Deepfake Technology and Criminal Liability: Need for a Comprehensive Legal Framework

Megharani Nikam
Modern Law College

***Abstract*—Deepfake technology, powered by artificial intelligence (AI) and deep learning algorithms, has transformed the creation and manipulation of digital content by producing highly realistic but fabricated audio, video, and images.**

Although the technology has legitimate applications in education, entertainment, healthcare, and accessibility, its misuse has created significant legal and ethical concerns.

Such misuse threatens individual rights, public trust, democratic institutions, and the integrity of criminal justice systems.

The rapid evolution of deepfake technology has outpaced existing legal frameworks in many jurisdictions. In India, offences involving deepfakes are currently addressed through fragmented provisions of the Bharatiya Nyaya Sanhita, 2023 (BNS), the Information Technology Act, 2000, copyright law.

However, these provisions do not specifically define or comprehensively regulate deepfake-generated content, resulting in challenges relating to criminal liability, evidentiary standards, attribution of offenders, jurisdiction, and victim protection.

The absence of a dedicated legislative framework further complicates investigation, prosecution, and international cooperation against transnational digital offences.

This research examines the criminal implications of deepfake technology by analysing the adequacy of existing legal provisions, judicial responses, and comparative legislative developments in selected jurisdictions such as the European Union, the United States, and China.

It also explores the evidentiary challenges associated with AI-generated digital content and the role of forensic technologies in detecting manipulated media.

The study argues that India requires a comprehensive legal framework that clearly defines deepfake-related offences, prescribes proportionate criminal sanctions, strengthens investigative mechanisms, ensures accountability of AI developers and digital intermediaries, and balances technological innovation with constitutional rights, including freedom of speech, privacy, and due process. Such a framework would enhance public confidence in digital communications while safeguarding the administration of criminal justice.

***Index Terms*—Deepfake, Artificial Intelligence, Cyber Crime, Criminal Liability, Digital Evidence, Privacy, Cyber Law.**

I. INTRODUCTION

The twenty-first century has witnessed a remarkable transformation in the field of digital technology, particularly with the rapid development of Artificial Intelligence (AI). The development of synthetic media has resulted in the emergence of deepfake technology, which represents one of the most controversial applications of Artificial Intelligence.

The term “deepfake” is derived from the combination of “deep learning” and “fake” and refers to artificially generated or manipulated digital content in which a person's face, voice, expressions, or actions are altered or recreated through AI-based techniques. Deepfake technology enables the creation of highly realistic videos, images, and audio recordings that portray individuals performing acts or making statements that never actually occurred.

The development of Generative Adversarial Networks (GANs) has significantly contributed to improving the quality and realism of deepfake content. As a result, deepfakes have created a new challenge for legal systems because the distinction between genuine and fabricated digital evidence has become increasingly difficult.

Criminal actors have increasingly exploited deepfake tools for various unlawful activities, including identity theft, financial fraud, cyber harassment, defamation, political misinformation, and non-consensual sexual exploitation. The creation of fake videos or audio

recordings of individuals without consent can severely affect personal dignity, reputation, and privacy rights.

One of the most concerning forms of deepfake misuse is the creation of non-consensual intimate content. Individuals, particularly women, have been targeted through AI-generated explicit images and videos, resulting in serious psychological, social, and legal consequences.

The misuse of deepfake technology has also created challenges in the political and democratic sphere. Artificially generated videos of political leaders or public figures can be used to spread false information, influence public opinion, and manipulate electoral processes.

The increasing use of deepfakes has created significant challenges for criminal justice systems. The first major challenge relates to the identification and prosecution of offenders. The borderless nature of cyber activities makes investigation and attribution difficult for law enforcement agencies.

Another important challenge concerns the establishment of criminal intention (*mens rea*). In deepfake cases, courts must determine whether the person responsible for creating or distributing synthetic content intended to deceive, harm, defame, or exploit another individual. The complexity of AI-generated content creates difficulties in applying traditional principles of criminal responsibility.

A further challenge arises regarding the admissibility and reliability of digital evidence. Courts rely upon electronic records to determine facts; however, deepfake technology creates possibilities for manipulation of audio-visual evidence. The increasing sophistication of AI-generated content requires specialised forensic techniques to verify authenticity and establish whether digital evidence has been altered. This creates a need for stronger digital forensic mechanisms and expert involvement in criminal investigations.

The existing legal framework in India provides certain remedies against deepfake-related offences through provisions of the Information Technology Act, 2000, Bharatiya Nyaya Sanhita, 2023, and data protection legislation. Provisions relating to identity theft, cheating by personation, privacy violations, obscenity, and defamation may apply to certain deepfake activities. However, these laws were not specifically designed to address AI-generated synthetic content, resulting in gaps relating to definition, liability, prevention, and victim protection.

The central research problem of this study is whether the existing legal framework in India is sufficient to regulate criminal activities involving deepfake technology or whether a specialised legal framework is required to address emerging AI-driven offences. The difficulty lies in balancing technological advancement with protection of fundamental rights, including privacy, dignity, freedom of expression, and personal security.

The significance of this research arises from the growing impact of deepfake technology on individuals, society, and legal institutions. A comprehensive legal framework is necessary to establish clear definitions of deepfake offences, determine criminal liability, impose responsibilities on digital platforms, strengthen digital evidence mechanisms, and provide effective remedies to victims.

Therefore, this research examines the relationship between deepfake technology and criminal liability and argues for the development of a specialised legal framework in India that can effectively regulate malicious use of synthetic media while encouraging responsible innovation in Artificial Intelligence.

II. CONCEPTUAL UNDERSTANDING OF DEEPFAKE TECHNOLOGY

2.1 Meaning and Origin of Deepfakes

2.1.1 Definition of Deepfake

The term “deepfake” refers to artificially generated or manipulated digital content created through advanced Artificial Intelligence (AI) techniques, particularly deep learning algorithms. The word “deepfake” combines two concepts: “deep learning,” which refers to a branch of machine learning based on artificial neural networks, and “fake,” indicating artificially created or altered content.

According to Chesney and Citron, deepfakes represent a significant challenge because they allow the creation of realistic false information that may affect individual rights, democratic institutions, and public confidence in digital communication.

The origin of modern deepfake technology can be traced to developments in artificial intelligence, especially deep neural networks and Generative Adversarial Networks (GANs). Initially, AI-generated

images and videos were mainly used for academic research; however, increased computing power, availability of online data, and open-source AI tools have made deepfake technology accessible to ordinary users.

Recent incidents demonstrate the growing misuse of deepfake technology.

In India, concerns increased after circulation of manipulated videos involving public personalities and political figures. The misuse of deepfake videos during election periods has highlighted the potential of synthetic media to spread misinformation and influence public perception. The Supreme Court of India and government authorities have also expressed concerns regarding the harmful impact of AI-generated misinformation and the necessity of regulatory measures to address such threats.

2.1.2 Role of Artificial Neural Networks

Artificial Neural Networks (ANNs) form the technological foundation of deepfake creation. Neural networks are computational models inspired by the functioning of the human brain. They consist of interconnected layers of artificial neurons that process information, identify patterns, and generate predictions based on available data.

Deepfake systems commonly use deep neural networks to analyse thousands of images, videos, or audio samples of a particular individual. Through repeated training, the system learns characteristics such as: Facial structure; Eye movements; Expressions; Speech patterns; Voice characteristics.

2.1.3 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are among the most important technologies behind modern deepfake creation. GANs were introduced by Ian Goodfellow and his colleagues in 2014 and consist of two competing neural networks:

1. Generator Network – Creates artificial content.
2. Discriminator Network – Evaluates whether the generated content appears real or artificial.

The generator continuously improves its output based on feedback from the discriminator until the synthetic content becomes highly realistic. This process enables AI systems to create convincing images, videos, and audio that may be difficult for human observers to distinguish from genuine material.

2.1.4 Audio, Video and Image Manipulation

Deepfake technology operates through different forms of digital manipulation:

(a) Image Manipulation

AI-based image manipulation involves altering photographs by replacing faces, modifying expressions, or creating completely artificial images. Such manipulated images may be used for identity theft, fake profiles, or reputational harm.

(b) Video Manipulation

Video deepfakes are among the most dangerous forms of synthetic media because they create the impression that a person performed an action or made a statement that never occurred. Face-swapping technology can replace one individual's face with another's while maintaining realistic movements and expressions.

(c) Audio Manipulation

Voice cloning technology allows AI systems to reproduce a person's voice by analysing audio samples. Criminals may use cloned voices to impersonate individuals for fraud, blackmail, or social engineering attacks.

The combination of audio and video manipulation creates highly convincing synthetic content, making detection and legal regulation increasingly challenging.

2.2 Types of Deepfake Applications

2.2.1 Face Swapping

It involves replacing the face of one individual in a photograph or video with another person's face while maintaining realistic facial movements and expressions.

Although face swapping has legitimate entertainment applications, it has also been misused for: Non-consensual sexual content; Fake identity creation; Defamation; Political misinformation.

The misuse of face-swapping technology raises serious questions regarding privacy rights and personal dignity.

2.2.2 Voice Cloning

Voice cloning refers to the use of AI technology to reproduce an individual's voice.

Voice cloning has become a significant concern in cyber fraud cases. Criminals may imitate the voice of company executives, relatives, or government officials to manipulate victims into transferring money or sharing confidential information.

2.2.3 Synthetic Videos

Synthetic videos involve AI-generated visual content where individuals appear to speak, act, or perform activities that never actually happened.

Such videos may be used for: False political statements; Fake news creation; Online harassment; Manipulation of public opinion.

The increasing realism of synthetic videos creates challenges for law enforcement agencies in distinguishing genuine evidence from fabricated material.

2.2.4 AI-Generated Impersonation

AI-generated impersonation involves creating a false digital identity by combining artificial images, voices, and behavioural patterns.

Such impersonation may result in: Financial fraud; Cyber stalking; Identity theft; Corporate deception.

Deepfake impersonation creates difficulties because traditional identity verification methods may become ineffective when criminals can digitally reproduce another person's appearance and voice.

2.3 Positive Uses of Deepfake Technology

Although deepfake technology is commonly associated with criminal misuse, it also has several legitimate and beneficial applications.

2.3.1 Education

Deepfake technology can enhance educational experiences by creating interactive learning materials. Historical personalities can be digitally recreated to explain events, languages can be taught through realistic virtual instructors, and personalised educational content can be developed using AI-generated avatars.

2.3.2 Healthcare

In healthcare, AI-generated synthetic media can support medical training and research. Deepfake-based simulations may assist doctors and students in understanding medical procedures through realistic visual demonstrations.

2.3.3 Entertainment Industry

The entertainment industry has widely adopted AI-generated content for: Film production; Visual effects; Character recreation; Voice enhancement; Digital avatars.

Deepfake technology can help filmmakers recreate historical characters, restore old films, and reduce production limitations.

However, ethical issues relating to consent, ownership, and personality rights must be considered when using digital replicas of individuals.

2.3.4 Digital Preservation

Deepfake technology can contribute to preservation of cultural and historical heritage. Digital reconstruction of historical figures, restoration of damaged audiovisual materials, and creation of virtual museum experiences are examples of beneficial applications.

III. DEEPFAKE TECHNOLOGY AS A TOOL FOR CRIMINAL ACTIVITIES

Deepfake technology has rapidly evolved from a research innovation into a powerful tool capable of facilitating sophisticated criminal activities.

The misuse of deepfake technology has expanded the scope of cybercrime by enabling identity theft, financial fraud, online sexual exploitation, political misinformation, and corporate deception. The widespread availability of AI-based tools has significantly reduced the technical expertise required to create sophisticated digital forgeries, thereby increasing the risk of deepfake-related offences and posing new challenges for law enforcement authorities worldwide.

Unlike conventional digital manipulation, deepfakes create convincing fabricated evidence capable of deceiving individuals, financial institutions, businesses, and even governments.

3.1 Identity Theft and Impersonation

Identity theft is among the most common criminal applications of deepfake technology. Individuals engaged in cybercrime may obtain publicly accessible images, videos, and voice samples from social media, interviews, virtual meetings, and other online sources. Such material can be manipulated using synthetic media technologies to construct convincing digital representations, which may subsequently be exploited to impersonate victims and mislead relatives, employers, banks, financial organisations, or public authorities.

(a) Misuse of Personal Data

Deepfake systems require extensive personal information such as facial images, voice samples, and behavioural characteristics. Since millions of users voluntarily upload videos and photographs on social

networking sites, cybercriminals can easily obtain the necessary data for generating convincing deepfakes.

The unauthorised exploitation of personal information through AI-enabled technologies may facilitate a range of offences, including: Identity theft, Fabrication of deceptive social media accounts, Unauthorised or fraudulent verification and authentication, Manipulation or fabrication of digital identities, Blackmail, coercion, and extortion, Harm to an individual's reputation and social standing

In contrast to conventional forms of identity theft, AI-generated impersonations can reproduce a person's appearance, manner of speaking, facial movements, and emotional expressions with remarkable accuracy. This enhanced realism can make such fraudulent identities significantly more difficult to recognise and verify.

(b) Financial Fraud

Criminals clone the voice or facial appearance of account holders, company executives, or bank officials to convince victims to transfer money or disclose confidential financial information.

Common methods include: Voice cloning for banking verification, Fake customer support representatives, AI-generated video messages requesting emergency fund transfers, Deepfake-based online loan applications, Fraudulent Know Your Customer (KYC) verification

Banks that increasingly rely on biometric verification systems face additional challenges because sophisticated deepfakes may circumvent facial recognition technologies.

(c) Digital Impersonation

Digital impersonation refers to creating fake digital identities that imitate real individuals. Deepfakes allow offenders to appear in virtual meetings, video conferences, and online interviews as another person.

Criminals may impersonate: Corporate executives, Lawyers, Government officials , Police officers, Family members, Celebrities

Such impersonation enables social engineering attacks, cyber extortion, phishing campaigns, and unauthorized access to confidential information.

3.2 Cyber Fraud and Financial Crimes

Deepfake technology has transformed cyber fraud by making deception more convincing than traditional phishing or spoofing attacks. AI-generated audio and video significantly increase the credibility of fraudulent communications, resulting in substantial financial losses.

(a) Fake Video Calls

Modern AI software enables criminals to conduct live video conversations using deepfake technology. During these calls, offenders may impersonate senior executives, relatives, business partners, or government authorities.

Fake video calls are commonly used for: Emergency money requests, Business payment approvals, Confidential information extraction, Remote identity verification fraud, Social engineering attacks

Because participants observe realistic facial movements and hear cloned voices, victims often fail to recognize the deception.

(b) Business Email Compromise (BEC)

Traditionally, Business Email Compromise (BEC) attacks primarily involved the use of deceptive or spoofed email accounts to impersonate legitimate individuals or organisations and induce victims to disclose information or authorise fraudulent transactions. Deepfake technology has expanded these attacks by combining fake emails with cloned voice calls or AI-generated video meetings.

Through Business Email Compromise and related impersonation techniques, offenders may attempt to: Direct employees to make unauthorised fund transfers, Alter or replace legitimate vendor banking and payment information, Approve fictitious or manipulated invoices, Authorise unauthorised or confidential financial transactions, Acquire sensitive business, financial, or corporate information

(c) Corporate Fraud

Deepfake technology can also be misused to support sophisticated forms of corporate fraud by creating false communications, documents, or audiovisual material that may affect internal corporate processes and financial decisions.

Potential forms of misuse include: Falsified recordings of board or management meetings, Fabricated statements or announcements purportedly issued by shareholders or company officials, False representations concerning mergers, acquisitions, or business negotiations, Manipulated financial or earnings presentations, Impersonation of directors or senior executives during business negotiations

Such deceptive activities can interfere with corporate governance, mislead investors and stakeholders, adversely affect market confidence or share prices, and

potentially result in significant financial and economic losses.

3.3 Deepfake Pornography and Sexual Exploitation

One of the most harmful applications of deepfake technology is the creation of non-consensual sexually explicit content. Victims are often women, journalists, public figures, students, and private individuals whose photographs are manipulated into fabricated intimate videos or images without their consent.

The rapid spread of such material through social media and messaging applications causes severe emotional, psychological, and reputational harm.

(a) Non-Consensual Intimate Images

AI-based deepfake applications can be misused to create fabricated sexually explicit content by digitally placing or altering an individual's facial features onto another person's body without obtaining the individual's consent. Such manipulated material may cause serious violations of privacy, dignity, and personal reputation.

Consequences include: Online harassment, Extortion, Blackmail, Emotional trauma, Employment-related consequences, Permanent reputational damage

Victims often have limited legal remedies because such content spreads rapidly across multiple digital platforms.

(b) Gender-Based Digital Violence

Women are disproportionately targeted by deepfake pornography. AI-generated explicit content has become a new form of technology-enabled gender-based violence.

The abuse includes: Revenge pornography, Cyberstalking, Sexual harassment, Online intimidation, Coercion, Public humiliation

Researchers consistently report that the overwhelming majority of publicly circulating deepfake pornographic content targets women without their knowledge or consent.

(c) Privacy Violations

Deepfake pornography constitutes a serious invasion of privacy because it exploits an individual's facial identity without authorization.

The misuse and circulation of fabricated intimate content can give rise to significant privacy and personal harms, including: Erosion of individual autonomy and control over personal information, Unauthorised use or manipulation of biometric data, Exploitation of an individual's identity and likeness, Emotional and psychological distress, Persistent harm to one's digital reputation and online identity

Individuals targeted by such fabricated content may face serious emotional and social consequences, including anxiety, depression, humiliation, loss of social confidence, and isolation, particularly when the material is widely circulated online.

3.4 Political Manipulation and Election Crimes

Deepfake technology poses significant risks to democratic governance by enabling the creation of fabricated political speeches, manipulated campaign advertisements, and misinformation capable of influencing public opinion.

The realistic appearance of AI-generated political content makes it difficult for voters to distinguish authentic information from fabricated media.

(a) Fake Political Speeches

Deepfake technology can be misused to produce convincing audio or video recordings that falsely depict political leaders delivering statements, expressing opinions, or making announcements that they never actually made.

These fabricated speeches may: Incite violence, Create diplomatic tensions, Damage public confidence, Mislead voters, Spread communal hatred, Influence electoral outcomes

Because such content often circulates rapidly before verification, its impact can be substantial.

(b) Disinformation Campaigns

Deepfakes strengthen coordinated disinformation campaigns by combining manipulated videos with social media bots, fake news websites, and algorithm-driven amplification.

Such campaigns aim to: Polarize society, Damage political opponents, Undermine public trust, Manipulate election narratives, Spread false information during crises

The speed and scale of online dissemination make timely correction difficult.

(c) Threat to Democratic Processes

Deepfakes threaten electoral integrity by reducing public confidence in authentic information. They also create the "liar's dividend," whereby genuine recordings may be dismissed as fake, enabling wrongdoers to deny authentic evidence.

Potential consequences include: Voter manipulation, Election interference, Declining trust in democratic institutions, Increased political polarization,

Challenges for election regulators, Weakening of public confidence in digital evidence

IV. EXISTING LEGAL FRAMEWORK IN INDIA

The rapid advancement of deepfake technology has created complex legal challenges because synthetic media can be used for multiple forms of cybercrime, including identity theft, financial fraud, privacy violations, sexual exploitation, and misinformation.

Indian law does not currently contain a specific statute exclusively regulating deepfakes; however, several provisions under existing cyber laws, criminal laws, and data protection legislation can be applied to address different forms of misuse.

In India, legal responses to offences involving deepfake technology are drawn from multiple statutory provisions, including the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023 (BNS), and the Digital Personal Data Protection Act, 2023 (DPDP Act).

4.1 Information Technology Act, 2000

The Information Technology Act, 2000 (IT Act) is the principal legislation governing cyber offences in India.) Although enacted before the emergence of artificial intelligence-based manipulation, several provisions of the Act can address criminal activities involving deepfake content.

4.1.1 Section 43 – Unauthorized Access and Damage

Section 43 of the IT Act provides civil liability for unauthorized access to computer systems, downloading data, introducing malware, disrupting computer networks, and causing damage to digital resources.

In the context of deepfakes, this provision may apply where offenders: Illegally access personal devices or online accounts to obtain photographs, videos, or voice samples, Extract personal data without authorization for creating synthetic media and Damage or interfere with digital systems used for storing personal information..

4.1.2 Section 66 – Computer-Related Offences

Section 66 converts certain acts mentioned under Section 43 into criminal offences when they are committed dishonestly or fraudulently.

Deepfake-related activities may fall within this provision where offenders use computer resources for: Unauthorized manipulation of digital information, Fraudulent alteration of electronic content, Creation of deceptive digital identities, Causing harm through dishonest use of computer systems.

4.1.3 Section 66C – Identity Theft

Section 66C of the IT Act specifically criminalizes identity theft involving fraudulent or dishonest use of another person's electronic signature, password, or unique identification features.

Deepfake technology creates new dimensions of identity theft by allowing criminals to replicate: Facial identity, Voice patterns, Biometric characteristics, Digital appearance

Although Section 66C does not expressly mention AI-generated identities, its broad interpretation may allow prosecution where digital identity elements are misused for fraudulent purposes.

4.1.4 Section 66D – Cheating by Personation Using Computer Resources

Section 66D is one of the most relevant provisions for deepfake-related fraud. It punishes cheating by personation through communication devices or computer resources.

Deepfake crimes commonly involve digital impersonation, such as: AI-generated video calls pretending to be senior officials, Voice cloning to deceive family members, Fake online profiles using another person's identity and Fraudulent business communications.

This provision provides an effective legal mechanism for addressing financial fraud facilitated through deepfake impersonation.

4.1.6 Section 66E – Violation of Privacy

Section 66E criminalizes intentional capturing, publishing, or transmitting images of a person's private area without consent and in circumstances violating privacy.

Deepfake pornography and synthetic intimate content raise significant privacy concerns. Although Section 66E was drafted before AI-generated explicit content became common, it may apply where: Private images are unlawfully obtained, Personal photographs are manipulated and distributed and Individuals are subjected to unauthorized exposure.

However, the provision has limitations because many deepfake victims are harmed through fabricated images rather than actual recordings of private areas. Therefore, additional legal provisions may be required to specifically address synthetic sexual content.

4.1.7 Sections 67, 67A and 67B – Obscene and Sexually Explicit Content

The IT Act provides stronger protection against online sexual exploitation through Sections 67, 67A, and 67B.

Section 67 – Publishing Obscene Material

Section 67 penalizes publishing or transmitting obscene material in electronic form.

Deepfake content involving sexually explicit manipulation may attract liability where obscene synthetic content is circulated online.

Section 67A – Sexually Explicit Acts

Section 67A deals with publishing or transmitting material containing sexually explicit acts.

It may apply to deepfake videos depicting fabricated sexual acts involving real individuals without consent.

Section 67B – Child Sexual Content

Section 67B provides protection against online child sexual exploitation material.

Deepfake technology creates additional risks by enabling offenders to generate artificial sexual content involving children. Such misuse represents a serious violation of child protection principles.

Despite these provisions, the IT Act faces challenges because it does not specifically define deepfake technology or synthetic media, creating uncertainty regarding prosecution and evidentiary requirements.

4.2 Bharatiya Nyaya Sanhita, 2023

The Bharatiya Nyaya Sanhita, 2023 (BNS) replaced the Indian Penal Code, 1860 and provides a contemporary criminal-law framework for addressing various offences, including conduct that may arise in or involve the digital environment.

Although BNS does not contain a separate offence for deepfake creation, several provisions may apply depending upon the nature and consequences of the misuse.

A. Cheating

Deepfake-enabled fraud frequently involves deception and dishonest inducement.

BNS provisions relating to cheating may apply where offenders: Misrepresent their identity, Create false digital appearances, Induce victims to deliver money or property, Obtain confidential information through deception.

For example, an AI-generated video impersonating a business owner to approve fraudulent payments may constitute cheating under BNS.

B. Forgery and False Electronic Records

Deepfakes create significant challenges regarding authenticity of digital evidence. Manipulated videos, fabricated audio recordings, and altered electronic documents may amount to forgery when created with the intention to deceive.

Relevant offences include: Creation of false electronic records, Use of fabricated digital material as genuine evidence and Manipulation of records for unlawful benefit.

Deepfake technology expands the traditional concept of forgery by allowing criminals to fabricate realistic audiovisual evidence rather than merely altering documents.

C. Defamation

Deepfake content can seriously damage an individual's reputation by portraying false statements, actions, or behaviour.

Defamation provisions under BNS may apply where:

a. False videos are created and circulated.
b. Individuals are falsely shown making offensive statements.
c. Professional or personal reputation is harmed.

Public figures, professionals, and private individuals may become victims of reputational attacks through manipulated content.

D. Criminal Intimidation

Deepfake technology may also be used for threats, blackmail, and coercion.

Examples include: Threatening to release fabricated intimate images, Extorting money through manipulated videos, Using fake recordings to intimidate victims.

Such conduct may attract criminal intimidation provisions under BNS.

4.3 Digital Personal Data Protection Act, 2023

The Digital Personal Data Protection Act, 2023 establishes a statutory framework for the collection, processing, protection, and lawful use of digital personal data within India, with the objective of strengthening individuals' control over their personal information.

Since deepfake creation often depends upon unauthorized collection and processing of personal information, the Act has significant relevance.

A. Protection of Personal Information

Deepfake generation requires access to personal data such as: Facial images, Voice recordings, Videos and Biometric-related information

The DPDP Act establishes obligations regarding lawful processing and protection of digital personal data.

Organizations collecting personal information must ensure appropriate safeguards against misuse and unauthorized access.

B. Consent-Based Processing

A fundamental principle of the DPDP Act is that personal data processing should generally occur with informed consent of the data principal.

Deepfake creation without consent violates this principle because individuals usually have no knowledge or control over the use of their: Photographs, Videos, Voice samples, Digital identities Unauthorized use of personal data for generating synthetic media raises serious questions regarding informational autonomy and privacy rights.

C. Liability for Misuse of Personal Data

The DPDP Act imposes obligations on data fiduciaries and processors to prevent unauthorized processing and protect personal information.

Liability may arise where organizations fail to: Implement reasonable security measures, Prevent data breaches, Protect personal information from misuse.

However, challenges remain regarding individual criminals who collect publicly available data and create deepfakes outside regulated data-processing environments.

V. CRIMINAL LIABILITY CHALLENGES IN DEEPFAKE CASES

Deepfakes present a distinctive challenge to criminal law because the technology can manipulate a person's identity, speech, appearance or conduct in a manner that may be difficult to distinguish from authentic material.

In India, there is presently no comprehensive and independent criminal offence specifically titled "deepfake offence." Depending upon the facts and circumstances, deepfake-related conduct may attract provisions of the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023 (BNS), laws relating to defamation, obscenity, impersonation, privacy and intellectual property, as well as other sector-specific legislation.

This fragmented legal framework creates significant difficulties in establishing individual criminal responsibility.

5.1 Identification of the Offender

5.1.1 Anonymous Perpetrators

One of the primary difficulties in prosecuting deepfake offences is identifying the individual who created or disseminated the manipulated content. A perpetrator may operate through a pseudonymous account, temporary email address, virtual private network (VPN), anonymous cryptocurrency transaction, compromised account or multiple intermediary platforms. In such circumstances, the identity visible on a social-media account may not correspond with the actual person responsible for the offence.

The problem becomes more complicated when the offender deliberately removes metadata or repeatedly transfers the content between different platforms. A deepfake may be created by one person, uploaded by another and subsequently circulated by thousands of users. Criminal investigation must therefore distinguish between the originator, modifier, first disseminator and subsequent sharers.

Indian courts have recognised the importance of electronic trails in establishing criminal conduct. In *Shreya Singhal v. Union of India*, the Supreme Court examined the constitutional dimensions of online speech and emphasised the need to distinguish protected expression from legally punishable conduct. Although the case did not concern deepfakes, its reasoning is relevant because criminal regulation of online content must operate within constitutional limitations, particularly freedom of speech under Article 19(1)(a).

The investigation of anonymous deepfake offenders may require obtaining IP logs, subscriber information, device identifiers, account-registration details, server records, payment information and platform-generated metadata. However, the availability and reliability of such information depend significantly upon the cooperation of intermediaries and service providers.

5.1.2 Use of Encrypted Platforms

End-to-end encrypted communication services create an additional investigative barrier. Where manipulated material is circulated through encrypted messaging services, investigators may be unable to directly access the contents of communications. Even when the content itself cannot be obtained, investigators may

potentially rely upon surrounding digital evidence such as device records, timestamps, account information, screenshots, downloaded files and communication metadata, subject to applicable law.

This creates a delicate balance between privacy, encryption and legitimate criminal investigation. The Supreme Court's decision in *K.S. Puttaswamy (Retd.) v. Union of India* recognised privacy as a constitutionally protected fundamental right under Article 21. Accordingly, investigative measures involving digital devices and private communications must satisfy constitutional requirements of legality, necessity and proportionality.

Thus, the mere existence of encrypted communication cannot automatically be treated as evidence of criminality. Investigators must establish a legally sustainable connection between the accused and the creation or dissemination of the deepfake.

5.1.3 Cross-Border Challenges

Deepfake offences may involve several jurisdictions simultaneously. The offender may be located in India, the manipulated content may be stored on a server outside India, the victim may be located in another country and the social-media platform may be incorporated in a fourth jurisdiction.

This creates difficulties concerning: Identification of the offender; preservation and acquisition of electronic evidence; service of legal requests upon foreign entities; determination of applicable criminal law; recognition of investigative requests by foreign authorities; and admissibility of foreign-obtained digital evidence before Indian courts.

The cross-border character of cyber offences therefore makes traditional territorial concepts of criminal jurisdiction increasingly difficult to apply.

5.2 Requirement of *Mens Rea*

Criminal liability ordinarily requires examination of the mental element accompanying the prohibited conduct. In deepfake cases, establishing *mens rea* is particularly important because the same technological act may produce completely different legal consequences depending upon the accused's intention and knowledge.

5.2.1 Intention Behind Creation and Dissemination

A person may create synthetic content for legitimate purposes such as entertainment, research, satire, education or film production. Conversely, the same technology may be employed to impersonate another person, sexually exploit an individual, spread misinformation, facilitate fraud, damage reputation or threaten a victim.

Therefore, the criminal law must distinguish between technological creation as an activity and criminal misuse of that technology.

Where the prosecution alleges intentional creation of a harmful deepfake, relevant circumstances may include: the nature of the manipulated material; the identity of the person depicted; communications surrounding its creation; previous threats or disputes; the manner in which the material was uploaded; efforts to conceal the offender's identity; captions or accompanying statements; financial or other benefits obtained from dissemination; and subsequent conduct of the accused.

The Supreme Court has repeatedly emphasised the importance of the mental element where an offence requires a particular intention or knowledge. In *Nathulal v. State of Madhya Pradesh*, the Court recognised the significance of *mens rea* as a general principle of criminal liability, subject to statutory exclusion or modification.

Consequently, the prosecution in a deepfake case should not rely solely upon proof that the accused's account was associated with the manipulated content. It must establish the mental element required by the particular offence invoked.

5.2.2 Difference Between Creation and Sharing

An important legal distinction exists between creating a deepfake and sharing or forwarding an existing deepfake.

The creator may possess the technical knowledge and specific intention required to manufacture the manipulated material. A person who subsequently forwards the material may, however, have acted without knowing that it was artificially generated.

For example, an individual who receives a manipulated video through a messaging platform and forwards it believing that it is genuine may stand on a different legal footing from a person who deliberately manufactures the video and circulates it to harm the victim.

Therefore, criminal liability should be assessed with reference to the accused's: knowledge of falsity; intention; awareness of the circumstances; purpose of dissemination; and specific offence alleged.

This distinction is particularly important in mass-distribution cases because treating every subsequent

recipient as equally criminally liable could lead to disproportionate criminalisation.

5.3 Evidentiary Challenges

Digital evidence constitutes the foundation of most deepfake prosecutions. However, the evidentiary problem is twofold: the prosecution must establish both what the digital material contains and whether the material is authentic or manipulated.

5.3.1 Authenticity of Digital Evidence

A deepfake can be exceptionally convincing. Consequently, visual appearance alone cannot establish authenticity. Courts may need to examine metadata, source files, device information, hash values, timestamps, platform records, forensic reports and other technical indicators.

The Supreme Court's jurisprudence concerning electronic evidence has established important safeguards. In *Anvar P.V. v. P.K. Basheer*, the Supreme Court held that electronic records sought to be admitted through a computer-generated output are governed by the statutory requirements concerning electronic evidence. The subsequent decision in *Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal* reaffirmed and clarified the principles concerning certification of electronic records.

The evidentiary significance of these decisions is substantial for deepfake litigation. A prosecution cannot simply produce a downloaded video and assume that its contents are self-proving. The court may have to determine the source of the material, the manner in which it was generated or retrieved and whether its integrity has been preserved.

5.3.2 Section 65B Principles

The former Indian Evidence Act, 1872 contained Section 65B, which specifically addressed the admissibility of electronic records. The Supreme Court's decisions in *Anvar P.V.* and *Arjun Panditrao Khotkar* made the certification requirement an important component of electronic-evidence jurisprudence.

However, following the enactment of the Bharatiya Sakshya Adhiniyam, 2023 (BSA), which replaced the Indian Evidence Act framework, researchers must carefully distinguish between the historical Section 65B jurisprudence and the corresponding provisions of the new evidentiary regime.

The principles developed under Section 65B remain highly relevant for understanding the judicial approach to electronic records, particularly with respect to authenticity, reliability and the manner of proving computer-generated material. However, contemporary research should cite the applicable provisions of the BSA for proceedings governed by the new law.

5.3.3 AI-Generated Evidence

AI introduces a further evidentiary difficulty because AI itself can generate, alter or analyse evidence. A forensic expert may use AI-based detection software to determine whether a video is synthetic. However, the output of an AI detection tool cannot automatically be equated with conclusive proof.

AI detection systems may produce false positives and false negatives. Their conclusions may also depend upon the dataset, algorithm, model and methodology used. Therefore, courts may need to examine: the reliability of the detection methodology; qualifications of the forensic expert; source and preservation of the digital material; chain of custody; technical methodology adopted; possibility of alteration after acquisition; and whether the expert's conclusion can be independently verified.

Indian courts have traditionally treated expert evidence as assistance to judicial determination rather than as an unquestionable substitute for judicial evaluation. The principle assumes greater importance in AI-related prosecutions because the technology used to detect manipulation may itself be technically complex and probabilistic.

5.4 Jurisdictional Issues

5.4.1 Territorial Jurisdiction

Cyber offences challenge the conventional concept that an offence occurs within a single physical territory. In a deepfake case, the person creating the content may be located in one State or country, the victim in another and the platform hosting the material elsewhere.

Indian law has attempted to address certain extra-territorial cyber activities through the Information Technology Act, 2000. Section 75 provides for application of the Act to offences or contraventions committed outside India in circumstances specified by the legislation.

At the same time, procedural jurisdiction under criminal law requires careful consideration of the place where the offence was committed, where its consequences occurred and where relevant evidence is located.

The Supreme Court's cyber-jurisdiction jurisprudence, including *Suhas Katti v. State of Tamil Nadu*, demonstrates the practical importance of establishing the territorial connection between online conduct and the jurisdiction in which criminal proceedings are initiated. Although the case predates contemporary deepfake technology, its significance lies in demonstrating that cyber offences can be prosecuted through evidence concerning electronic communications and their consequences.

In deepfake cases, jurisdiction may therefore arise from factors such as: location of the accused; location of the victim; place where the content was uploaded; place where the content was accessed; place where harm or deception occurred; and location of relevant computer systems or evidence.

5.4.2 International Cooperation

Effective investigation of transnational deepfake offences frequently depends upon cooperation between law-enforcement agencies in different countries. Traditional investigative procedures may become inadequate when evidence is stored by foreign service providers or when the accused operates outside Indian territory.

International cooperation may involve: requests for preservation of electronic evidence; disclosure of subscriber information; obtaining IP logs; production of account information; seizure or forensic imaging of foreign devices; examination of witnesses located abroad; and transfer of relevant evidence to Indian investigative authorities.

The increasing international nature of cybercrime therefore requires procedural mechanisms that permit rapid preservation of volatile electronic evidence.

5.4.3 Mutual Legal Assistance Treaties (MLATs)

Mutual Legal Assistance Treaties provide a formal mechanism through which one country can request another country to undertake investigative or evidentiary measures. In deepfake investigations, MLAT mechanisms may become relevant where essential evidence is located outside India.

However, traditional MLAT procedures may be time-consuming in comparison with the speed at which digital evidence can disappear. Social-media accounts may be deleted, cloud records may change and content may be repeatedly uploaded to different platforms.

Consequently, an effective framework for deepfake prosecution requires closer international cooperation and faster mechanisms for: preservation of electronic evidence; identification of anonymous users; authentication of foreign digital records; cross-border forensic cooperation; and lawful disclosure of platform-held information.

The future development of Indian cybercrime law must therefore address not only the substantive criminalisation of harmful AI-generated content but also the procedural mechanisms necessary to investigate and prosecute offenders operating across national boundaries.

VI. COMPARATIVE LEGAL ANALYSIS

The legal response to deepfakes differs considerably across jurisdictions because countries have approached the problem through different regulatory philosophies. In the United States, deepfake-related concerns have largely been addressed through a combination of existing civil and criminal remedies, targeted federal statutes, state-level legislative initiatives, and constitutional protections grounded in the First Amendment.

The European Union has adopted a broader regulatory framework that combines artificial intelligence governance with platform regulation and mandatory transparency. A comparative assessment of these regulatory models reveals that addressing deepfakes requires a careful balance between technological advancement and the protection of individual rights, freedom of expression, privacy, electoral processes, and society from digital harms.

6.1 United States

6.1.1 Deepfake Accountability Laws

A significant recent development is the TAKE IT DOWN Act, enacted as federal law in May 2025. The legislation specifically addresses the intentional online publication of non-consensual intimate visual depictions, including certain AI-generated or digitally altered images. The Act defines a “digital forgery” to include an intimate visual depiction created or modified through artificial intelligence, machine learning, software, or other technological means. It also establishes criminal prohibitions in specified circumstances and requires covered online platforms to establish a notice-and-removal mechanism for qualifying material.

The American approach also illustrates the limitations of technology-specific legislation. Deepfakes can be

employed for numerous purposes, including political manipulation, fraud, impersonation, harassment, defamation, identity misuse, and commercial exploitation. A law directed at one category of deepfake harm cannot automatically address the entire range of possible consequences. Consequently, conventional legal doctrines such as defamation, fraud, privacy, copyright, right of publicity, intentional infliction of emotional distress, and consumer-protection law continue to play an important role.

The federal approach is therefore best understood as harm-specific rather than technology-wide regulation. Instead of treating the creation of every synthetic representation as unlawful, the law tends to intervene when the use of the technology produces a legally recognizable injury.

6.1.2 State-Level Regulations

State legislation has particularly focused on election-related deepfakes and non-consensual sexual imagery. Some states have introduced restrictions on the creation or distribution of manipulated political content during specified periods surrounding elections, while others have created civil or criminal remedies for synthetic intimate imagery.

Election-related deepfake regulation illustrates this tension particularly clearly. Political deepfakes may undermine electoral processes by falsely portraying candidates as making statements or engaging in conduct that never occurred. However, legislation that prohibits false political expression must be drafted carefully because political communication receives particularly strong constitutional protection in the United States.

For example, Minnesota's restrictions on certain political deepfakes have faced constitutional litigation concerning their interaction with the First Amendment.

The resulting legal landscape is consequently fragmented. Some state laws emphasize prevention and removal, whereas others rely upon civil remedies or criminal sanctions. The diversity of these approaches demonstrates the difficulty of creating a uniform legal definition of a “harmful deepfake.”

6.1.3 First Amendment Concerns

The most distinctive feature of the American regulatory model is the constitutional significance of the First Amendment. Regulation of deepfakes cannot be evaluated solely on the basis of whether the material is false. American constitutional doctrine does not automatically place all false statements outside the protection of freedom of expression. The legal question is therefore whether a particular restriction serves a sufficiently important governmental interest and is appropriately tailored to the harm it seeks to prevent.

The constitutional challenge becomes even more significant where legislation regulates the **creation** of synthetic content rather than merely its fraudulent or harmful use. A blanket prohibition on creating artificial representations, without considering their intended purpose or context, could unintentionally restrict legitimate activities such as artistic experimentation, entertainment, journalism, academic research, and political satire.

Accordingly, American constitutional analysis tends to favour conduct-based regulation over unrestricted suppression of expression. Regulation is more constitutionally defensible where the deepfake is connected to a legally cognizable harm, such as fraud, non-consensual sexual exploitation, threats, impersonation, or other unlawful conduct.

The United States therefore demonstrates that effective deepfake regulation must take account of not only technological capability but also constitutional principles. Any future federal legal framework should carefully differentiate between deceptive conduct that causes harm and forms of expression protected by law, while ensuring effective legal remedies for individuals whose identities are used or manipulated without their consent.

6.2 European Union

The European Union has adopted a substantially different regulatory philosophy. Rather than relying principally upon isolated criminal or civil prohibitions, the EU has constructed a multi-layered regulatory system involving artificial intelligence regulation, online platform governance, risk management, and transparency requirements.

6.2.1 Artificial Intelligence Act

The European Union Artificial Intelligence Act (EU AI Act) represents a major development in the regulation of artificial intelligence. Rather than treating AI-generated content as a single category, the legislation establishes obligations according to the nature and risk of the AI system or application.

Article 50 requires relevant deployers of AI systems that generate or manipulate image, audio, or video

content constituting a deepfake to disclose that the content has been artificially generated or manipulated. The legal framework also recognizes the distinctive nature of artistic, creative, satirical, fictional, and similar forms of expression, ensuring that transparency obligations do not unduly restrict or impede their creation, dissemination, or enjoyment. This approach is significant because it does not necessarily prohibit the creation of a deepfake. Instead, it seeks to ensure that individuals encountering synthetic content are given information necessary to understand its nature. The regulatory emphasis is therefore placed on transparency and informed consumption.

The AI Act also reflects a broader European regulatory philosophy based on risk management. AI regulation is connected with fundamental rights, safety, accountability, and human oversight.

6.2.2 Digital Services Act

The Digital Services Act (DSA) works alongside the AI Act by establishing obligations and accountability mechanisms for online platforms and intermediary service providers. Its significance for deepfakes arises from the fact that the harm associated with synthetic content is frequently amplified through online distribution rather than through creation alone.

The DSA provides mechanisms through which the European Commission can investigate platform compliance and require access to information, including information concerning algorithms and recommender systems.

Recent enforcement activity demonstrates that the DSA is not merely a disclosure framework. The European Commission has investigated platform systems concerning manipulated sexually explicit material and has examined whether platforms adequately assessed and mitigated risks associated with the dissemination of such content.

The DSA consequently introduces an important concept into deepfake governance: platform responsibility. The legal focus is no longer restricted to the person who creates synthetic media. It also extends to the digital infrastructure through which harmful material is disseminated and amplified.

6.2.3 Transparency Obligations

Transparency constitutes one of the central pillars of the European model. The underlying principle is that users should not be required to determine independently whether highly realistic media is authentic when technology makes such determination increasingly difficult.

Under the AI Act, disclosure of artificially generated or manipulated deepfake content seeks to preserve the informational autonomy of individuals. A viewer who knows that an image or video has been synthetically generated can evaluate it differently from apparently authentic material.

The European approach therefore shifts part of the regulatory burden from prohibition to disclosure. Instead of attempting to eliminate every deepfake, the legal system seeks to establish conditions under which synthetic content can circulate while its artificial character remains identifiable.

The DSA adds another layer by requiring platforms to address systemic risks and maintain greater transparency concerning the operation of online systems. The European Commission's continuing enforcement activity demonstrates that transparency is increasingly treated as an operational obligation rather than merely a voluntary ethical principle.

However, transparency alone cannot eliminate deepfake-related harm. Labels may be ignored, removed, misunderstood, or technically circumvented. Moreover, once false content has been widely disseminated, a disclosure label may not fully repair reputational or electoral damage. Consequently, the European framework is strongest when transparency is combined with platform accountability, risk assessment, user remedies, and effective enforcement.

VII. NEED FOR A COMPREHENSIVE DEEPFAKE REGULATION FRAMEWORK IN INDIA

The rapid development of generative artificial intelligence has created a significant challenge for the Indian legal system. Deepfakes can be used for legitimate purposes such as entertainment, education, artistic expression, simulation, and research; however, the same technology can also facilitate impersonation, fraud, defamation, sexual exploitation, misinformation, electoral manipulation, and violations of privacy and personal dignity. The central difficulty is that existing Indian laws were largely developed before highly realistic AI-generated audio-visual content became easily accessible to the general public. Currently, legal action against the misuse of deepfake technology may be addressed through multiple legal frameworks, including information technology laws,

criminal statutes, intellectual property rights, privacy protections, defamation principles, and regulations governing online intermediaries. Although these provisions may apply to particular forms of harmful synthetic content, they do not constitute a unified legal framework specifically addressing the creation, dissemination, detection, attribution, and removal of deepfakes. This fragmented approach may create uncertainty regarding liability, enforcement, evidentiary standards, and remedies available to victims.

India therefore requires a comprehensive and technology-neutral regulatory framework capable of addressing deepfake-related harms without unnecessarily restricting legitimate innovation or freedom of expression. Such a framework should adopt a risk-based and harm-oriented approach, distinguishing between lawful synthetic media and synthetic content that causes or is reasonably likely to cause substantial harm.

7.1 Enactment of Specific Deepfake Legislation

7.1.1 Definition of Deepfake Offences

The first requirement for an effective regulatory framework is a clear statutory definition of a deepfake. The definition should be framed broadly enough to encompass AI-generated or materially AI-altered audio, video, images, and other digital content that deceptively represents a person, event, statement, or situation as genuine or authentic. However, the legislation should avoid treating every form of manipulated media as unlawful. A distinction should be made between creation of synthetic content and harmful use of synthetic content. For example, an AI-generated fictional character or clearly labelled parody should ordinarily not attract criminal liability merely because artificial intelligence was used in its creation.

Deepfake offences could therefore be structured around the intention, manner of dissemination, and resulting or foreseeable harm. Criminal liability may be considered where a person knowingly creates or distributes a synthetic representation for purposes such as fraud, impersonation, sexual exploitation, extortion, deliberate reputational injury, identity misuse, or serious interference with electoral processes.

The proposed legislation should also recognize aggravated forms of deepfake offences. Greater punishment may be justified where the victim is a child, where intimate material is involved, where the offender seeks financial gain, where the conduct forms part of organized cybercrime, or where the dissemination is intended to influence a significant public process.

A carefully drafted definition would reduce ambiguity and assist law-enforcement agencies, courts, technology companies, and victims in determining whether particular conduct falls within the regulatory framework.

7.1.2 Criminal Penalties

The criminal-law component of proposed deepfake legislation should establish proportionate penalties based upon the seriousness of the conduct. The law should distinguish between accidental or negligent dissemination and deliberate creation or distribution undertaken with knowledge of its falsity and harmful purpose.

Possible aggravating circumstances could include: financial or commercial exploitation; non-consensual sexual or intimate deepfakes; impersonation for obtaining money or property; threats, extortion, or blackmail; targeting of children or vulnerable persons; deliberate electoral manipulation; repeated dissemination after receiving notice of falsity; and coordinated dissemination through multiple accounts or platforms.

Punishment should not be based exclusively upon the technological method used. Instead, the severity of the penalty should correspond to the nature of the harm and the culpability of the offender. This would make the proposed legislation technologically neutral and capable of remaining relevant as generative-AI techniques evolve.

At the same time, criminal law should remain a measure of last resort. Excessively broad criminalization may discourage legitimate technological experimentation and creative expression. Accordingly, the legal framework should include adequate safeguards to prevent criminal provisions from being misused against individuals engaged in ordinary disputes, satire, journalistic activities, or bona fide criticism.

7.1.3 Civil Remedies

Criminal prosecution alone cannot adequately address the consequences suffered by deepfake victims. A person whose face or voice has been manipulated may experience reputational damage, professional consequences, social stigma, financial loss, and

emotional distress even when the perpetrator is never successfully prosecuted.

A dedicated statutory framework should therefore provide accessible civil remedies. Depending upon the circumstances, courts could be empowered to grant: injunctions preventing further dissemination; orders requiring removal or disabling of access; compensation for financial and reputational loss; damages for infringement of privacy and dignity; corrective or clarification notices in appropriate cases; and orders requiring preservation of relevant digital evidence.

Civil liability should also extend, in appropriate circumstances, to persons or entities that knowingly continue distributing unlawful synthetic content after receiving legally sufficient notice.

The availability of civil remedies would provide victims with a faster avenue of relief than waiting for the conclusion of a criminal prosecution.

7.2 Mandatory AI Content Disclosure

7.2.1 Watermarking AI-Generated Content

Compulsory disclosure of AI-generated content can strengthen transparency and reduce the potential for deception, as sophisticated synthetic material may closely resemble authentic content and therefore be difficult for users to identify as artificially created. A legal requirement for appropriate identification would therefore increase transparency and enable users to make more informed judgments.

India could introduce a standardized system for identifying AI-generated or substantially AI-manipulated content. Such identification may take the form of visible labels, machine-readable metadata, cryptographic provenance information, or other reliable technical indicators.

Watermarking should ideally operate at two levels. Visible disclosure would provide an immediately understandable indication to ordinary users, while technical provenance information could assist platforms, investigators, and forensic experts in establishing the origin and history of the material.

However, watermarking should not be treated as a complete solution. Sophisticated actors may attempt to remove or manipulate watermarks. Consequently, disclosure requirements should be supported by technical standards, platform-level verification systems, and appropriate penalties for deliberate removal or falsification of required identifiers.

7.2.2 Detection Mechanisms

India should also encourage the development of reliable deepfake detection mechanisms. Detection technologies can examine inconsistencies in facial movement, audio characteristics, image structure, metadata, compression patterns, or other technical indicators. However, no detection system should be regarded as infallible.

AI technology is developing rapidly, and detection tools may themselves become less reliable as generative systems improve. The legal framework should therefore avoid establishing a rule that automatically treats the output of one technological detector as conclusive proof.

Instead, detection should function as one component of a broader authentication process. Where the authenticity of material is disputed in legal proceedings, technical examination should be supplemented by provenance information, source records, witness testimony, device evidence, metadata, and expert opinion.

Government agencies, universities, forensic institutions, and private technology organizations should be encouraged to cooperate in developing standardized Indian protocols for deepfake detection.

7.3 Platform Responsibility

7.3.1 Due Diligence Obligations

Online platforms play a critical role in determining the extent to which deepfake content reaches the public. A harmful synthetic video may cause relatively limited damage if it remains private, but its consequences can become substantial when recommendation algorithms and social-media sharing mechanisms distribute it to thousands or millions of users.

Platform responsibility should therefore constitute an important part of India's deepfake regulatory framework.

Platforms could be required to adopt reasonable due-diligence measures, including: mechanisms for identifying suspected synthetic content; systems for receiving complaints from affected individuals; procedures for verifying high-risk content; preservation of relevant evidence following a lawful complaint; cooperation with authorized investigative agencies; transparency regarding content-moderation practices; and appropriate measures against repeated malicious dissemination.

The obligations should be proportionate to the size, functionality, and systemic influence of the platform. A small online service should not necessarily bear the same compliance burden as a platform with hundreds of millions of users.

Such a differentiated approach would protect innovation while ensuring that entities possessing greater technical and organizational capacity assume greater responsibility for preventing foreseeable harm.

7.3.2 Faster Removal Mechanism

Deepfake-related harm is often highly time-sensitive. A defamatory or sexually explicit synthetic video may be copied and redistributed across multiple platforms within minutes. Consequently, a remedy that becomes available only after several weeks may be practically ineffective.

India should consider establishing a rapid-response mechanism for clearly unlawful deepfake content. Victims should have access to an easily identifiable reporting channel through which they can submit complaints and supporting evidence.

For high-risk categories—particularly non-consensual intimate deepfakes, child sexual content, extortion-related material, impersonation used for financial fraud, and content presenting an immediate and serious threat—platforms should be required to act within a significantly shorter period after receiving a sufficiently substantiated complaint.

At the same time, expedited removal procedures should contain safeguards against false complaints and abuse. Platforms should maintain records of decisions and provide appropriate review mechanisms where content has been removed erroneously.

The objective should be to create a system in which speed and procedural fairness operate together.

7.4 Strengthening Digital Evidence Law

Deepfake litigation presents a distinctive evidentiary challenge. Traditional digital evidence principles generally focus on establishing the authenticity and integrity of electronic records. With generative AI, however, the central question may be whether the electronic material itself is artificially generated or manipulated.

This creates a need to strengthen India's digital-forensics infrastructure and evidentiary procedures.

7.4.1 AI Forensic Laboratories

India should establish or strengthen specialized forensic laboratories capable of examining AI-generated and manipulated media. Such laboratories should have expertise in: synthetic-image detection; synthetic-video analysis; voice-cloning identification; metadata examination; digital provenance analysis; cryptographic verification; device and source analysis; and reconstruction of digital manipulation.

Specialized AI forensic facilities would reduce dependence on general digital-forensics infrastructure that may not possess adequate expertise in rapidly evolving generative technologies.

These laboratories should also develop standardized procedures for acquisition, preservation, examination, and reporting of suspected deepfake evidence. Standardization is essential because inconsistent forensic methodologies may produce conflicting expert conclusions and reduce judicial confidence in technical evidence.

7.4.2 Expert Authentication

Courts may increasingly encounter disputes concerning whether an audio recording, photograph, or video is authentic or AI-generated. Judges and investigating officers may not always possess the technical expertise necessary to resolve such disputes independently.

Accordingly, the legal system should encourage the use of appropriately qualified experts in AI and digital forensics. Expert evidence should explain not only the conclusion reached but also the methodology used, the material examined, the limitations of the technique, and the degree of confidence attached to the finding.

Importantly, expert authentication should not replace judicial assessment. A forensic opinion should assist the court rather than become automatically conclusive. Courts must exercise caution by recognizing that AI-based detection tools may produce both erroneous positive findings and erroneous negative findings. A sophisticated legal framework must therefore evaluate the reliability of the particular forensic method used in each case.

7.5 Victim Protection Mechanisms

Deepfake regulation should ultimately be measured not only by the number of prosecutions achieved but also by the effectiveness of protection available to victims. Victims may face severe reputational, economic, professional, social, and psychological consequences even before legal proceedings commence.

7.5.1 Compensation

A comprehensive framework should establish meaningful compensation mechanisms for victims. Compensation may address direct financial loss, professional damage, reputational injury, privacy violations, and other demonstrable consequences.

In cases involving commercial exploitation, fraud, or deliberate dissemination for financial benefit, courts should have appropriate authority to consider the economic gain obtained by the offender when determining compensation or other monetary relief.

A victim should not be required to establish every consequence with mathematical precision where the nature of the harm is inherently difficult to quantify. Nevertheless, safeguards should ensure that compensation remains based on legally recognized injury and evidence.

7.5.2 Immediate Takedown Orders

Victim protection requires remedies that operate before irreversible harm occurs. Once a deepfake has become widely distributed, subsequent removal may not eliminate copies already downloaded or redistributed by other users.

The legal framework should therefore permit courts or competent authorities to issue urgent takedown and disabling-access orders in appropriate circumstances. Such orders could be particularly important for non-consensual intimate deepfakes, child-related material, extortion, threats, and other high-risk categories.

An effective system should also permit applications through simplified procedures where delay itself may substantially increase the harm. At the same time, judicial or administrative oversight should be maintained to prevent arbitrary suppression of lawful speech.

7.5.3 Psychological Support

The impact of deepfake victimization is not limited to legal or financial injury. Victims may experience humiliation, social isolation, professional anxiety, fear of continued exposure, and loss of personal control over their identity.

A comprehensive legal framework should therefore incorporate referral mechanisms for psychological and social support. Victims should be informed about available legal remedies, evidence-preservation procedures, reporting mechanisms, and appropriate support services.

Particular attention should be given to minors and victims of sexually explicit deepfakes. For these categories, the legal response should prioritize confidentiality, dignity, and prevention of secondary victimization.

7.6 Proposed Model for India

India should consider adopting a multi-layered Deepfake Regulation Framework rather than relying upon a single regulatory measure. The proposed model may be structured around five interconnected pillars:

First, substantive regulation:

legislation should clearly identify serious forms of harmful deepfake conduct and establish proportionate criminal and civil consequences.

Second, transparency:

AI-generated and substantially manipulated content should, where appropriate, carry reliable disclosure or provenance information.

Third, platform accountability:

online intermediaries should undertake proportionate due diligence, maintain effective complaint mechanisms, and provide expedited responses to high-risk unlawful synthetic content.

Fourth, evidentiary preparedness:

courts and investigative agencies should have access to specialized AI forensic laboratories, trained experts, and standardized authentication protocols.

Fifth, victim-centred remedies:

affected persons should receive rapid takedown mechanisms, compensation where legally justified, preservation of evidence, and appropriate psychological and legal support.

This model would allow India to move from a largely reactive approach to a preventive and rights-oriented regulatory system. Regulation should not attempt to prohibit synthetic media as such. Instead, it should identify the circumstances in which synthetic media creates unacceptable risks to individual rights, public institutions, and society.

7.7 Bill For Specific Deepfake Legislation

THE DEEPFAKE PREVENTION, REGULATION AND ACCOUNTABILITY BILL, 2026

A Bill

to provide for the prevention, regulation and penalisation of the creation, publication, distribution and misuse of deepfake and other materially deceptive synthetic media; to protect individual dignity, privacy, reputation, identity, intellectual property and electoral integrity; to impose appropriate duties upon users,

developers and intermediaries; to establish mechanisms for detection, reporting, removal and redressal of unlawful deepfake content; and for matters connected therewith or incidental thereto.

PREAMBLE

WHEREAS advances in artificial intelligence, machine learning and generative technologies have enabled the creation of highly realistic synthetic audio, visual and audiovisual material;

AND WHEREAS such technology may be used for legitimate purposes including education, research, entertainment, artistic expression and innovation, but may also be misused to impersonate individuals, disseminate false information, facilitate fraud, violate privacy and dignity, manipulate public discourse and cause serious social and economic harm;

AND WHEREAS existing legal remedies may address particular consequences of deepfake misuse but may not always provide a comprehensive and technology-specific framework governing its creation, dissemination, attribution, detection and removal;

AND WHEREAS the Information Technology Rules have increasingly addressed synthetic or AI-generated information, demonstrating the need for transparency and intermediary responsibility in the digital environment;

AND WHEREAS it is necessary to establish a balanced legal framework that prevents harmful manipulation without unnecessarily restricting legitimate technological innovation, artistic expression, journalism, academic research or constitutionally protected speech;

BE it enacted by the Legislature as follows:

CHAPTER I

PRELIMINARY

1. Short title, extent and commencement

1. This Act may be called the Deepfake Prevention, Regulation and Accountability Act, 2026.
2. It shall extend to the whole of India.
3. It shall come into force on such date as the Central Government may, by notification in the Official Gazette, appoint.

2. Definitions

In this Act, unless the context otherwise requires—

(a) "artificial intelligence system" means a computational system capable of generating, modifying, analysing or manipulating digital content through automated or machine-learning techniques;

(b) "deepfake" means any digitally created, altered or synthesised image, video, audio, voice, text or combination thereof which substantially imitates the appearance, identity, voice, conduct or characteristics of a real person, event or object in a manner capable of misleading a reasonable person as to its authenticity or origin;

(c) "deepfake content" includes synthetic or materially manipulated content that falsely represents a person as having spoken, acted, participated in or endorsed an event, statement, transaction or activity which did not actually occur;

(d) "creator" means any person who creates, generates, materially modifies or commissions the creation of deepfake content;

(e) "distributor" means any person who knowingly publishes, uploads, transmits, forwards, sells, advertises or otherwise makes deepfake content available to another person;

(f) "intermediary" shall have the meaning assigned to it under the applicable information technology law;

(g) "consent" means free, specific, informed and voluntary permission given by a person for the creation or use of synthetic representation of their identity, likeness, voice or personal attributes;

(h) "materially deceptive content" means synthetic or manipulated content which is reasonably capable of causing a person to believe that an artificial representation is authentic and which is likely to cause legally recognisable harm;

(i) "victim" means a person whose identity, likeness, voice, reputation, privacy, property or legally protected interests have been adversely affected by unlawful deepfake content;

(j) "synthetic media" means digital content wholly or substantially generated or altered through artificial intelligence or other automated computational techniques.

CHAPTER II

PROHIBITED DEEPFAKE ACTIVITIES

3. Prohibition of malicious creation of deepfake content

No person shall knowingly create or materially alter synthetic media for the purpose of—

1. deceiving another person;
2. impersonating another person;
3. obtaining an unlawful financial or other advantage;
4. causing injury to reputation;

5. violating privacy or personal dignity;
6. facilitating fraud or identity theft;
7. interfering unlawfully with an election or democratic process; or
8. causing any other legally recognisable harm.

4. Prohibition of non-consensual intimate deepfake content

1. No person shall create, publish, transmit or distribute sexually explicit synthetic content depicting an identifiable person without that person's consent.
2. Consent to the creation or use of an ordinary image, video or recording shall not be deemed to constitute consent to the creation or distribution of sexually explicit synthetic content.
3. The offence under this section shall apply irrespective of whether the underlying image, voice or likeness was originally available publicly.

5. Deepfake impersonation for fraud

Any person who knowingly uses deepfake technology to impersonate another person for the purpose of obtaining money, property, confidential information, authentication credentials or any other unlawful benefit shall be guilty of an offence under this Act.

6. Deepfake-related identity misuse

A person shall be guilty of an offence where such person knowingly uses another individual's face, voice, biometric characteristics or other distinctive identity attributes in synthetic media without lawful authority and with the intention of deceiving, exploiting or causing substantial harm to that individual.

7. Electoral and democratic interference

No person shall knowingly create or disseminate a materially deceptive deepfake representing a candidate, public office holder, election authority or other relevant participant in an electoral process where the publication is intended to materially influence electoral conduct through deception.

Nothing in this section shall prohibit legitimate political criticism, satire, parody, commentary or journalistic reporting where the synthetic nature of the content is reasonably disclosed.

CHAPTER III

TRANSPARENCY AND DISCLOSURE

8. Mandatory disclosure of synthetic content

1. A person who publishes or distributes synthetic media shall provide a clear and reasonably noticeable disclosure where the content has been materially generated or manipulated by artificial intelligence.
2. The disclosure shall be sufficiently visible or audible to inform an ordinary viewer or listener that the content is synthetic or materially altered.
3. The disclosure requirement shall not apply in the same manner to content used exclusively for research, testing, education, satire, parody, fiction or artistic purposes where the context itself clearly communicates that the content is not authentic.

9. Digital provenance and watermarking

The Central Government may prescribe standards requiring specified categories of AI-generated or materially manipulated content to contain—

(a) machine-readable provenance information;

(b) digital watermarking;

(c) metadata identifying synthetic generation or material alteration; or

(d) other technically appropriate authenticity indicators.

Such standards shall, as far as practicable, remain technology-neutral and shall not mandate dependence upon a single detection technology.

CHAPTER IV

DUTIES OF AI DEVELOPERS AND INTERMEDIARIES

10. Duties of developers and providers of deepfake-generating systems

Providers of AI systems capable of generating realistic synthetic media shall adopt reasonable safeguards proportionate to the foreseeable risks associated with their systems, including—

1. mechanisms to discourage unlawful impersonation;
2. safeguards against the generation of non-consensual intimate content;
3. appropriate user reporting mechanisms;
4. preservation of relevant technical records where legally required; and
5. cooperation with lawful investigations.

11. Duties of intermediaries

An intermediary shall, in accordance with applicable law—

1. establish an accessible mechanism for reporting unlawful deepfake content;
2. provide an expedited mechanism for complaints involving impersonation, sexual exploitation, fraud and serious threats to personal safety;

3. take appropriate action upon receiving a valid legal order or otherwise becoming legally obligated to remove unlawful content;
4. preserve relevant evidence for a prescribed period where required by law; and
5. provide appropriate information to authorised investigative agencies subject to lawful procedure and privacy safeguards.

The obligations under this section shall be implemented in a manner that does not impose a general obligation upon intermediaries to monitor all user-generated content.

CHAPTER V

DEEPFAKE DETECTION AND REPORTING

12. Deepfake reporting mechanism

1. Every designated intermediary shall establish a readily accessible reporting mechanism for suspected unlawful deepfake content.
2. Complaints involving intimate synthetic content, financial fraud, identity impersonation or threats to personal safety shall receive priority treatment.
3. The complainant shall receive an acknowledgement and, where practicable, information regarding the status of the complaint.

13. Independent verification

Where the authenticity of disputed content is material to legal proceedings, an appropriate forensic examination may be undertaken by an accredited digital forensic laboratory.

No person shall be convicted solely on the basis of an automated deepfake-detection result without adequate corroborative evidence.

CHAPTER VI

OFFENCES AND PENALTIES

14. General penalty

Any person who knowingly creates, publishes or distributes unlawful deepfake content in contravention of this Act shall be punishable with imprisonment which may extend to three years, or with fine which may extend to ₹5,00,000, or with both.

15. Aggravated offences

Where the offence—

(a) involves sexually explicit content;

(b) is committed against a minor;

(c) is committed for financial gain;

(d) involves organised or repeated conduct;

(e) causes substantial financial loss;

(f) is intended to manipulate an election; or

(g) results in serious and foreseeable harm to the victim,

the offender may be punished with imprisonment which may extend to **seven years** and fine which may extend to ₹10,00,000.

16. Offences against children

Where deepfake technology is used to create, possess, transmit or distribute sexually explicit synthetic material involving a child, the matter shall additionally attract the provisions of applicable child-protection and criminal laws.

17. Corporate liability

Where an offence under this Act has been committed by a company or other body corporate, every person who, at the time of commission of the offence, was responsible for the conduct of its business shall be deemed liable, unless such person proves that the offence was committed without their knowledge or that reasonable preventive measures were exercised.

CHAPTER VII

RIGHTS AND REMEDIES OF VICTIMS

18. Right to removal and correction

A victim of unlawful deepfake content shall have the right to seek—

1. removal or disabling of access to the unlawful content;
2. correction or clarification where appropriate;
3. preservation of evidence;
4. identification of the responsible person through lawful process; and
5. compensation for demonstrable loss.

19. Civil remedies

A competent court may grant—

(a) interim and permanent injunctions;

(b) removal or disabling orders;

(c) damages for financial and reputational injury;

(d) compensation for infringement of privacy and dignity;

(e) orders requiring correction or clarification; and

(f) any other relief necessary to prevent continuing harm.

20. Emergency protection

Where publication of deepfake content presents an immediate and substantial threat to the safety, dignity, privacy or financial interests of a person, the competent authority or court may issue an expedited order for temporary restriction or removal, subject to subsequent review.

CHAPTER VIII

INVESTIGATION AND JURISDICTION

21. Investigation of offences

Offences under this Act shall be investigated by an officer authorised under the applicable law and possessing appropriate training in cybercrime and digital forensics.

22. Preservation of electronic evidence

Investigating authorities may require preservation of relevant electronic evidence in accordance with applicable law, subject to judicial safeguards and constitutional protections.

23. Jurisdiction over offences committed outside India

Where an offence under this Act is committed outside India but involves a victim, computer resource, digital platform or substantial consequence within India, the provisions of this Act may apply to the extent permitted by applicable law.

CHAPTER IX

SAFEGUARDS FOR FUNDAMENTAL RIGHTS

24. Protection of legitimate expression

Nothing in this Act shall be interpreted to prohibit or criminalise, merely because synthetic technology is used—

(a) legitimate artistic expression;

(b) satire or parody;

(c) academic research;

(d) scientific experimentation;

(e) journalism and public-interest reporting;

(f) political criticism; or

(g) fictional or clearly labelled entertainment content.

However, the protection under this section shall not extend to conduct that independently constitutes an offence under this Act or any other applicable law.

25. Proportionality and due process

Any restriction, removal order, investigation or penalty under this Act shall be proportionate to the nature and seriousness of the alleged harm and shall be subject to applicable procedural safeguards.

CHAPTER X

DEEPFAKE REGULATORY AUTHORITY

26. Establishment of Authority

The Central Government may establish a specialised body to be known as the Deepfake Regulation and Digital Authenticity Authority for the purposes of—

1. developing regulatory standards;
2. coordinating with law-enforcement agencies;
3. promoting digital forensic capacity;
4. issuing technical guidance;
5. facilitating research on synthetic media;
6. maintaining standards for provenance and authenticity technologies; and
7. publishing periodic reports regarding deepfake-related risks and enforcement.

27. Advisory and technical committees

The Authority may constitute committees consisting of experts in—

- law;
- artificial intelligence;
- cybersecurity;
- digital forensics;
- human rights;
- media and journalism;
- child protection; and
- constitutional law.

CHAPTER XI

OFFENCES BY PERSONS ACTING IN BAD FAITH

28. False complaints

A person who knowingly and maliciously files a false complaint under this Act with the intention of suppressing lawful speech, damaging another person's reputation or causing unlawful economic loss may be liable to appropriate penalties prescribed by rules.

No person shall be penalised merely because a complaint is ultimately found to be unsubstantiated, provided that it was made honestly and in good faith.

CHAPTER XII

RULE-MAKING AND REVIEW

29. Power to make rules

The Central Government may, by notification, make rules for carrying out the purposes of this Act, including rules concerning—

(a) disclosure and labelling standards;

(b) technical provenance mechanisms;

(c) reporting and grievance procedures;

(d) preservation of electronic evidence;

(e) accreditation of forensic laboratories;

(f) duties of specified categories of intermediaries; and

(g) procedures for victim assistance.

30. Periodic review

The Central Government shall review the operation of this Act at reasonable intervals in consultation with relevant stakeholders, including technology experts, civil society organisations, legal professionals, digital platforms and academic institutions.

CHAPTER XIII
MISCELLANEOUS
31. Act to operate in addition to other laws
The provisions of this Act shall be in addition to, and not in derogation of, any other law relating to privacy, defamation, fraud, identity theft, sexual offences, child protection, intellectual property, information technology or criminal liability.
32. Protection of action taken in good faith
No legal proceeding shall lie against an authorised officer for anything done in good faith in pursuance of this Act.
33. Power to remove difficulties
If any difficulty arises in giving effect to the provisions of this Act, the Central Government may, consistent with the provisions of this Act, take such measures as may be necessary to remove the difficulty.

STATEMENT OF OBJECTS AND REASONS
The rapid development of generative artificial intelligence has substantially increased the ability of individuals and organisations to create realistic synthetic images, videos, voices and other forms of digital media. Although such technologies have legitimate applications, their malicious use can facilitate impersonation, fraud, misinformation, non-consensual intimate imagery, reputational injury, identity misuse and interference with democratic processes.

India has progressively addressed synthetic and AI-generated content through the information technology regulatory framework. The Ministry of Electronics and Information Technology has introduced amendments concerning synthetically generated information, reflecting the growing regulatory importance of transparency, intermediary responsibility and digital-content governance.
However, a comprehensive statutory framework specifically addressing deepfake-related conduct can provide greater clarity concerning offences, liability, victim remedies, evidentiary standards, disclosure obligations and institutional responsibility.
The proposed legislation therefore seeks to establish a dedicated legal framework based on five principal objectives:

1. Prevention of malicious deepfake creation and dissemination;
2. Transparency regarding synthetic and materially manipulated content;
3. Accountability of persons and entities responsible for unlawful use;
4. Protection and remedies for individuals whose rights are violated; and
5. Protection of legitimate expression and technological innovation through appropriate exceptions and procedural safeguards.

The proposed framework adopts a harm-based approach rather than prohibiting artificial intelligence technology itself. This distinction is important because the objective is to regulate harmful conduct while preserving legitimate technological, artistic, educational, scientific and journalistic uses.

FINANCIAL MEMORANDUM
The implementation of this legislation may require expenditure for the establishment of regulatory, technical and forensic infrastructure, including digital forensic laboratories, reporting mechanisms, training programmes and technological research.

The exact financial implications shall be determined by the Government after consultation with relevant authorities and stakeholders.

VIII. CONCLUSION

The proposed Deepfake Prevention, Regulation and Accountability Act, 2026 seeks to fill the conceptual and institutional gap between existing technology regulation and the rapidly evolving harms associated with synthetic media. Rather than imposing an absolute prohibition on deepfake technology, the Bill adopts a differentiated approach based on intent, deception, consent, harm, transparency and proportionality.

Such an approach is particularly important for India because regulation of synthetic content must reconcile competing interests: protection of individual dignity and privacy, prevention of fraud and misinformation, technological innovation, freedom of speech and expression, journalistic activity and democratic participation. The Bill therefore attempts to create a framework in which harmful manipulation is subject to meaningful legal consequences while legitimate uses of artificial intelligence remain protected

PRELIMINARY BIBLIOGRAPHY (SLIC STYLE)

[1] Danielle K. Citron, Sexual Privacy, 128 Yale Law Journal 1870 (2019).
[2] Bobby Chesney & Danielle Citron, “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security”, 107 California Law Review 1753 (2019).
[3] Robert Chesney and Danielle Citron, “Deepfakes and the New Disinformation War”, Foreign Affairs, 2019.
[4] Information Technology Act, 2000.
[5] Digital Personal Data Protection Act, 2023.
[6] Bharatiya Nyaya Sanhita, 2023.
[7] United Nations, Declaration of Basic Principles of Justice for Victims of Crime and Abuse of Power, 1985.
[8] European Union, Artificial Intelligence Act, 2024.
[9] NIST, Artificial Intelligence Risk Management Framework, 2023.
[10] European Union, Digital Services Act, 2022.