DEEPFAKE TECHNOLOGY AND THE LAW: ADDRESSING EMERGING THREATS IN THE DIGITAL ERA

Aditya Singh Rajput, Amity Law School, Amity University, Patna, Bihar.

ABSTRACT

Deepfake technology, powered by artificial intelligence (AI) and machine-learning methods, has regenerated the production and alteration of digital audio-visual content. While deepfakes have valid uses in areas like entertainment, education and communication their harmful application has led to serious legal challenges. These include individualism theft, defamation, privacy breaches, fraud, spread of false information, impersonation and the creation of intimate content without consent. The growing complexity of fake media has made it challenging to tell difference between real content from fake content, creating problems for individual rights and reliability of digital information. In India, there is currently no single law that specifically governs deepfakes. Instead their misuse could fall under provisions of the Information Technology Act, 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 and the Bharatiya Nyaya Sanhita, 2023.For example, Section 356 of the Bharatiya Nyaya Sanhita, 2023 deals with defamation through words, signs, or visual representations, which can be relevant in cases that involves harmful deepfake content.[1] The Digital Personal Data Protection Act, 2023 also provides a legal framework for the handling and protection of digital personal information, including rights relating to correction and remove data.[2] The regulatory approach has developed alongside technological progress. The Ministry of Electronics and Information Technology (MeitY) has issued guidelines on synthetic media and deepfakes, including requirements for labelling and identifying content that has been artificially created.[3][4] Further, the Information Technology Rules were updated in February 2026 to include specific provisions on synthetically generated information (SGI), showing a greater focus on responsibility and transparency for platforms. [5][6] This paper examines the new risks brought by

[1] Bharatiya Nyaya Sanhita, 2023, § 356, India Code.
[2] Digital Personal Data Protection Act, 2023, §§ 4–15, India Code
[3] Ministry of Electronics and Information Technology, Advisory on Deepfakes/Synthetic Content, Dec. 26,
[4] (requiring intermediaries facilitating synthetic creation to provide appropriate labelling or permanent metadata/identifiers).
[5] Ministry of Electronics and Information Technology, Advisory on Deepfakes/Synthetic Content, Dec. 26,
[6] (requiring intermediaries facilitating synthetic creation to provide appropriate labelling or permanent metadata/identifiers).

deepfake technology and imposes whether India's current legal system is sufficient. It looks into issues like privacy, freedom of expression, responsibilities of online platforms, defamation, consent, digital evidence and support for victims. The study argues that effective regulation needs a balanced and efficient strategy that includes clear legal rules, accountability for online platforms, tools to detect artificial content, transparency and public awareness, while making sure that genuine uses of AI and constitutional freedoms are not overly restricted.

Keywords: Deepfake Technology, Artificial Intelligence, Synthetic Media, Synthetically Generated Information (SGI), Cyber Law, Data Protection, Digital Evidence, Misinformation, Identity Theft, Non-Consensual Intimate Imagery, Defamation, Financial Fraud, AI Regulation, Technology Governance.

INTRODUCTION

The fast growth of Artificial Intelligence (AI), especially generative AI, has changed how digital content is made, shared and used. One of the most notable examples of this advancement is deepfake technology which allows people to create or alter audio, images and also videos to make very realistic versions of individuals, events or statements that may not have actually happened. While synthetic media has useful uses in areas like entertainment, education, accessibility, advertising and creative fields, its improper use has caused major worries about privacy, reputation, identity, false information, fraud, fair elections and personal freedom. [7]

Deepfakes are legally important because they make it hard to tell real information from fake.[8] A altered video could show someone saying something they never did in reality and fake audio might copy someone's voice to trick others or commit fraud.AI-made intimate images can also be created without the person's knowledge or permission. These issues challenge traditional legal ideas because current laws were made for real people and basic digital changes, not for advanced synthetic content.

In India, the harms caused by deepfakes are dealt with using existing cyber,

[7] Information Technology Act, No. 21 of 2000, India Code; Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021; Bharatiya Nyaya Sanhita, No. 45 of 2023, India Code.
[8] Radha Ranjan, "Deep Fake Porn: Duplicity and Dystopia," in *Law in the Age of Disruption: Understanding the Impact of Technology* (Lex Assisto Media & Publications, 2024).

criminal, privacy and intermediary rules, not by a single law focused only on deepfakes.[9]

The Information Technology Act, 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, and the Bharatiya Nyaya Sanhita, 2023 could be relevant depending on the nature of the actions and harm involved.[10] Recognizing the increasing dangers of AI-generated and altered content, the Ministry of Electronics and Information Technology (MeitY) updated the IT Rules in February 2026 to include a special framework for "synthetically generated information" (SGI).[11]

This new framework is a major step in Indian cyber law. However, technology continues to develop quickly, making it a constant challenge for lawmakers, courts, regulators, tech companies and individuals. The key issue is not just whether deepfakes can be controlled, but whether the current legal tools are clear, effective and balanced with constitutional rights to deal with the special problems caused by synthetic media.

1. RESEARCH BACKGROUND AND THE EVOLUTION OF SYNTHETIC MEDIA

Synthetic media refers to digital content that is created, altered, or modified using computational technologies, especially artificial intelligence and machine learning.[12]

Deepfakes are among the most advanced forms of synthetic media because they can alter a person's face, voice, expressions or actions in a way that looks real. The advancement of deep-learning techniques and generative models has significantly lowered the level of technical skill needed to produce such content.

The development of synthetic media can be seen as a shift from basic digital editing to more autonomous and realistic AI-generated content. In the past, manipulation of digital content typically required specialized skills and a lot of human involvement. Now, modern

[9] Radha Ranjan et al., "Policies Against Fraud and Cybercrime: Strategic, Legal, and Technological Approaches," *Policies Against Fraud and Cybercrime: Strategic, Legal, and Technological Approaches* (IGI Global, 2026), DOI: 10.4018/979-8-3373-5992-2.ch007.

[10] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Gazette of India, G.S.R. 120(E) (Feb. 10, 2026).

[11] Ministry of Elecs. & Info. Tech., Frequently Asked Questions: Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, 1–2 (Feb. 10, 2026).

[12] Radha Ranjan Jivantika Gulati, "Catalytic Role of Artificial Intelligence in Ameliorating Criminal Justice System: A Critical Analysis," in *Empowerment and Equality: Navigating Human Rights Law in a Complex World* (Institute of Legal Education, 2023), pp. 47–54.

generative AI systems can create realistic images, clone voices, change facial expressions, and produce synthetic videos with much less technical knowledge.

This technological progress brings both benefits and challenges. When used responsibly, synthetic media can support education, accessibility, multilingual communication, creative industries, simulations and other useful activities. MeitY acknowledges that synthetic information can help drive innovation, improve accessibility, enable personalized learning, enhance public services, and promote economic growth. [13] However, the same technology can also be misused to spread false information, impersonate people, commit financial fraud, produce non-consensual intimate content, harm reputations, and influence public opinion.[14]

As synthetic media becomes more realistic, it creates a specific challenge in terms of evidence. Visual or audio material that was once considered reliable may no longer be trusted automatically. This change has impacts not only on individuals but also on journalism, democratic processes, law enforcement, and legal systems. Therefore, regulating synthetic media has become a key issue in modern cyber law and technology governance.

India's regulatory approach has been shaped by these changes. In October 2025, MeitY introduced amendments to the IT Rules, focusing on synthetic content and highlighting deepfakes as a major source of misinformation, privacy breaches, impersonation, financial fraud and other problems. [15] These proposals led to the Information Technology

(Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, which were officially published on February 10, 2026, and came into effect on February 20, 2026.[16]

2. STATEMENT OF THE PROBLEM: THE LEGAL LAG IN TECHNOLOGY REGULATION

[13] Ministry of Elecs. & Info. Tech., Explanatory Note, supra note 1, at 1–2.
[14] Ministry of Elecs. & Info. Tech., Explanatory Note, supra note 1, at 1–2.
[15] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026); Ministry of Elecs. & Info. Tech., Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (updated Feb. 10, 2026).
[16] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026); Ministry of Elecs. & Info. Tech., Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (updated Feb. 10, 2026).

The main issue this research focuses on is the delay between the development of technology and the creation of laws to regulate it. Technology often advances much faster than new laws can be made. By the time a specific type of digital harm is recognized and addressed through laws or court rulings, new ways of manipulation may have already been developed through technological progress.

Deepfakes clearly show this issue. Traditional legal terms like defamation, fraud, impersonation, privacy violation, obscenity and forgery might be used to deal with some forms of deepfake misuse, but these legal terms were not created specifically for AI- generated media. Therefore, there are questions about how to properly categorize deepfake actions, who is responsible, what standards of evidence are needed, the responsibility of online platforms, which legal system applies, and what kind of support victims can get.

Another major challenge is figuring out who is responsible among the various people involved in creating and sharing synthetic media. A harmful deepfake might be created by one person, uploaded by another, distributed by a platform, developed using an AI system, and then shared by other users. Determining how much each party should be held responsible is a difficult legal and policy issue.

This problem is made even more complicated by the need to balance regulation with the rights protected by the Constitution. Overly strict rules on synthetic content might limit legal forms of artistic expression, satire, political speech, research, journalism, and technological progress. However, not enough regulation could allow deepfakes to do serious and lasting damage to people and society.

The 2026 changes to the IT Rules aim to close this regulatory gap by introducing requirements for synthetically generated content, including ways to help users identify such content and improve the responsibility of online platforms.[17] However, questions still remain about how effectively these rules can be applied in practice, the accuracy of the technology used, the involvement of international platforms, the support available to victims, privacy concerns, the freedom to express oneself and the ability of regulators to keep up with fast-changing AI systems. Therefore, the main problem is not just the lack of laws, but whether current legal methods are sufficient, flexible, and enforceable in dealing with quickly

[17] Ministry of Elecs. & Info. Tech., Frequently Asked Questions, supra note 4.

changing synthetic media technologies.

3. METHODOLOGY

The research adopts a doctrinal and qualitative legal research methodology. Primary legal sources, including statutes, rules, regulations, government notifications and judicial decisions, form the principal basis of the study.

Secondary sources such as scholarly articles, books, research papers, policy documents, government reports and credible institutional publications are examined to understand technological developments and existing academic debates concerning deepfakes.

The study follows an analytical approach by examining the relationship between technological developments and existing legal principles. It also adopts a comparative approach, where appropriate, to examine regulatory responses in selected foreign jurisdictions and identify practices that may be relevant to the Indian legal framework.

The research ultimately seeks to determine whether India's existing legal framework provides adequate protection against deepfake-related harms and, where gaps exist, to propose measures that balance individual rights, public interest, technological innovation and constitutional freedoms.

TECHNICAL LANDSCAPE & TAXONOMY OF DEEPFAKES

Deepfake technology is a rapidly evolving area of artificial intelligence ("AI") and synthetic media. It has the ability to create, modify or alter digital content in ways that can make fake material look real. While the term "deepfake" is often linked with manipulated videos of people, current synthetic media technologies now cover audio, images, video and even text- based content. The development of generative models has made the production of realistic synthetic content more efficient and accessible, which has led to both useful uses and notable risks. It is important to understand the technical basis, classifications, availability, and limitations of deepfakes in order to explore their legal issues. The technical nature of a deepfake directly affects questions related to authenticity, consent, privacy, digital evidence, attribution, the responsibilities of intermediaries, and the detection of illegal content.

1. GENERATIVE ADVERSARIAL NETWORKS (GANS) AND

DIFFUSION MODELS

One key technology in the development of deepfakes is the Generative Adversarial Network (GAN). GANs were introduced as a machine learning framework that mainly involves two competing neural networks: a generator and a discriminator. The generator tries to produce synthetic content that looks like real data, while the discriminator tries to determine whether the content is genuine or artificially created. Through repeated training, the generator becomes better at producing more realistic outputs.[18]

GAN-based methods have played a major role in the development of facial manipulation, face-swapping, facial reenactment, and other forms of synthetic visual media. By training models on large datasets, the technology can learn facial features and replicate or change them in another image or video. This has made it possible to create increasingly convincing representations of individuals.

A recent advancement is the rise of diffusion models. These models create content by starting with noisy data and gradually removing the noise until a clear output is formed. Diffusion- based systems are especially important in modern generative AI due to their ability to produce highly detailed and realistic images and increasingly, video content.[19]

The shift from GAN-based systems to newer generative architectures shows the fast development of synthetic media. These technologies have enhanced the quality of generated content while making the creation of complex digital manipulations easier for people with less technical expertise.

2. CLASSIFICATION OF DEEPFAKE MEDIA (AUDIO, VIDEO, IMAGERY, TEXT)

Deepfake media can be categorized that is based on the type of digital content that is created or altered. Even though video is the most commonly recognized form, advances in generative AI have expanded synthetic manipulation into several categories.

[18] Lan J. Goodfellow et al., Generative Adversarial Nets, in Advances in Neural Information Processing Systems 27 (2014).
[19] Jonathan Ho, Ajay Jain & Pieter Abbeel, Denoising Dff usion Probabilistic Models, 33 Advances in Neural Information Processing Systems 6840, 6840–48 (2020).

Video Deepfakes

Video deepfakes involve changing or replacing a person's face, facial expressions, body movements or actions in video clips.

This technology can make it appear as if a person said something or did something that never actually happened. Video manipulation is especially concerning because the combination of facial appearance, movement and speech can create a strong impression of authenticity.

Audio Deepfakes

Audio deepfakes refer to the artificial generation or modification of speech to mimic a specific person’s voice. Voice-cloning systems can replicate features such as tone, pitch, pronunciation, and speaking style. Although this technology has useful applications in entertainment and accessibility, it can also enable impersonation, fraud, fabricated communications, and damage to someone's reputation.

Image Deepfakes

Image-based deepfakes involve the creation or manipulation of still images. AI can alter facial features, body shapes, clothing, surroundings, or other visual elements. One particularly harmful use is the creation of non-consensual intimate imagery, where an individual's likeness is digitally placed into explicit material without permission.[20]

Text-Based Synthetic Content

Generative AI can also produce highly realistic written content, including fabricated statements, social media posts, articles, reviews, messages and other forms of communication. Although text-generated material is not usually referred to as a traditional "deepfake," it is part of the broader synthetic media ecosystem.

The expansion from video manipulation to multiple forms of synthetic content has important legal implications. India’s regulatory framework now uses the broader term

[20] Bobby Chesney & Danielle Keats Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1759–70 (2019).

"synthetically generated information" ("SGI"), which includes artificially created, altered, or modified audio, visual, and audio-visual information.[21]

3. THE DEMOCRATIZED ACCESS: COMMERCIAL APPS VS. MALICIOUS SOFTWARE

The growth of deepfake technology has led to its democratisation, meaning that tools once limited to researchers and tech-savvy individuals are now more widely available to regular users. Commercial platforms and online AI services offer features like face swapping, voice cloning, image generation, avatar creation and video editing with minimal technical knowledge.

This widespread access brings several positive uses. Synthetic media can be applied in filmmaking, advertising, education, entertainment, accessibility, translation, gaming and digital communication. These uses show that deepfake technology itself is not illegal by nature; its legality often depends on how it is used, the intent behind it, whether consent is given and the consequences of its use.

However, the same ease of access can also be used for harmful purposes. People may use freely available or specialized software to create fake identities, impersonate public figures, alter political or business messages, commit financial fraud, spread false information, or produce non-consensual intimate content.[22]

The difference between legal uses and harmful software is becoming more important for legal control. A complete ban on synthetic media could harm legitimate innovation and free expression. Instead, laws should target harmful uses, ensure informed consent, promote transparency, hold people accountable, and establish good ways to report and remove illegal synthetic content.

The widespread availability of deepfake technology also causes issues with attribution. A harmful deepfake may go through many stages, involving an AI developer,

[21] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r. 2(1) (wa), as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026).

[22] Ministry of Elecs. & Info. Tech., Explanatory Note: Proposed Amendments to the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 in Relation to Synthetically Generated Information (Oct. 22, 2025).

software provider, creator, uploader, a platform, and people who share it again. Deciding who should be held legally responsible among these parties is a major challenge for regulation.

4. TECHNICAL LIMITATIONS: STRUCTURAL VULNERABILITIES IN FAKE MEDIA

Even with fast technological progress, deepfake media may still have technical flaws that can help identify fake content. These flaws may include unnatural facial movements, irregular blinking, distorted features, inconsistent lighting, odd shadows, unnatural body movements, or mismatches between lip movements and speech. [23]

Audio deepfakes may also show strange pronunciation, unnatural pauses, inconsistent breathing, or changes in tone and rhythm. However, these signs are not always reliable, as newer AI systems are becoming better at producing realistic content that can bypass many traditional ways of detecting fakes.

Another key concern is metadata and digital provenance. Digital files can include information about when they were made, changed or where they came from. Even though this information can be altered or removed, provenance tools can help confirm the history and authenticity of digital content. Detecting deepfakes now often involves a mix of automated systems, digital forensics, metadata analysis, provenance tools and human input. Studies have shown that both people and machines can struggle to consistently tell the difference between highly realistic synthetic media and real content.[24]

These technical challenges have major legal effects. If courts, investigators, journalists, or regular users can't determine whether audiovisual content is real, it raises questions about the authenticity and legal value of digital evidence. As a result, technological detection and authentication tools are becoming vital in the legal regulation of synthetic media.

The technical nature of deepfakes shows that laws alone can't solve the issue. Effective regulation needs collaboration between law, artificial intelligence, digital forensics, tech companies, regulators and other stakeholders. Understanding how deepfakes are made,

[23] Siwei Lyu, Deep Fake Detection: Current Challenges and Next Steps, IEEE Int'l Conf. on Multimedia & Expo Workshops (2020).
[24] Matthew Groh et al., Deepfake Detection by Human Crowd and Machines, 119 Proc. Nat'l Acad. Sci. 1 (2022).

shared, and identified is essential for creating a legal system that can handle both current and future forms of synthetic media.

CATEGORIZATION OF EMERGING SOCIO-LEGAL THREATS

Deepfake technology has moved from being an experimental use of artificial intelligence to becoming a powerful tool that can affect individuals, organizations and society.

While synthetic media can offer genuine benefits, its misuse presents a wide range of socio- legal threats that go beyond regular cybercrime. Deepfakes can be used to influence political discussions, interfere with elections, threaten national security, create non-consensual intimate images, support financial fraud and identity theft, and reduce trust in digital evidence. The unique danger of deepfakes is that they can make false information look real, increasing the chances that people and organizations will act on fabricated content. These threats also raise complex legal issues around privacy, freedom of expression, criminal responsibility, liability of online platforms, standards of evidence and protection for vulnerable people. A clear system of categorizing these new threats is necessary to fully understand the broader legal implications of deepfake technology.

1. POLITICAL DISINFORMATION, ELECTION MANIPULATION, AND NATIONAL SECURITY

Deepfakes can be a major threat to democracy because they can produce realistic but false images of political candidates, government officials, journalists or other influential people. A synthetic video may falsely show a candidate saying something harmful, accepting a bribe, supporting a certain policy or engaging in inappropriate behavior. Similarly, fake audio can be spread close to an election, leaving little time for verification or correction.

The effect of political deepfakes is not just about the false content itself. Their effectiveness can come from when they are shared, how they make people feel, and how quickly they spread through social media platforms. False content posted right before an election can change public opinions even if it is later proven to be false. This becomes a bigger issue when users share content without checking where it comes from.

Deepfakes can also add to political division and manipulate public views. Both foreign and domestic actors might use synthetic media to create confusion, damage institutions, or

influence debates. As a result, deepfake technology has serious implications for democratic fairness and national security. Scholars have noted that synthetic media poses a risk to privacy, democracy, and national security because it can damage trust in information and institutions. [25]

The challenge from a legal perspective is to stop malicious political manipulation without limiting acceptable political speech. Satire, parody, criticism, artistic expression and political commentary can also use altered or synthetic media. Therefore, regulations must be careful to differentiate between harmful deceptive content and protected forms of expression.

In India, the regulation of synthetic content has become increasingly important in this issue.The 2026 changes to the Information Technology Rules aim to improve transparency about synthetic content and place more responsibilities on online platforms. [26] These steps are meant to reduce the chance that users will mistake AI-generated or altered content for real material.

2. NON-CONSENSUAL INTIMATE IMAGERY (NCII) AND GENDER-BASED DIGITAL VIOLENCE

One of the most serious uses of deepfake technology is the creation and spread of Non- Consensual Intimate Imagery (NCII). AI tools can alter photos or videos to place someone's face on nudity material without their knowledge or permission. Unlike traditional intimate- image abuse, deepfake technology can produce sexual content even when the victim has no existing intimate photos or videos.

NCII can have serious consequences for victims, including embarrassment, emotional suffering, social stigma, career problems, harassment, and threats to personal safety. Women and girls are especially at risk from this type of abuse. Deepfake pornography can be seen as a form of gender-based digital violence, especially when it is used to harass, intimidate, silence, or control victims.

A major legal challenge is that the fake nature of the image does not always lessen

[25] Bobby Chesney & Danielle Keats Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1759–70 (2019).
[26] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026).

the harm felt by the victim. The victim may suffer damage to their reputation and social standing even if the original sexual activity never happened. As a result, traditional legal approaches that focus on the authenticity of the content may not fully address the harm caused by synthetic sexual imagery. [27]

Deepfake NCII also raises key questions about privacy, dignity, consent, and the right to one's identity. The unauthorized use of someone's face or identity can interfere with their control over their digital presence. Indian legal principles regarding privacy and dignity, especially those established under Article 21 of the Constitution, offer a crucial framework for dealing with the harms caused by deepfake technology.[28]

Effective regulation goes beyond simply criminalizing such acts. Victims need quick and accessible ways to remove illegal content, preserve proof, identify those responsible, and access both civil and criminal legal remedies. Online platforms also play an important role in stopping the repeated spread of such harmful material.

3. FINANCIAL FRAUD, IDENTITY THEFT, AND CORPORATE EXTORTION

Deepfake technology has opened new avenues for financial fraud and identity-based deception. AI-generated voice cloning allows criminals to mimic executives, family members, customers, or other trusted individuals. A fake voice message or video call can trick a victim into transferring money, revealing confidential details, or approving a transaction.

Identity theft is another major concern. By replicating someone's face or voice, synthetic media can be used to create fake identities. This issue becomes even more dangerous when deepfakes are combined with stolen personal information, phishing, social engineering, or other cybercrime tactics.

Businesses and corporations are also vulnerable. A fake video or audio clip of a senior executive can be used to manipulate employees, investors, customers, or the market. Deepfakes can also be used for corporate extortion, where criminals threaten to release false

[27] Radha Ranjan & Jivantika Gulati, "The Sexual Harassment against Women in Cyber Space: Concept, Reality and the Law," Kashmir Journal of Legal Studies, Vol. X, No. 1 (2023), pp. 79–90.
[28] K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1

or altered material unless money or other demands are met.

The economic impact can go beyond the immediate victims. Financial institutions, companies, and consumers may suffer from financial losses, damage to their reputation, costs of investigations, and the need for stronger security measures. The growing availability of synthetic media tools makes this issue particularly challenging to address, as advanced impersonation can be done without high technical skills.

Indian criminal and cyber laws may apply to various aspects of these activities, depending on the situation. Offenses such as cheating, impersonation, forgery, fraudulent electronic records, and misuse of identity may fall under the Bharatiya Nyaya Sanhita, 2023 and the Information Technology Act, 2000.[29] However, applying traditional legal definitions to AI-generated actions may raise issues regarding how to classify the offense, attribute it to the perpetrator, prove intent and determine jurisdiction.

4. EVIDENTIARY INTEGRITY: THE "LIAR'S DIVIDEND" IN COURTS OF LAW

The increasing availability of deepfakes poses a unique threat to the integrity of digital evidence. Historically, photographs, audio and video recordings have been powerful forms of proof. However, the ability to create convincing synthetic material brings uncertainty about whether digital evidence is authentic.

This problem is referred to by scholars as the "liar's dividend." It describes a situation where the existence of deepfake technology enables someone to question the authenticity of genuine evidence. [30] In other words, deepfakes not only create false evidence but also allow people to doubt real evidence.

The liar's dividend can have serious consequences for courts and the justice system. A person accused of some wrongdoing may claim that real audio, video, or photo evidence was faked. Alternatively, a bad actor might create fake evidence to wrongly accuse someone else. These situations complicate the ability of courts to determine the truth.

[29] Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 318, 319, 335–340, India Code (2023); Information Technology Act, No. 21 of 2000, §§ 66C–66D, India Code (2000).
[30] Chesney & Citron, supra note 1, at 1782–84.

This challenge is especially important in today's world where digital evidence is widely used in criminal investigations and legal disputes. Courts may need to place more emphasis on authentication, chain of custody, metadata, forensic analysis, expert testimony and supporting evidence when reviewing potentially altered digital material.

Indian evidence laws already recognize the importance of electronic and digital records. The Bharatiya Sakshya Adhiniyam, 2023 provides the current legal structure for handling electronic and digital records. [31] However, with the growing sophistication of deepfakes, courts and investigators may need to develop more specialized methods for verifying and authenticating digital evidence.

In the end, the deepfake issue creates a paradox: the more advanced synthetic media becomes, the harder it may be to prove that false evidence is false and to confirm that real evidence is real. Therefore, maintaining the integrity of evidence will require ongoing interaction between legal standards, forensic science, technological verification and judicial knowledge.[32]

GLOBAL LEGAL & REGULATORY FRAMEWORKS: A COMPARATIVE ANALYSIS

The regulation of deepfake technology has evolved differently in various countries, influenced by differences in fundamental rights, laws, approaches to technology, and rules about responsibility of online platforms.

In India, efforts to manage deepfakes mainly involve intermediary obligations and criminal law. In contrast, the European Union has taken a more systematic and general approach through the EU Artificial Intelligence Act, especially its rules about transparency for content that is created or altered by AI. The United States uses a mix of federal laws, state laws, rights related to public figures, and specific sectoral actions. Recently, new federal laws have increased the response to deepfakes that involve non-consensual intimate images. Comparing these frameworks is important because deepfake content can spread across international borders quickly, making national laws alone inadequate. This comparison also

[31] Bharatiya Sakshya Adhiniyam, No. 47 of 2023, §§ 61–63, India Code (2023).
[32] See Danielle Keats Citron & Robert Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1782–84 (2019).

shows a key challenge in regulation: how to hold online platforms accountable without limiting freedom of expression or slowing progress in technology.

1. THE INDIAN LANDSCAPE: IT RULES AMENDMENTS (TAKEDOWN TIMELINES) AND THE BHARATIYA NYAYA SANHITA (BNS)

India currently lacks a single law that comprehensively addresses deepfakes. Instead, it uses a combination of the Information Technology Act, 2000, the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, the Bharatiya Nyaya Sanhita, 2023 (BNS), and other related privacy and data protection laws.

A major change came with the Information Technology (Intermediary Guidelines and Digital

Media Ethics Code) Amendment Rules, 2026, which introduced new rules dealing with Synthetically Generated Information (SGI). These rules broadly define SGI as audio, visual, or audio-visual content that is artificially or algorithmically made, changed or altered so that it looks real or genuine and it shows a person or an actual event in a way that might be hard to tell from real content. [33]

The updated rules also place specific responsibilities on intermediaries that help create or spread SGI. These intermediaries must use suitable technical tools and methods to prevent the spread of harmful synthetic content. The rules particularly cover content related to child sexual abuse, non-consensual intimate images, pornography, privacy breaches and fake documents or electronic records.[34][35][36] The changes also improve the time limits for removing harmful content and handling complaints. For certain types of harmful content like content involving private parts, nudity, sexual acts or fake images, intermediaries must remove or block access within two hours of getting a complaint. [36] Other complaints may take up to thirty-six hours to resolve.[37]

[33] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r.

[34] (1)(wa), as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E)

[35] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rule 3(1)(b)–(c) — intermediary due-diligence obligations relating to unlawful/SGI content.

[36] IT Rules, 2021, Rule 3(2)(b) — the 2-hour takedown timeline for speci□ed harmful content, including certain intimate/impersonation content.

[37] IT Rules, 2021, Rule 3(2)(a) — the 36-hour grievance-resolution/removal timeline for speci□ed complaints.

The BNS further supports this regulation by offering criminal penalties that could apply to misuse of deepfakes. Depending on the situation, laws related to fraud, criminal breach of trust, forgery, false electronic records, and defamation might be relevant.[38] However, the use of these laws depends on the specific circumstances and whether the conditions for each crime are met. India's approach blends rulemaking for online platforms with traditional criminal law. The main strength of this approach is the ability to enforce quick compliance from intermediaries. However, issues remain concerning enforcement, technological precision, identifying who is responsible, handling foreign platforms and the risk of taking down legitimate content. [39]

2. THE EUROPEAN UNION APPROACH: HORIZONTAL REGULATION UNDER THE EU AI ACT (ARTICLE 50 TRANSPARENCY OBLIGATIONS)

The European Union has taken a structured approach through the Artificial Intelligence Act ("EU AI Act"), which sets up a broad regulatory framework for artificial intelligence based on varying levels of risk.

Article 50 of the EU AI Act is especially relevant to synthetic media. It requires transparency from providers and deployers of certain AI systems. Providers of AI systems that create synthetic audio, image, video, or text content must ensure that their outputs are clearly marked in a machine-readable format and are identifiable as artificially generated or altered, with some exceptions and considering technical feasibility.[40]

The Act also specifically deals with deepfakes. Article 50(4) mandates that deployers of AI systems generating or altering image, audio, or video content that constitutes a deepfake must disclose that the content has been artificially created or modified. There are certain exceptions, including specific situations related to law enforcement activities. Artistic, creative, satirical or fictional works have less strict disclosure rules to prevent transparency requirements from interfering unnecessarily with their presentation. [41]

[38] Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 318, 319, 335–340, 356, India Code (2023).
[39] Radha Ranjan Pallavi Singh "Cyber Crime Against Women in Cyber Space: A Critical Analysis of Indian Legislations," *Kanpur Philosophers*, Vol. X, Issue I(A) (2023).
[40] Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024, Laying Down Harmonised Rules on Arti□cial Intelligence, art. 50(2), 2024 O.J. (L 1689) 1, 76.
[41] EU AI Act, Article 50(4) — disclosure/transparency obligations for deployers of AI systems generating or manipulating deepfake image, audio, or video content.

The EU approach, therefore, primarily focuses on transparency and traceability instead of banning synthetic media entirely. It acknowledges that AI-generated content can have valid uses while aiming to ensure that users are not misled about its artificial origin. Article 50 also highlights the role of technological standards in AI regulation. Providers are expected to use solutions that are effective, interoperable, robust, and reliable, taking into account technical feasibility and current technological advancements.[42] This creates an important regulatory model for jurisdictions such as India, as it links legal requirements with technological authentication and detection methods. The EU framework therefore offers a clearer distinction between the creation of synthetic content and the misleading or hidden use of such content. Its focus on transparency, machine-readable markings, and disclosure provides an alternative to regulation based purely on content removal.

3. THE UNITED STATES FRAMEWORK: FRAGMENTED STATE STATUTES, RIGHT OF PUBLICITY, AND FEDERAL INITIATIVES

The United States has historically taken a more fragmented approach to regulating deepfakes. Instead of relying on a single, comprehensive federal law, regulation has developed through a mix of state laws, existing criminal and civil laws, privacy protections, right-of-publicity doctrines and federal efforts.

Several U.S. states have introduced laws that address specific types of synthetic media, particularly election-related deepfakes and non-consensual intimate imagery. These laws vary significantly in how they define deepfakes, what behavior, they prohibit, the remedies available, and how they are enforced. This results in a complex legal landscape where the legal consequences may depend on the state where the activity occurs or where the victim is located.

The right of publicity is another important legal tool. It generally safeguards an individual's ability to profit from their name, image, likeness, voice, or other recognizable aspects of their identity. Deepfakes that use someone's identity without permission may lead to right-of- publicity issues, although the extent of protection varies by state. This legal principle becomes especially relevant when synthetic representations are used for commercial

[42] EU AI Act, Article 50(2) — obligations concerning machine-readable marking/detectability of AI- generated or manipulated content.

purposes or to exploit someone's identity.

A major new federal development is the TAKE IT DOWN Act, which was signed into law on May 19, 2025.[43] The law addresses the intentional sharing of non-consensual intimate visual content, including AI-generated or altered material. It requires covered platforms to provide a process for victims to request the removal of such content and known copies.

The platform-removal provisions became enforceable on May 19, 2026, requiring covered platforms to remove qualifying non-consensual intimate images within 48 hours of a valid request.[44] The Federal Trade Commission is responsible for enforcing these requirements.

The U.S. framework thus reflects a shift from fragmented state-level regulation toward more targeted federal actions. However, it remains less comprehensive than the EU's horizontal AI framework because different types of deepfake-related harm may still be governed by different federal and state laws.

4. CRITICAL GAPS: SAFE HARBOUR EXEMPTIONS VS. PLATFORM LIABILITY

One of the most challenging issues in deepfake regulation is determining the appropriate level of responsibility platforms should bear. Online platforms are typically not the original creators of harmful deepfake content, but they may enable its rapid distribution to millions of users in a short time. This raises a policy question about whether platforms should only respond to illegal content after being notified or whether they should take proactive steps to detect and stop harmful synthetic media.

Traditional safe harbour protections generally shield platforms from liability for user- generated content, provided certain conditions are met. These protections are meant to promote free expression and prevent platforms from being held responsible for all user-uploaded content.

However, deepfakes challenge this model because they can be created and shared

[43] TAKE IT DOWN Act, Pub. L. No. 119-12, 139 Stat. 227 (2025).
[44] Federal Trade Commission, FTC Begins Enforcing the TAKE IT DOWN Act (May 19, 2026).

quickly through automated systems, making highly realistic harmful material widespread in a short time. A purely reactive notice-and-takedown approach may not be enough if the harm occurs before the victim has a chance to report it.

India's 2026 IT Rules mark an important development as they impose specific technical and due diligence requirements on intermediaries dealing with synthetic generated images (SGI).

At the same time, the updated framework clearly states that removing or blocking access to SGI in compliance with the Rules does not automatically violate the conditions for intermediary protection under Section 79 of the Information Technology Act.[45] This aims to balance platform responsibility with continued safe harbour protections.

The EU also takes a similar approach by increasing transparency and traceability requirements while maintaining a broader regulatory structure based on an AI system's role and risk. The U.S. has increasingly introduced targeted platform obligations, especially in relation to non-consensual intimate imagery through the take it down Act.

The main regulatory gap is in determining the right balance between full platform immunity and unlimited platform liability. Excessive liability could lead to the removal of legal speech, satire, political criticism, and artistic expression. On the other hand, excessive protection may allow platforms to avoid meaningful responsibility, even though they have the technological ability to detect and limit harmful synthetic content.

A sustainable regulatory framework should therefore adopt a risk-based and proportionate approach. Platforms are expected to put in place reasonable technological measures to protect against misuse, ensure clear and open reporting processes, quickly remove content that is clearly illegal, and provide effective ways for users to resolve disputes. At the same time, they should maintain necessary legal protections to support genuine compliance with laws and the right to express opinions freely. Experience from similar situations shows that the development of regulations around deepfakes will likely depend on finding a balance

between encouraging innovation, protecting individual rights, holding platforms

[45] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r. 7(1B), as amended by G.S.R. 120(E) (Feb. 10, 2026).

responsible, and preserving freedom of expression.

PROPOSED LEGAL STRATEGIES & RECOMMENDATIONS

The nature of deepfake technology, which spans multiple countries and evolves quickly, shows that traditional legal methods may not be enough to deal with the new risks it brings. Deepfake content can be made in one country, uploaded via a platform in another, and viewed by people in many different places, creating major challenges with respect to determining where laws apply, gathering evidence, and enforcing legal decisions. This is made worse by the fact that new tech can quickly render current rules outdated. A lasting solution will need both international cooperation and flexible domestic regulations. The following suggestions aim at creating coordinated ways to handle cross-border enforcement and building legal systems that can respond quickly to new kinds of synthetic media harm while protecting basic rights and supporting real technological progress.

1. UNIFORM INTERNATIONAL TREATIES FOR CROSS-BORDER CYBER JURISDICTIONS

Deepfake-related crimes are often transnational in nature. The person making the deepfake might be in one country, the platform might be based in another, and the victim might be in a third. This division makes it hard to find offenders, get digital evidence, serve legal papers, and carry out court orders. Current differences in national laws further complicate prosecution as actions that are illegal in one country may not be treated the same way elsewhere.

An effective long-term solution would involve more development of international legal cooperation regarding synthetic media and AI-driven cybercrime. A multilateral agreement could set out common principles for defining offences, determining jurisdiction, keeping digital evidence safe, helping each other with legal requests, identifying wrongdoers, and removing clearly illegal content across borders.

The Budapest Convention on Cybercrime offers an existing example of international cooperation in handling cybercrime. It has processes for working together, extradition, legal help, and securing electronic evidence. [46] While it doesn't specifically cover

[46] Convention on Cybercrime, Nov. 23, 2001, 2296 U.N.T.S. 167.

deepfakes, its structure shows how countries can work together when digital crimes cross borders.

A future international agreement could build on this by specifically focusing on AI-generated and altered content. It could set minimum standards for crimes involving harmful deepfakes, especially when used for fraud, identity theft, non-consensual intimate images, election interference, or serious damage to reputation. However, such an agreement should not seek to criminalize all forms of synthetic media. Uses that are acceptable, such as satire, artistic expression, education, research, entertainment, and political discussions, must remain protected. The treaty should focus on harmful actions and intentions rather than just the creation of synthetic media.

International cooperation should also include common standards for digital evidence and content verification. Since deepfake investigations may involve servers, accounts, metadata, and communications located in different countries, investigators need reliable ways to preserve and share electronic evidence. Standardized rules could reduce delays and improve the effectiveness of cross-border investigations.

International cooperation is especially important because individual country laws alone cannot stop deepfake content from spreading through global digital platforms. A coordinated approach will therefore make it easier to hold people accountable and close jurisdictional gaps.

2. DYNAMIC REGULATORY TEXT FOR RAPID-RESPONSE TAKEDOWN MECHANISMS

Another key recommendation is the adoption of a dynamic regulatory system that can adapt to technological changes without needing lengthy legal changes each time a new form of synthetic media abuse appears. Traditional laws often take a long time to pass or change, while AI can develop very quickly. This risk means that laws may become outdated before they can effectively address new types of manipulation.

Dynamic regulation should therefore create technology-agnostic legal principles that let regulators update technical standards, prohibited categories, reporting procedures, and platform responsibilities as needed. The goal should not be to rewrite main laws each time a

new AI method is developed, but to give regulators clear powers with checks from courts and the constitution.

Quick-response takedown mechanisms should be part of this system. When synthetic content clearly includes illegal material—such as non-consensual intimate images, child sexual abuse content, serious impersonation, or clearly fraudulent content—platforms should have simple ways for users to report issues and take swift action.

India's amended IT Rules show the benefits of setting clear timeframes for how intermediaries should respond. The 2026 framework introduces faster obligations for specific types of harmful content, such as a two-hour timeline for certain complaints involving intimate or impersonation-related material.[47] These measures can help shorten the time harmful deepfakes stay online and prevent them from spreading further.

However, fast removal processes must include procedural safeguards. Removing content automatically without proper review could lead to legitimate material being incorrectly taken down, especially when the content is part of satire, journalism, political speech, or artistic expression. Therefore, platforms should offer users notice, ways to appeal, fair review processes, and opportunities to challenge wrongful removals.

A flexible regulatory system should also support the use of technological detection and source verification tools. Platforms might be required to identify or mark AI-generated content using machine-readable tags, metadata, or other reliable verification methods where possible. The EU AI Act offers a helpful example through its requirements for transparency regarding artificially created or altered content.[48]

In addition, regulatory bodies should regularly examine how effective takedown systems are. These reviews should take into account elements like response times, incorrect removals, repeated uploads, advances in technology, transparency reports, and the success of support mechanisms for victims. The main goal should be to develop a responsive yet rights-protective regulatory environment. Regulation needs to be adaptable enough to handle new

[47] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, r. 3(2)(b), as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, G.S.R. 120(E) (Feb. 10, 2026).

[48] Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024, Laying Down Harmonised Rules on Arti□cial Intelligence, art. 50, 2024 O.J. (L 1689) 1.

technological risks while also being predictable, clear, fair, and subject to legal supervision. Combining international collaboration, adaptable regulatory tools, quick removal processes, technological protections, and effective support options would offer a more solid approach to dealing with deepfake-related issues in the digital age.

Here is a conclusion suitable for your research paper, summarizing the technical, socio-legal, comparative, and recommended legal aspects of deepfake regulation.

CONCLUSION

Deepfake technology represents both a significant advancement in artificial intelligence and a growing challenge for law, society, and democratic structures. Its ability to generate highly realistic synthetic audio, images, videos, and other digital content has fundamentally changed the way digital deception is carried out. The issue is not the technology itself, but its misuse for political manipulation, non-consensual intimate imagery, financial fraud, identity theft, reputational damage, and the fabrication or denial of digital evidence. As deepfake tools become more accessible and advanced, the ability to distinguish between real and fabricated content is becoming increasingly difficult.

The study shows that current legal systems are often unable to keep up with the rapid development of this technology. In India, the Information Technology Rules and the Bharatiya Nyaya Sanhita offer mechanisms that can address various forms of deepfake-related misconduct, while recent regulatory changes have introduced stricter requirements for synthetic content and quicker removal of harmful material. Nevertheless, challenges such as enforcement, attribution, cross-border jurisdiction, detection, and the protection of fundamental rights remain unresolved. A comparative analysis reveals that there is no single, universally accepted approach to deepfake regulation.

The European Union's AI Act prioritizes transparency and disclosure, while the United States uses a mix of federal and state laws, civil remedies, and sector-specific regulations. India’s approach is increasingly centered on intermediary responsibility combined with criminal and technology-related laws. These differences highlight the need for regulatory systems that are flexible enough to fit different legal environments while still ensuring consistent standards for dealing with serious harms. An important concern is the potential conflict between regulation and freedom of expression. Deepfake laws should not

be used to suppress legitimate political commentary, satire, artistic expression, journalism, research, or other lawful uses of synthetic media. Regulation should target actual harm, unlawful intent, lack of consent, deception, and the effects of distribution, rather than simply banning AI-generated content.

The study also emphasizes the importance of tackling the "liar's dividend", where the presence of deepfake technology allows individuals to discredit real evidence as fake. This risk calls for stronger digital forensic tools, dependable authentication methods, provenance tracking, and judicial awareness of synthetic media. Legal institutions must develop the technical capacity needed to accurately identify real digital evidence from manipulated content. Ultimately, deepfake regulation cannot rely solely on traditional criminal law or content-removal strategies. A multi-layered approach is needed, combining international collaboration, technology-neutral legislation, appropriate platform responsibilities, fast-response mechanisms, digital literacy, detection tools, victim-focused remedies, and strong judicial oversight. International cooperation is especially critical because deepfake creation and sharing often cross-national borders, making domestic solutions insufficient.

The future legal framework should therefore aim to balance technological progress with individual protection. Artificial intelligence should continue to support creativity, communication, research, and economic growth, while people and organizations must be shielded from harmful uses of synthetic media. A flexible, rights-conscious regulatory model, backed by global cooperation and technological safeguards, can offer a more effective response to the ongoing threat of deepfakes. Deepfake technology ultimately challenges one of the core assumptions of the digital age—that what is seen or heard can be believed at face value. The law must adapt to this new reality by creating mechanisms that protect truth, privacy, dignity, democratic integrity, and the reliability of evidence, without unduly limiting lawful technological innovation. The success of future deepfake regulation will depend not on eliminating synthetic media, but on ensuring that its creation and use align with the principles of accountability, consent, transparency, proportionality, and the rule of law.