechanism should be

[27] Harshvardhan Mudgal, 'The Deepfake Dilemma: Detection and Decree' *Bar and Bench* (18 November 2023) https://www.barandbench.com/columns/deepfake-dilemma-detection-and-desirability accessed 17 July 2026
[28] Britt Paris and Joan Donovan, *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence* (Data & Society Research Institute 2019).
[29] Shyam Divan and Arghya Sengupta, *Electronic Evidence in India: Law and Practice* (Oxford University Press 2016).

established to ensure the prompt removal of harmful deepfake content, particularly where it involves financial fraud, identity theft, electoral misinformation, or non-consensual intimate imagery. Specified timelines for responding to complaints and removing unlawful content would improve victim protection while reducing the harmful consequences associated with the rapid dissemination of synthetic media.

6.5 Promoting Public Awareness and Digital Literacy

The effectiveness of any legal framework regulating deepfake technology depends not only upon legislative measures but also upon public awareness and digital literacy. A significant challenge in combating deepfake-related offences is the inability of many individuals to distinguish manipulated synthetic media from authentic digital content. This lack of awareness increases the risk of financial fraud, identity theft, cyber harassment, and the rapid dissemination of misinformation. To address this concern, the Government should implement a nationwide Digital Literacy and Deepfake Awareness Programme aimed at educating citizens about the nature of AI-generated synthetic media, methods of identifying manipulated content, and mechanisms for reporting suspicious material to the appropriate authorities. Educational institutions should also incorporate digital literacy and media verification into their curricula to equip students with the skills necessary to critically evaluate online information and recognise AI-enabled misinformation.

6.6 Strengthening International Cooperation

The borderless nature of cyberspace presents significant challenges in investigating and prosecuting deepfake-related offences. AI-generated synthetic media may be created in one jurisdiction, hosted on servers located in another, and disseminated globally within a matter of seconds. Such transnational characteristics often impede effective law enforcement and complicate the collection and preservation of digital evidence. Accordingly, India should strengthen international cooperation through mutual legal assistance mechanisms and cybersecurity partnerships with foreign governments and international organisations. Enhanced collaboration with agencies such as INTERPOL and Europol would facilitate cross-border investigation, intelligence sharing, and the identification of individuals involved in transnational deepfake-related crimes[30]. Cooperation with major technology companies,

[30] Robert Chesney and Danielle Keats Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107(6) *California Law Review* 1753.

including social media platforms and artificial intelligence developers, should also be encouraged to develop globally accepted standards for detecting, authenticating, and removing malicious synthetic media. Furthermore, India should actively participate in international discussions on artificial intelligence governance under multilateral forums, including the United Nations, to contribute towards the development of harmonised legal principles governing AI-generated content. Such coordinated international efforts would strengthen the global response to deepfake-enabled offences while supporting effective enforcement of India's criminal law framework.

7. Conclusion

The emergence of deepfake technology has introduced unprecedented challenges to India's criminal justice system by enabling the creation and dissemination of highly realistic yet fabricated digital content capable of facilitating identity theft, financial fraud, cyber harassment, electoral misinformation, and violations of privacy and reputation. Although the Bharatiya Nyaya Sanhita, 2023 modernises India's substantive criminal law, it does not expressly recognise deepfake technology or AI-generated synthetic media as distinct criminal offences. Consequently, liability for deepfake-related misconduct must presently be established through the application of existing provisions relating to forgery, cheating, personation, defamation, offences against women, and electronic records. While these provisions offer partial legal remedies, they were not enacted with artificial intelligence-enabled offences in contemplation and therefore present significant interpretative and enforcement challenges.

This study demonstrates that the existing legal framework, though capable of addressing certain forms of deepfake misuse, remains fragmented and insufficient to effectively regulate the evolving nature of AI-generated synthetic media. The absence of specific statutory definitions, dedicated offences, uniform evidentiary standards, and clear investigative procedures creates uncertainty in the prosecution of deepfake-related crimes. Moreover, challenges relating to digital forensic detection, intermediary liability, cross-border enforcement, and the protection of privacy and personality rights further complicate the administration of criminal justice in the digital age. A balanced and forward-looking regulatory approach is therefore essential. Legislative reform should include the express recognition of malicious deepfakes as a distinct category of criminal conduct under the Bharatiya Nyaya Sanhita, 2023 or through a separate

specialised legislation. Such reform should be complemented by robust digital forensic infrastructure, specialised training for investigating agencies, enhanced intermediary accountability, effective mechanisms for the prompt removal of unlawful synthetic content, and greater public awareness regarding the misuse of artificial intelligence. At the same time, any regulatory framework must carefully preserve constitutional guarantees of freedom of speech and expression by distinguishing malicious deepfakes from legitimate uses of artificial intelligence for education, research, artistic expression, satire, and other lawful purposes. As artificial intelligence continues to transform digital communication, the law must evolve correspondingly to address emerging forms of technological misuse. An effective legal response should not merely react to individual instances of deepfake-related harm but establish a comprehensive framework capable of ensuring accountability, protecting individual rights, preserving the integrity of digital evidence, and maintaining public confidence in the criminal justice system. A proactive and technologically informed approach will enable India to effectively address the challenges posed by deepfake technology while fostering responsible innovation and strengthening the rule of law in the digital era.