The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2026

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

LEGAL CHALLENGES OF DEEPFAKE TECHNOLOGY IN INDIA

Aradhya Singh

INTRODUCTION

In the age of artificial intelligence, seeing is no longer believing. Within just a few years, AI has advanced from a tech-hobbyist's playground into a highly sophisticated factory capable of cloning voices, generating photorealistic images, and altering video footage with terrifying precision. For a long time, we treated these "deepfakes" as harmless entertainment—a viral video of a celebrity dancing, or a deceased actor "brought back to life" for a film cameo. But that innocent phase is officially over. Today, deepfakes have weaponized digital media, leaving a trail of compromised elections, ruined reputations, targeted cybercrimes, and an unprecedented violation of women's privacy.[1]

This threat hits with unique force in India. As home to one of the largest and fastest-growing digital populations on the planet, millions of Indians are logging onto the internet for the first time every month, often via cheap smartphone data. This rapid digital expansion has outpaced public awareness, creating a massive, vulnerable ecosystem where synthetic misinformation spreads like wildfire across messaging apps before anyone thinks to fact-check it.

Ultimately, the deepest crisis we face isn't just the existence of fake content itself. It is the creeping, systemic collapse of trust in *all* digital evidence. When anything can be perfectly faked, authentic proof can easily be dismissed as a hoax, leaving our legal system, our media, and our public discourse stranded in a world where truth itself becomes a matter of opinion.

UNDERSTANDING DEEPFAKE TECHNOLOGY

[1] N.M. Benny, *Navigating the Unseen: Deepfakes in Indian Politics*, Strategic Foresight Research Forum (2024)

At its core, a deepfake functions like an advanced digital mask where artificial intelligence (AI) studies thousands of facial movements and voice patterns to imitate a real person almost perfectly. Driven by machine learning—specifically Generative Adversarial Networks (GANs)—this technology relies on two competing neural networks to generate hyper-realistic, deceptive media.[2] Through **face-swapping**, an algorithm maps and seamlessly grafts a target's face onto another body in a video; through **voice cloning**, it processes brief audio samples to replicate someone's exact vocal signature, synthesizing any written script with realistic inflection.

Synthetic media exists on a spectrum of intent and public harm:

[Harmless Editing / Cinema] ⟶ [Satire / Memes] ⟶ [Misinformation / Forwards] ⟶ [Malicious Deepfakes / Fraud]

In India's massive digital ecosystem, this technology has escalated into an active crisis across four destructive fronts:[3]

- **FAKE POLITICAL SPEECHES:** Manipulated audio and video are weaponized during elections to alter voter perceptions via closed networks like WhatsApp.
- **CELEBRITY MORPHED VIDEOS:** High-profile individuals routinely have their faces grafted onto compromising footage, violating their personality and privacy rights.
- **ONLINE VOICE FRAUD:** Scammers clone family members' voices from public social media clips to stage urgent emergencies and dupe relatives into making rapid UPI financial transfers.
- **NON-CONSENSUAL EXPLICIT CONTENT:** Predominantly targeting women, malicious actors generate hyper-realistic, sexually explicit images of ordinary citizens, causing irreversible psychological and social trauma.

Ultimately, the deepest legal hurdle is the systemic collapse of digital trust. When any piece of evidence can be simulated perfectly, authentic proof can easily be dismissed as a hoax, undermining judicial proceedings and public truth.

[2] Abhishek Kukreti, Dr. Vivek Kumar & Dr. Avishek Raj, *The Impact of Deepfake Technology on Legal Systems: A Global Perspective*, 7 **ShodhKosh: J. Visual & Performing Arts** 530, 534–38 (2026)
[3] S. Kashyap, *The Digital Mirage: India's Evolving Legal Battle Against Deepfake Technology*, 22 **SCRIPTed** 112, 118–24 (2025).

THE RISE OF DEEPFAKES IN INDIA: WHY THE THREAT IS SERIOUS

The distinct danger of deepfakes in India lies in how seamlessly they exploit human psychology, pulling us into a "post-truth digital society where emotions spread faster than facts." In a culture built heavily on digital community, synthetic media acts as an accelerant. It triggers immediate outrage, fear, or religious empathy, bypassing logical filters before truth can catch up. This psychological vulnerability fractures society across four core vectors:

- **POLITICAL MANIPULATION:** Democratic processes face algorithmic warfare. Weaponized deepfakes, like fabricated audio of political candidates making derogatory remarks, are unleashed hours before voting deadlines.[4] They stoke communal tensions on closed messaging platforms where emotional momentum renders late-stage fact-checking entirely obsolete.
- **CYBER FRAUD:** Traditional phishing text messages have shifted to hyper-personalized voice cloning scams. By harvesting brief vocal snippets from an individual's public Instagram or YouTube accounts, automated tools simulate authentic distress calls, tricking family members into executing immediate UPI financial transfers during artificial emergencies.
- **GENDERED EXPLOITATION:** Deepfakes act as targeted weapons against women's safety and autonomy. Through non-consensual explicit content,[5] malicious actors generate hyper-realistic, sexually explicit images of ordinary women. This synthetic revenge pornography functions as a modern tool for extortion and harassment, resulting in irreversible societal isolation.
- **THREAT TO JOURNALISM:** For newsrooms, synthetic media dismantles editorial reliability. When verifying digital evidence demands heavy forensic computation, reporters cannot quickly distinguish authentic breaking-news footage from deepfakes. This systematic manipulation compromises investigative journalism, creating a culture where genuine evidence can simply be dismissed as a hoax.

CONSTITUTIONAL CONCERNS RAISED BY DEEPFAKES

[4] Bente Kalsnes, *Protecting Democracy: Comparing Codes of Conduct for Generative AI Use During Elections*, 29 **Info. Comm. & Soc'y** 1, 4–9 (2026).
[5] S. Shaji, *Legal Framework Governing Deepfake & Personality Rights in India: A Critical Analysis*, 6 **Indian J. Legal Res.** 417, 422–28 (2026).

Technology has outpaced legislation. As hyper-realistic synthetic media blurs the line between fact and fiction, the Indian legal system faces a fundamental crossroad. Deepfakes weaponize realism itself, and the law struggles where algorithms evolve overnight. Existing laws treat the downstream side effects of technology rather than its structural source. This statutory vacuum triggers complex constitutional tensions across fundamental guarantees.

THE SPEECH DILEMMA: ARTICLE 19(1)(A) VS. ARTICLE 19(2)

Under the Constitution of India, AI-generated content may technically qualify as a protected form of expression. Article 19(1)(a) guarantees the freedom of speech and expression, a right that the Supreme Court has consistently extended to modern digital mediums.[6] Satirical deepfakes, parodies, or creative AI cinema fall under this expressive umbrella.

However, when deepfakes are weaponized to engineer election misinformation—as seen in viral incidents where political candidates appear to withdraw from races hours before polling or to incite communal tension via fabricated audio clips, the state must intervene. These malicious applications test the boundaries of "reasonable restrictions" codified under Article 19(2).[7]

CONSTITUTIONAL RIGHT: ARTICLE 19(1)(A)	**STATE RESTRICTIONS: ARTICLE 19(2) BOUNDARY**
• Creative AI Art and Expression	• Public Order Disruption & Fake News
• Political Satire and Parody	• Criminal Defamation & Character Assassination
• Protected Digital Speech	• Incitement to a Cognizable Offence

The state can legally curb synthetic media if it directly threatens public order, morality, or constitutes criminal incitement and defamation. Yet, enforcing blanket algorithmic bans risks over-censorship. The Ministry of Electronics and Information Technology (MeitY) addressed

[6] INDIA CONST. art. 19(1)(a); *Shreya Singhal v. Union of India*, (2015) 5 S.C.C. 1 (India).
[7] INDIA CONST. art. 19(2).

this by introducing the IT Amendment Rules, which enforce a strict three-hour takedown window for prohibited synthetic content. However, this aggressive timeline risks creating a chilling effect, forcing platforms to pre-emptively censor legitimate parodies out of regulatory fear.

IDENTITY AS DATA: ARTICLE 21 AND THE PUTTASWAMY PRECEDENT

The more severe constitutional injury lies in the violation of Article 21, which guarantees the right to life, personal liberty, and individual dignity. In the landmark judgment *Justice K.S. Puttaswamy v. Union of India*, the Supreme Court recognized the right to privacy as an intrinsic component of Article 21,[8] establishing strict tests for state and private intrusions. Deepfakes systematically dismantle this framework because they convert identity itself into manipulable data.

When a malicious actor crafts non-consensual explicit deepfakes of ordinary citizens or high-profile actors cloning their bodies and faces without consent they do not merely steal data; they violate informational privacy and shatter bodily autonomy. The judiciary has stepped in to protect this boundary. For example, in *Anil Kapoor v. Simply Life India*, courts issued strict injunctions against the unauthorized AI cloning of an individual's face and voice traits. Ultimately, deepfakes inflict severe reputational damage, weaponizing a person's biometric likeness against their will and turning individual identity into a tool for public exploitation.

EXISTING LEGAL FRAMEWORK IN INDIA: A FRAGMENTED SHIELD

Indian jurisprudence does not have a single, unified statute explicitly targeting artificial intelligence or synthetic media. Instead, courts and law enforcement must rely on a fragmented patchwork of legacy cyber regulations, overhauled criminal codes, data frameworks, and intellectual property doctrines. While this multi-layered approach offers various avenues for legal recourse, it acts more as a reactive shield than a proactive deterrent against rapid algorithmic evolution.

THE INFORMATION TECHNOLOGY ACT, 2000

[8] *Justice K.S. Puttaswamy v. Union of India*, (2017) 10 S.C.C. 1 (India).

The primary statutory text for digital offenses remains the Information Technology (IT) Act, 2000.[9] Prosecutors routinely invoke a specific sequence of provisions to address deepfake manipulation:

- **SECTION 66C (IDENTITY THEFT):** Penalizes the fraudulent or dishonest use of another person's unique electronic identification features.
- **SECTION 66D (CHEATING BY IMPERSONATION):** Targets individuals who use a computer resource or communication device to cheat by pretending to be someone else.
- **SECTIONS 67 AND 67A (OBSCENE/EXPLICIT CONTENT):** Prohibits publishing or transmitting obscene or sexually explicit material in electronic form.
- **SECTION 69A (BLOCKING ORDERS):** Grants the Central Government authority to issue emergency blocking directives to remove content that threatens national security or public order.

THE STRUCTURAL VULNERABILITY: The foundational architecture of the IT Act was drafted a quarter-century ago—long before Generative Adversarial Networks (GANs) democratized synthetic media. Because these provisions assume a human bad actor manually stealing standard credentials or distributing static files, they struggle to address the automated, friction-free generation of synthetic content. For instance, when an AI tool synthesizes a highly realistic video from scratch, no traditional identity data is technically "stolen" or broken into; rather, a new, simulated reality is constructed.

Furthermore, while the IT Intermediary Rules enforce strict, accelerated takedown windows for flagged deepfakes, this approach remains inherently reactive. The primary psychological and reputational damage of a malicious deepfake is achieved within the first few minutes of going viral, rendering delayed downstream platform removals structurally inadequate.

THE BHARATIYA NYAYA SANHITA (BNS), 2023

With the implementation of the Bharatiya Nyaya Sanhita (BNS), 2023, India updated its penal framework to better address digital offenses,[10] replacing legacy provisions of the Indian Penal Code (IPC):

[9] Information Technology Act, 2000, No. 21, Acts of Parliament, 2000 (India).
[10] Bharatiya Nyaya Sanhita, 2023, No. 45, Acts of Parliament, 2023 (India).

OFFENCE CATEGORY	KEY BNS PROVISION	APPLICATION TO DEEPFAKES
CRIMINAL DEFAMATION	**SECTION 356**	Replaces Section 500 IPC. Applies when synthetic videos or morphed traits are deployed to intentionally target and systematically destroy an individual's public reputation.
FORGERY OF ELECTRONIC RECORDS	**SECTIONS 335 & 336**	Replaces Sections 465 & 469 IPC. Prosecutes the malicious manipulation, alteration, or fabrication of digital media intended to deceive an audience or cheat a target.
CHEATING BY PERSONATION	**SECTION 319**	Replaces Section 416 IPC. Targets individuals who use synthetic voice clones or facial masks to assume false identities for financial or social deception.
CYBER HARASSMENT & STALKING	**SECTION 78**	Replaces Section 354D IPC. Applies when automated tools are utilized to repeatedly monitor, target, or generate unwanted digital imagery of a woman.

The transition to the BNS signals a clear legislative recognition of digital harms. However, a major enforcement gap remains: establishing *mens rea* (criminal intent). In a distributed open-source pipeline, separating the developer who creates the AI tool from the user who inputs the malicious prompt complicates traditional criminal attribution.

COPYRIGHT LAW AND PERSONALITY RIGHTS

Deepfakes do not merely break criminal laws; they actively violate intellectual property by exploiting an individual's unique voice, image, and physical likeness. Because traditional

copyright frameworks primarily protect fixed creative expressions rather than raw human traits, the judiciary has had to step in.

Through landmark orders—such as *Anil Kapoor v. Simply Life India*—Indian High Courts have recognized robust "personality rights" under the umbrella of common law and Article 21.[11] These evolving doctrines allow public figures to secure sweeping ex parte interim injunctions against tech platforms, blocking the unauthorized commercial utilization or synthetic cloning of their distinct vocal attributes, names, and physical likenesses.

DATA PROTECTION CONCERNS: THE DPDP ACT, 2023

The statutory center of gravity for regulating the building blocks of deepfakes is the Digital Personal Data Protection (DPDP) Act, 2023.[12] Under this framework, an individual's unique facial architecture and vocal signatures are classified as personal data.

Consequently, scraping public social media profiles to train generative AI engines or run face-swapping algorithms without explicit, verifiable consent constitutes a severe compliance violation. The DPDP Act places significant legal and financial liability on data fiduciaries, empowering the Data Protection Board to impose penalties of up to ₹250 crore for unauthorized biometric processing.

Nevertheless, a critical statutory blind spot persists: enforcement. While the DPDP Act provides a strong framework for holding legitimate domestic enterprises accountable, it remains largely ineffective against anonymous, decentralized applications operating outside Indian jurisdictions, leaving victims with a clear right but an elusive remedy.

MAJOR LEGAL CHALLENGES IN REGULATING DEEPFAKES

The rapid proliferation of sophisticated Artificial Intelligence (AI) and deepfake technology poses unprecedented challenges to India’s criminal justice system, particularly concerning the admissibility and authentication of digital evidence. While the newly enacted Bharatiya Sakshya Adhiniyam, 2023 (BSA) which repealed the Indian Evidence Act, 1872 broadened the definition of electronic records under Section 2(1)(d) to incorporate digital communications,[13] it remains fundamentally silent on specific forensic standards for validating AI-generated

[11] *Anil Kapoor v. Simply Life India*, 2023 SCC OnLine Del. 6914 (India).
[12] Digital Personal Data Protection Act, 2023, No. 22, Acts of Parliament, 2023 (India).
[13] Bharatiya Sakshya Adhiniyam, 2023, No. 47, Acts of Parliament, 2023 (India).

content. Consequently, malicious actors can seamlessly fabricate highly realistic audio-visual records, such as doctored confessions, false alibis, or manipulated video clips.

This legislative lacuna undermines judicial integrity, as the conventional certificate-based authentication process under Section 63 of the BSA is ill-equipped to detect deepfakes that bypass standard digital signatures. To mitigate these risks, the Ministry of Electronics and Information Technology (MeitY) enacted the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026. These rules introduce the concept of "Synthetically Generated Information" (SGI) and enforce stringent, compressed compliance windows—including a strict two-hour takedown mandate for high-risk, non-consensual deepfakes.[14]

Nevertheless, a profound systemic gap persists between online platform regulation and judicial evidentiary standards. Investigating agencies face severe capacity deficits, and courts lack specialized, credentialed forensic infrastructure to definitively distinguish authentic digital footprints from algorithmically altered ones. Without explicit statutory updates embedding advanced cryptographic and metadata verification into the BSA, the fundamental truth-seeking function of Indian courts remains highly vulnerable to technological manipulation.

ROLE OF SOCIAL MEDIA PLATFORMS AND AI COMPANIES

The explosive rise of generative AI has fundamentally altered the digital landscape, shifting the burden of truth onto corporate gatekeepers. In this new ecosystem, the responsibility of platforms and AI developers is no longer just ethical it is existential.

Historically, social media networks have operated on a reactive model, deploying moderation teams or issuing corrections after a toxic piece of media has already achieved viral status. However, this delay is weaponized by algorithmic amplification. Recommendation engines are mathematically optimized to prioritize user engagement and emotional outrage. Because hyper-realistic synthetic media inherently triggers these psychological responses, platform algorithms actively accelerate the spread of unverified deepfakes, turning structural design flaws into a vector for chaos.

To mitigate this, tech governance demands a dual approach of proactive friction and rapid takedown systems:

[14] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Gazette of India, pt. II sec. 3(i) (2026).

- Upstream Governance: AI companies must implement mandatory cryptographic AI watermarking and structural metadata tracking at the point of synthesis.[15]
- Downstream Governance: Social platforms must mandate clear content labeling for synthetically generated information (SGI) and enforce hyper-compressed takedown windows (such as India's three-hour mandate) during critical events.

The Pacing Crisis: The speed of misinformation is now faster than the speed of legal remedy.

When an algorithm can distribute a damaging deepfake to millions of feeds in seconds, traditional judicial processes subpoenas, hearings, and injunctions become functionally obsolete. By the time a court can issue a legal remedy, the digital fabrications have already permanently altered public perception, sabotaged reputations, or disrupted elections. To safeguard digital discourse, tech companies must transition from passive intermediaries to active custodians, designing safety into their code rather than treating democratic stability as collateral damage.

COMPARATIVE PERSPECTIVE

While the global community moves aggressively toward structured AI governance, legislative strategies diverge significantly by region.

In the United States, decentralized gridlock has shifted the regulatory burden. Congress took a targeted national step by enacting the tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks (TAKE IT DOWN) Act, establishing strict platform-level takedown rules for non-consensual intimate media. However, granular control remains fractured at the state level, where over forty states have independently passed laws criminalizing explicit deepfakes and enforcing mandatory disclaimers on deceptive political campaign advertisements ahead of election cycles.

Conversely, the European Union relies on a highly centralized, risk-based architecture. Under Article 50 of the European Union Artificial Intelligence Act,[16] providers face strict transparency obligations and AI disclosure rules. Generative systems must mark synthetic text, audio, and video in machine-readable formats, forcing deployers to clearly label deepfakes unless they are obviously satirical or artistic.

[15] N.R. Minhas, *Exploring Legal and Technical Challenges of Deep Fakes in India*, 6 **Int'l J. Multidisciplinary Res.** 2160, 2165–70 (2024).
[16] Regulation (EU) 2024/1689, Artificial Intelligence Act, art. 50, 2024 O.J. (L 1689).

Taking state control a step further, China has codified the world's most rigid, uniform provenance system. The Cyberspace Administration of China (CAC) enforces mandatory labeling of AI-generated content through strict national standards.[17] Service providers must deploy a dual-layer strategy: highly visible, explicit watermarks on public-facing content and implicit, tamper-resistant cryptographic identifiers embedded directly within file metadata.

THE INDIAN ENFORCEMENT DISCONNECT

This international mapping exposes a critical structural vulnerability: India currently relies more on general cyber laws than specialized AI regulation. While global peers implement preemptive, AI-specific compliance, India's legal enforcement relies on retrospective penal codes. This patchwork approach attempts to stretched-fit foundational cyber frameworks to combat an exponential tech threat that demands a dedicated, upstream legislative architecture.

THE WAY FORWARD: SOLUTIONS FOR INDIA

To navigate the asymmetric threat of synthetic media, India must shift from retrospective prosecution to proactive structural deterrence. Technology has outpaced legislation; resolving this systemic vulnerability requires a highly coordinated approach combining legislative updates, supply-chain accountability, and foundational capacity building.[18]

THE PILLARS OF AI MITIGATION

└── 1. STATUTORY BASIS

  ├── Specific Laws

  └── Legal Definition

└── 2. TECH CONTROLS

  ├── Provenance Data

  └── Watermarking

└── 3. HUMAN CAPACITY

  ├── Media Literacy

[17] Measures for the Labelling of Artificial Intelligence Generated Synthetic Content (Cyberspace Administration of China et al. 2025).
[18] S. Vig, *Regulating Deepfakes: An Indian Perspective*, 17 **J. Strategic Sec.** 73, 78–84 (2024).

└── Cyber Cells

- **COMPREHENSIVE STATUTORY OVERHAUL:** India must enact dedicated deepfake legislation that moves beyond administrative guidelines to formalize a clear, statutory definition of synthetic media. This framework should penalize the malicious upstream creation, deployment, and open-source hosting of generative models rather than simply chasing downstream side effects.
- **MANDATORY TECHNICAL PROVENANCE:** The law must mandate corporate accountability from AI developers. Regulations should require cryptographic AI watermarking and embedded metadata logs at the point of media generation, allowing for instant, automated verification across social networks.
- **HYPER-COMPRESSED GRIEVANCE PORTALS:** Social media platforms must optimize their internal response units. Establishing rapid grievance redressal mechanisms with fast-tracked verification protocols ensures deceptive content is identified and contained before achieving viral scale.
- **INSTITUTIONAL AND JUDICIAL PREPAREDNESS:** State-level police forces must upgrade to specialized cyber cells equipped with dedicated AI forensic experts. Concurrently, ongoing judicial training programs must be instituted to educate judges and investigators in verifying complex digital records under the Bharatiya Sakshya Adhiniyam, 2023.
- **PUBLIC MEDIA LITERACY PROGRAMS:** The ultimate systemic defense requires a digitally resilient citizenry. Incorporating widespread digital literacy campaigns that educate the public on reverse-image routing, media verification techniques, and general fact-checking principles will help turn individual critical thinking into the country's strongest barrier against algorithmic manipulation.

CONCLUSION

Deepfakes are not merely technological disruptions or isolated cyber offenses; they represent a structural assault on the bedrock of modern society. By distorting reality at scale, hyper-realistic synthetic media poses an existential threat to democratic processes, individual consent, institutional trust, and the very concept of objective authenticity. When biometric likeness and human expression can be automated and weaponized seamlessly, the entire architecture of digital citizenship begins to fracture.

Resolving this crisis demands a systemic shift in how we approach technology governance. India can no longer rely on a reactive patchwork of legacy regulations and administrative updates. Navigating this algorithmic landscape requires a robust legal framework that embeds cryptographic accountability, clear statutory definitions, and forensic resilience directly into our judicial and technological infrastructure.

As we cross into an era where seeing is no longer believing, our response will determine whether the internet remains an open town square or transforms into an unverified echo chamber of synthetic manipulation. Ultimately, the real danger of deepfakes lies not only in fake videos, but in a future where society stops believing even the truth. India's challenge is not simply to regulate artificial intelligence, but to preserve human trust in the digital age.