VOLUME 6 | ISSUE 4

INTERNATIONAL JOURNAL OF ADVANCED LEGAL RESEARCH

DOCTRINAL COHERENCE IN THE AGE OF DEEPFAKES: INTELLECTUAL PROPERTY, CONSENT, AND PUBLIC INTEREST IN INDIA

- Ms. Reshma Hayat[1] & Prof. (Dr) Naseem Ahmed[2]

Abstract

Deepfakes collapse the distance between what looks real and what is true. For India, this shift raises urgent questions about who controls identity, how creative works are used, and what safeguards protect the public sphere. This paper explores deepfake technology through an intellectual property perspective while engaging directly with privacy, consent, defamation, and public order. It explains the mechanics that make deepfakes so lifelike, traces their legitimate uses and abuses, and tests India's current legal toolkit, the IT Act and Intermediary Rules, copyright and performers' rights, and court-driven personality rightsagainst real-world harms. The analysis surfaces four fault lines: the absence of deepfake-specific definitions and graded harms; lack of explicit, revocable consent for likeness and voice; ambiguity around the IP status of AI outputs and the lawful use of copyrighted inputs; and enforcement friction around detection, attribution, rapid takedown, and cross-border cooperation. Drawing on comparative models from the US, EU, UK, and China, the paper proposes a layered reform agenda tailored to India's constitutional commitments: precise statutory definitions; identity autonomy and consent as the baseline; clarified IP treatment for inputs and outputs; risk-tiered platform duties with fast removal, stay-down, and transparent appeals; standardized forensic and evidentiary protocols; provenance and labelling requirements; capacity-building for police, prosecutors, and courts; and public media literacy. The aim is not to ban synthetic media, but to place clear guardrails that deter abuse, especially intimate-image fakes, official impersonation, electoral interference, and fraud, while supporting responsible creativity, journalism, and research.

[1] Research Scholar, Faculty of Law, Integral University, Lucknow.
[2] Professor (Dean & Head), Faculty of Law, Integral University, Lucknow.

Keywords: *Deepfakes, Intellectual Property rights, consent and privacy, personality rights, platform responsibility.*

1. Introduction

Deepfake technology is an advanced and sophisticated development at the intersection of artificial intelligence (AI) and machine learning. It enables the creation of highly realistic yet entirely fabricated digital content, including images, videos, and audio recordings. The term *deepfake* is derived from *deep learning*, a subset of AI involving neural networks, and *fake*, reflecting the deceptive nature of the content produced. The core technology behind deepfakes involves generative adversarial networks (GANs), where two AI networks-the generator, which creates synthetic media, and the discriminator, which evaluates the authenticity-continuously learn from each other to produce increasingly convincing fabrications. This iterative process allows deepfake videos and images to become nearly indistinguishable from genuine footage, raising significant concerns for individuals, organizations, and governments alike.[3]

Globally, the rise of deepfake technology is propelled by increased accessibility to AI tools and the exponential growth of digital communication platforms, enabling rapid distribution of synthetic media to millions. India, given its vast population, rising digital connectivity, and expanding internet penetration, has witnessed a surge in deepfake cases. This growth threatens the integrity of public discourse, individual rights, and national security, making it imperative to analyses and address these risks within a robust legal and ethical framework.[4]

In the Indian context, deepfakes raise intricate legal questions, particularly around intellectual property rights. Deepfakes often involve the unauthorized use of copyrighted works, including images, videos, and audio clips, as well as a person's likeness without permission, creating complex liability and enforcement issues. Currently, Indian laws do not provide specific regulations targeting synthetic media or AI-generated content, leading to significant gaps. The extant intellectual property and cyber laws, developed before the rise of deepfake technology, are tested by unprecedented questions regarding ownership, consent, and rights in the age of synthetic media.

[3]Floridi, L. (Ed.). (2020). The ethics of artificial intelligence: Principles, challenges, and opportunities. Oxford University Press.

[4] Press Information Bureau. (2023, December 26). India is well-equipped to tackle evolving online harms and cyber crimes. Government of India. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2154268

This research focuses on exploring these challenges comprehensively. It aims to evaluate the adequacy of India's intellectual property laws in responding to deepfake technology, examine the existing regulatory gaps, and propose necessary judicial and policy reforms. The study seeks to clarify how deepfakes intersect with Indian IP protections, identify where current legislations fall short, and offer recommendations to ensure effective protection of individuals' rights and societal interests in an increasingly digital and AI-driven environment.

2. Understanding Deepfake Technology: Technical Foundations, Varieties of Synthetic Media, and Dual-Use Applications

Deepfake technology is fundamentally powered by advances in artificial intelligence, specifically deep learning techniques that enable machines to autonomously generate convincing synthetic content. Central to this process are Generative Adversarial Networks (GANs), which consist of two competing neural networks-the generator and the discriminator. The generator creates synthetic media by learning from vast datasets comprising real images, videos, or audio, while the discriminator's role is to distinguish between authentic and artificial content. Through continuous feedback and improvement, these networks produce media that is increasingly realistic, challenging the ability of humans and traditional detection tools to discern authenticity.[5]

There are various types of deepfake content that exploit this technology. The most common include videos where an individual's face or expressions are convincingly superimposed onto another body or scene. Audio deepfakes manipulate voice recordings to make it sound as if a person said words they never spoke. [6]Image deepfakes create or alter photographs, placing people in fabricated contexts or transforming their appearances. Additionally, emerging forms of synthetic text content, generated by AI language models, contribute to misinformation by impersonating individuals or creating false narratives.

The applications of deepfake technology are twofold. On the legitimate side, deepfakes enhance creativity and efficiency in entertainment, allowing filmmakers to create visual effects and resurrect deceased actors digitally, or to localize content through realistic dubbing. Educational programs utilize deepfakes for immersive learning, while the advertising sector experiments with personalized marketing. However, the darker side involves malicious

[5] Kietzmann, J., Lee, L., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135–146. https://doi.org/10.1016/j.bushor.2019.11.006

[6]Tolosana, R., Vera-Rodríguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131148. https://doi.org/10.1016/j.inffus.2020.06.014

misuse, including fabricated political speeches, revenge pornography, fraud, and harassment. This dual nature necessitates a balanced understanding of the technology's capabilities alongside robust safeguards against abuse.

3. Intellectual Property Law Challenges in the Indian Context: Copyright Infringement, Personality Rights, and Liability Complexities

The rise of deepfake content raises serious challenges for intellectual property law in India. One primary concern is copyright infringement. Deepfakes frequently involve unauthorized use of copyrighted works, such as videos, photographs, or audio recordings, manipulated to create synthetic media. Indian copyright law, though robust in protecting original works, lacks specific provisions addressing AI-generated content or synthetic transformations. Consequently, establishing clear ownership and infringement in deepfake cases becomes legally ambiguous, complicating enforcement and remedies.[7]

Equally critical are issues related to personality rights and the right of publicity. These legal concepts protect individuals from unauthorized commercial exploitation of their likeness, image, or voice. Deepfakes often violate these rights by using an individual's identity without consent, potentially damaging reputations or causing emotional distress. In India, courts have begun recognizing such violations, particularly concerning celebrities, but comprehensive statutory protections covering all individuals remain absent, leaving many without adequate legal protection.[8]

4. Privacy, Consent, and Defamation: Legal and Ethical Fault Lines in the Age of Deepfakes

Deepfake technology creates profound tensions at the intersection of privacy, consent, and reputation. At its core, a deepfake that fabricates a person's image, voice, or actions without authorization intrudes upon decisional and informational privacy, undermining an individual's control over how their identity is represented and shared. The harms range from intimate-image abuse and coerced *digital nudity* to fabricated speech or conduct that a person never consented to portray. Even when no physical intrusion occurs, the appropriation and manipulation of personal likeness constitute a grave dignitary harm, often experienced as a violation of autonomy, humiliation, and loss of agency.

---

[7] World Intellectual Property Organization. (2023). WIPO issues paper on intellectual property policy and artificial intelligence (2nd ed.). World Intellectual Property Organization. https://www.wipo.int/edocs/pubdocs/en/wipo-pub-1055.pdf

[8] Delhi High Court. (2023). Anil Kapoor v. Simply Life India &Ors., 2023 SCC Online Del 6531.

Consent is the legal and ethical fulcrum. In practice, meaningful consent requires that the subject know what content will be created, how it will be altered, where it will be published, and for how long it will remain accessible.[9] Deepfakes routinely bypass these requirements: images scraped from social profiles, film clips, and voice samples are repurposed at scale without notification, let alone informed permission. Further complexity arises with "consent laundering," where platforms' broad terms of service are misused to imply permission for synthetic manipulation-despite users never intending to authorize identity appropriation. A rights-respecting regime must treat consent as explicit, specific, revocable, and time-bound, recognizing that permission for one use (e.g., a headshot) is not blanket consent for another (e.g., synthetic pornography or political speech).

Defamation risks surge in a deepfake environment because the technology can convincingly attribute harmful statements or acts to a person who never made them. A faked confession, a fabricated hate speech clip, or a staged video of misconduct can inflict immediate reputational damage, jeopardize employment, trigger social ostracism, and expose the subject to legal peril. Traditional defamation standards, false statement, publication, harm, and fault, must be applied to synthetic media that appears authentic to an ordinary viewer. Remedies should include swift takedowns, retraction and correction mechanisms, court-ordered de-indexing, and damages proportionate to the speed and scale of viral spread. Importantly, the law should recognize that the burden of disproving a realistic fake cannot rest solely on the victim, and procedural tools should facilitate rapid evidentiary relief.[10]

Privacy and consent concerns are magnified in intimate deepfakes. Non-consensual sexually explicit deepfakes weaponize identity for coercion or harassment, often targeting women and public figures. These abuses demand heightened protections: expedited injunctions, anonymity for complainants, mandatory platform response timelines, and aggravated penalties for creators and distributors of intimate synthetic media. Beyond punitive measures, survivor-centred processes, confidential reporting, trauma-informed investigation, and lasting content suppressionare essential to meaningful redress.

A further challenge is the *residual harm* of deepfakes: even after removal, screenshots, mirrors, and reposts perpetuate the injury. Effective responses require persistent takedown protocols, hash-based matching to block re-uploads, and cross-platform coordination. Public

[9] European Data Protection Board. (2020). Guidelines 05/2020 on consent under Regulation 2016/679 (Version 1.1). https://edpb.europa.eu

[10] Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954

education also matters: equipping audiences to recognize manipulation reduces the potency of defamatory or privacy-violating fakes and counters the "liar's dividend," where wrongdoers dismiss authentic evidence as fake to evade accountability.[11]

6. Enforcement and Evidentiary Challenges in Addressing Deepfake Technology

A primary hurdle is reliable detection. High-quality deepfakes leave few visible artifacts; even trained viewers may be misled. Technical detection methods exist, but their performance can vary with the generator model, compression level, platform processing, and the domain (image, video, or audio). Contemporary detection typically blends multiple approaches: pixel-level artifact analysis (e.g., frequency-domain inconsistencies, lighting and shadow mismatches), physiological and behavioural cues (e.g., eye-blink patterns, lip-speech synchronization, head pose dynamics), device and encoding forensics (metadata coherence, recompression traces), and model-based classifiers trained on diverse fake and real corpora. None is foolproof in isolation, and adversaries adapt quickly-switching generation methods, laundering files through re-encoding, or adding adversarial noise to degrade detectors. This creates a moving-target problem where evidentiary reliability must be demonstrated not only by the output (a classification) but also by documented methodology, error characteristics, and validation history.Because lay perception is insufficient, expert testimony becomes central.[12] Courts increasingly rely on digital forensics specialists who can explain analytic workflows in accessible terms: what features were tested, what tools were applied, how confidence was calculated, what controls were used, and what alternative explanations were ruled out. Credibility turns on transparency: repeatable procedures, known error rates, peer-reviewed techniques where possible, and chain-of-custody discipline. Judges must evaluate competing expert claims and reconcile conflicting conclusions when parties present different tools or datasets. This places a premium on court literacy with technical evidence and, where feasible, neutral court-appointed experts or agreed protocols to reduce a "battle of experts."

Attribution poses distinct evidentiary challenges beyond detecting falsity. A case may require showing who created the fake, who first uploaded it, who amplified it knowingly, and whether intermediaries acted diligently after notice. Here, investigators look for provenance markers: original file hashes, platform upload logs, time stamps, IP addresses, account linkages, device fingerprints, and any embedded watermarks or provenance signals (where

[11] Paris, B., & Donovan, J. (2019). Deepfakes and cheap fakes: The manipulation of audio and visual evidence. Data & Society. https://datasociety.net

[12]Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932. https://doi.org/10.1109/JSTSP.2020.3002101

present). However, anonymization tools, VPNs, throwaway accounts, and cross-platform reposting can obscure origins. When creators cannot be identified, liability theories may focus on downstream actorsdistributors who shared content despite red flags, or platforms that failed to act after credible notice. The evidentiary record must therefore capture not only the media artifact but also the timeline of dissemination, notice events, and platform responses.[13]

Preservation of digital evidence is another pressure point. Synthetic content can be altered, deleted, or compressed by platforms' pipelines, degrading forensic signals. Effective preservation begins with immediate capture of the highest-quality version available (preferably the source file rather than a streamed copy),cryptographic hashing to lock the artifact's integrity, and detailed logging of acquisition steps. Parallel to capturing the media, parties should initiate prompt preservation requests to hosting platforms for associated logs (upload times, account IDs, IP data, prior versions, deletion records) before routine retention cycles purge them. Maintaining a clean chain of custody-from initial capture, through analysis, to submissionreduces admissibility disputes and prevents contamination claims.[14]

Jurisdictional complexity magnifies all these issues. Deepfakes can be created in one country, edited in another, hosted on servers in a third, and consumed globally. Victims and investigators must navigate divergent legal standards, data protection rules, and cooperation pathways to obtain logs or subscriber details. Time is critical: without swift, standardized preservation requests and cooperation frameworks, crucial attribution data may be lost. Practical strategies include using structured takedown and preservation templates, engaging platform trust-and-safety channels swiftly, leveraging mutual legal assistance mechanisms where available, and seeking court orders tailored to dynamic scenarios (e.g., orders that cover mirrors, re-uploads, and algorithmic propagation).

Mitigation and remedy design must reflect the speed and persistence of online harm. Even after a takedown, mirrors and reposts can perpetuate damage. Effective relief often includes dynamic injunctions covering derivative uploads and hashed matching to block re-appearance at scale, de-indexing orders to curb discoverability, and structured right-of-reply or correction

[13] Adobe, BBC, Microsoft, & Publicis Groupe. (2023). The Coalition for Content Provenance and Authenticity (C2PA) technical specification 2.0. C2PA. https://c2pa.org

[14] National Research Council. (2009). Strengthening forensic science in the United States: A path forward. National Academies Press. https://doi.org/10.17226/12589

mechanisms to counter reputational harm.[15]In intimate or safety-critical cases, accelerated timelines, complainant anonymity, and mandated platform response windows are crucial to meaningful redress.

Ultimately, meeting the enforcement and evidentiary challenge requires a layered architecture: specialized investigative capacity in digital forensics; judicial training to evaluate technical claims; standardized protocols for detection, preservation, and expert reporting; platform obligations for rapid preservation and action after notice; and streamlined cross-border cooperation. Without these coordinated elements, even well-crafted laws will struggle to deliver timely, reliable outcomes in an information environment where convincing falsehoods can be produced and propagated faster than traditional legal processes can respond.

7. National Security and Public Order Concerns

Deepfake technology introduces a profound shift in the threat landscape for national security and public order because it enables the rapid, low-cost fabrication of convincing audio-visual content designed to distort reality. In the electoral context, synthetic videos or audio can be strategically timed to appear just before voting or major political developments, fabricating scandals, or attributing inflammatory remarks to candidates. Even when such content is debunked, the damage to public perception can persist, weakening electoral integrity and trust in institutions. This phenomenon operates alongside the "liar's dividend," where genuine evidence is casually dismissed as fake, further corroding public confidence in authentic records and widening the space for manipulation.

From a security standpoint, realistic impersonation of public officials, ministers, senior military personnel, or police leadership can generate operational confusion and panic. A fabricated address announcing emergency measures, curfews, troop deployments, or financial restrictions can trigger public disorder, market disruption, or misallocation of critical resources. Real-time synthetic audio on calls or virtual briefings can also be weaponized to issue fraudulent directives, undermining command-and-control structures and complicating crisis response. These scenarios are intensified by the speed and virality of social media and encrypted messaging platforms, which can outpace official verification and rebuttal efforts.

[15]Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954

Communal harmony is particularly vulnerable to deepfakes that exploit religious sentiment or identity-based tensions. Manipulated content that appears to depict desecration, insults to faith, or targeted violence can catalyse retaliatory actions and mass mobilization before authorities can intervene. In such circumstances, the ability of law enforcement and civil administration to maintain order depends on rapid rumour control, verified communication channels, multilingual clarifications, and coordinated engagement with community leaders. Without these measures, localized incidents can rapidly escalate, straining policing capacity and public health and safety systems.[16]

Existing legal provisions relating to public order and cyber offenses provide partial recourse but are not tailored to the nuances of synthetic media. While laws addressing incitement, dissemination of obscene or harmful content, impersonation, and cyber fraud can be invoked, they often lack precise definitions that distinguish manipulated from fully synthetic media or recognize the gradations of harm- from satire and artistic expression to intentional deception designed to incite violence or interfere with elections. This can hinder timely intervention and complicate prosecutorial strategies, especially where intent, scale of dissemination, and resultant harm must be carefully demonstrated.[17]

A central challenge is balancing security imperatives with civil liberties. Overbroad regulation risks chilling legitimate expression, including satire, documentary reenactments, and journalism that responsibly uses synthetic techniques for storytelling or public education. Conversely, permissiveness can embolden malicious actors. A proportionate approach emphasizes context and intent: requiring clear disclosure for materially synthetic political content distributed at scale, focusing restrictions on high-harm scenarios such as incitement or impersonation of public authorities, and embedding due process through notice, appeals, and independent oversight. Such an approach preserves a space for free expression while enabling targeted, timely responses to genuine threats.

Technical measures can further reduce systemic risk. Provenance tools and cryptographic signing at the point of capture enable audiences to verify official communications. Watermarking standards for synthetic media and interoperable labelling help platforms and users identify manipulated content. Integrating detection tools into the vetting of material claiming to originate from authorities can prevent inadvertent amplification of fakes.

[16]Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework. Council of Europe. https://edoc.coe.int

[17] Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954

Collaborative hash-sharing and threat intelligence across agencies and platforms support quick suppression of re-uploads and variants, limiting residual harm.[18]

Cross-border dynamics complicate enforcement and mitigation. Deepfake campaigns can leverage foreign infrastructure, anonymized accounts, and global distribution channels, making attribution and data access difficult. Practical responses include rapid preservation requests to platforms, standardized data-disclosure templates, and participation in bilateral and multilateral cooperation frameworks for expedited log-sharing and lawful access. Engagement in international dialogues on provenance standards, labelling norms, and election-period safeguards helps align domestic practice with emerging global best efforts.

8. Comparative Analysis of Global Legal Approaches to Deepfake Regulation and Their Relevance to India

Around the world, regulators are converging on a common recognition: deepfakes demand targeted, layered responses that combine clear legal prohibitions for harmful uses, transparency obligations for synthetic media, and platform accountability. Yet approaches vary across jurisdictions in scope, emphasis, and the balance they strike between innovation and protection. Understanding these models helps identify actionable options for India that are consistent with constitutional guarantees and enforcement realities.

In the United States, the regulatory landscape is plural and largely state-driven, supplemented by thematic federal proposals. Several states criminalize specific high-harm categories, notably non-consensual sexually explicit deepfakes and election-related synthetic media within blackout windows preceding voting. Civil remedies often accompany criminal sanctions, enabling victims to seek injunctions, takedowns, and damages. This scheme operates alongside entrenched protections for free expression, which keeps the focus on intent (malicious deception), context (election interference, intimate imagery, fraud), and harm (reputational, economic, public order). The practical lesson is a calibrated, category-focused toolkit: narrowly tailored prohibitions where the risk is highest; safe harbours for satire, news, and artistic works when disclosed; and a strong role for civil liability to complement prosecution.

[18] International Telecommunication Union. (2024, May). AI watermarking: A watershed for multimedia authenticity. ITU Hub. https://www.itu.int/hub/2024/05/ai-watermarking-a-watershed-for-multimedia-authenticity/

The European Union proceeds through comprehensive, system-level instruments. Data protection rules constrain the collection and processing of biometric identifiers, such as faces and voices, without a lawful basis or clear, informed consent. Complementary AI governance measures emphasize transparency and traceability, requiring clear disclosure when content is AI-generated or materially manipulated and encouraging technical provenance (e.g., watermarking, cryptographic signatures) to help users and platforms distinguish synthetic media. This model's strength lies in imposing upstream duties on providers and deployers of generative systems, not only on end-users, which promotes responsible design, record-keeping, and labelling. It also harmonizes enforcement across member states, reducing fragmentation and creating predictable obligations for industry.[19]

China adopts a highly prescriptive posture with mandatory labelling, real-name and service-provider verification duties, and swift administrative enforcement against synthetic content deemed deceptive, destabilizing, or harmful. Providers must implement detection, takedown, and user controls by default. The result is a prevention-first regime that prioritizes social stability and rapid mitigation, with heavier ex ante compliance burdens on platforms and developers. While this ensures speed and clarity in enforcement, it reflects a governance philosophy that is not readily transferable to systems with strong speech protections, and thus must be selectively adapted if considered elsewhere.

Other jurisdictions illustrate additional tools. The United Kingdom embeds duties within online safety legislation requiring platforms to manage harmful content risksincluding intimate-image abuse using synthetic mediathrough risk assessments, user reporting pathways, and timely remedies. Some countries are experimenting with treating a person's likeness (face, voice) as a protectable interest akin to intellectual property or personality rights, strengthening individuals' control over synthetic use of their identity and clarifying commercial consent requirements. Sectoral measuressuch as election integrity codes, broadcaster/advertiser standards for disclosures, and age-assurance obligations for explicit contentround out a layered approach.

These comparative strands point to a set of concrete, India-relevant design choices:

1. Define and classify deepfakes with precision. Distinguish manipulated versus fully synthetic media; identify aggravating contexts (intimate imagery, impersonation of

[19] ArtificialIntelligenceAct.eu. (2025). EU Artificial Intelligence Act: Up-to-date developments and guidance. https://artificialintelligenceact.eu

public authorities, election interference, extortion/fraud); and expressly protect satire, parody, research, and documentary re-enactment when transparently disclosed. This reduces overreach while equipping authorities to act where harms are acute.

2. Anchor identity, autonomy, and consent. Statutorily recognize control over one's likeness and voice, require explicit, purpose-bound, revocable consent for commercial or high-risk synthetic uses, and create expedited civil remedies—interim injunctions, dynamic takedowns, stay-down orders using hash or perceptual matching—for victims. Extend protection beyond celebrities so ordinary citizens have clear recourse.
3. Mandate transparency for scaled distribution. Require conspicuous, durable labels for materially synthetic political or public-interest content distributed at scale, coupled with machine-readable provenance. Encourage or require watermarking/provenance at generation by major tools to support downstream detection, while preserving exemptions for protected speech that is clearly disclosed as synthetic.
4. Calibrate platform obligations by risk. Impose tiered duties on intermediaries: rapid removal service-level commitments for intimate or impersonation deepfakes; trusted-flagger fast lanes for verified victims and authorities; persistent stay-down for adjudicated harms; transparent reporting on detection, removals, appeals, and error rates; and user-facing tools to report and contextually flag suspected synthetic content.
5. Strengthen evidentiary and enforcement infrastructure. Establish baseline forensic standards and expert-practice guidelines for courts; formalize rapid preservation and disclosure protocols with platforms; and invest in public digital forensics capacity to reduce dependence on private tooling. Clear, uniform procedures minimize "*battle of experts*" disputes and accelerate relief in time-sensitive cases.
6. Safeguard constitutional rights through necessity and proportionality. Embed intent/context tests, narrow tailoring, and due process (notice, reasons, appeals) into takedown and blocking powers. Prefer labels and downranking over removal where feasible; reserve emergency removal for high-harm categories; and protect bona fide journalism, research, and artistic expression that uses synthetic techniques transparently and responsibly.
7. Address elections and public order with special measures. For election periods, consider limited-time rules targeting deceptive deepfakes about candidates, voting procedures, or institutions, with accelerated review, clear disclosure duties, and

independent oversight. For public order, formalize rumours-control protocols, verified official channels, and cross-platform coordination to contain high-risk synthetic triggers.

8. Coordinate internationally. Develop standardized preservation requests and cooperation templates for cross-border data; participate in dialogues on provenance standards and election-period safeguards; and align penalties for platform non-compliance with global practice to avoid regulatory arbitrage.

9. Legal Framework and Regulatory Measures in India Addressing Deepfake Technology: Current Status and Challenges

India's current approach to deepfakes is built on a patchwork of existing statutes, rulemaking for digital intermediaries, judicially recognized personality rights, and operational measures by cyber agencies, rather than a single dedicated law targeting synthetic media. This mosaic offers meaningful points of entry for redress and enforcement, yet important gaps remain because these instruments were not crafted with AI-generated content in mind and therefore struggle with the speed, scale, and anonymity that characterize deepfake misuse.[20]

At the core of India's cyber regime is the Information Technology Act, 2000 (IT Act), which provides substantive and procedural levers to address several deepfake harms. Provisions related to impersonation and cheating by personation in electronic communications can be invoked where synthetic media is used to deceive or defraud. Privacy-centric provisions penalize the capture, distribution, or publication of intimate imagery without consent, and content provisions addressing obscenity and sexually explicit material apply directly to non-consensual intimate deepfakes. Blocking powers permit targeted restriction of access to unlawful content in the interests of public order and security, while safe-harbour provisions set the conditions under which intermediaries remain shielded from liability-conditions that hinge on due diligence and prompt removal upon notice. These statutory anchors, though technology-neutral, give authorities and complainants a legal pathway to seek takedowns, initiate criminal action, and preserve evidence.

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, translate these statutory principles into concrete duties for platforms. Intermediaries must maintain grievance redress mechanisms, act expeditiously on complaints about

[20] Press Information Bureau. (2023, December 26). India well-equipped to tackle evolving online harms and cybercrimes. Government of India. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2154268

morphed, impersonated, or otherwise unlawful content, and retain essential records for investigative cooperation. Failure to follow due diligence can erode safe-harbour protection, creating stronger incentives for timely action. In practice, this means that platforms are expected to remove flagged deepfakes within prescribed timelines, preserve logs on request, and support law enforcement with originator information where legally permissible. Amendments and government advisories in recent years have further emphasized accountability for synthetic media, signalling a policy direction that treats deepfakes as a priority online harm.[21]

Beyond platform governance, traditional intellectual property protections continue to play a role. The Copyright Act, 1957, offers civil and criminal remedies when copyrighted images, videos, and audio recordings are reproduced, adapted, or communicated to the public without authorization in synthetic outputs. This is particularly relevant where deepfakes are built on identifiable copyrighted source material- film footage, photographs, sound recordingsor where the resulting work is derivative of protected expression. Although Indian copyright law does not yet articulate a comprehensive doctrine for AI-generated outputs, the combination of exclusive rights, authors' moral rights, and remedies for infringement equips rightsholders to challenge unauthorized uses embedded in deepfakes. [22]Personality rights, while primarily judge-made, bridge another gap by protecting an individual's name, image, voice, and signature traits against unauthorized commercial exploitation. Courts have issued sweeping injunctions restraining synthetic misuse of celebrity likenesses, underscoring that identity appropriation via AI or morphing can be restrained to preserve dignity and prevent unjust enrichment.

Institutionally, India's cyber-response bodies support enforcement at the operational layer. Agencies tasked with incident response and cybercrime coordination disseminate advisories on synthetic media risks, guide preservation and investigation practices, and facilitate cooperation among state police units, platforms, and technical experts. These actions matter because deepfake cases often turn on rapid response, swift flagging, preservation of high-

[21] Ministry of Electronics and Information Technology. (2021, February 25; amended 2022 & 2023). Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Government of India.

[22] International Journal of Scientific Development and Research. (2024, March). The evolution of deepfake laws: Adapting copyright and intermediary liability frameworks (IJSDR2403094). https://ijsdr.org/papers/IJSDR2403094.pdf

quality artifacts, log retention, and coordinated takedownto mitigate harm before content metastasizes across platforms.[23]

Despite these strengths, structural challenges persist. The most significant is the absence of a statutory definition of "deepfake" and a graded, harm-based framework tailored to synthetic media. Without precise definitions that distinguish manipulated from fully synthetic content, and without calibrated categories that reflect intent and impact (e.g., intimate-image abuse, impersonation of public authorities, electoral interference, fraud/extortion), enforcement can be inconsistent and reactive. The technology-neutral posture of current laws helps them endure, but it also leaves ambiguity at crucial junctures: whether and how training-data ingestion implicates copyright, how to assess the status of AI-synthesized outputs (derivative, transformative, or unauthorized reproduction), and what standard should govern consent for use of likeness and voice in synthetic content.

Enforcement is further complicated by the velocity and virality of deepfakes. Takedown timelines often lag the lifecycle of online harm, and the same clip may reappear via mirrors and re-uploads unless persistent stay-down measures, hash-matching, and perceptual detection are in place. Attribution is difficult: creators can hide behind anonymization and cross-border infrastructure, shifting the burden to downstream actors and platforms while victims scramble for quick relief. The evidentiary layer also requires modernization- uniform guidance on digital preservation, chain of custody, and admissibility of expert forensic analysis- so that courts can confidently evaluate authenticity, falsity, and responsibility within compressed timelines typical of public-order or reputational crises.

To close these gaps, a dedicated, narrowly tailored statute or a set of targeted amendments could anchor a modern deepfake governance framework. Such a measure would define manipulated and fully synthetic media; establish aggravated offenses for high-harm categories (intimate deepfakes, impersonation of public officials, election interference, coordinated fraud); and codify explicit, purpose-bound, and revocable consent for identity use in synthetic content. It would also clarify the intellectual property status of AI outputs and input use, aligning with fair dealing while protecting creators' and performers' rights. On the platform side, risk-tiered obligations could mandate: accelerated response times for high-harm categories; trusted-flagger fast lanes for verified victims and authorities; transparent

[23] Vivekananda International Foundation. (2025, April 28). Bharatiya laws against deepfake cybercrime: Opportunities and challenges. https://www.vifindia.org/article/2025/april/28/Bharatiya-Laws-Against-Deepfake-Cybercrime-Opportunities-and-Challenges

labelling and provenance for materially synthetic content distributed at scale; and stay-down requirements for adjudicated violations using hash-based and perceptual matching. Procedurally, fast-track civil remedies- interim and dynamic injunctions that cover mirrors and re-uploads, complainant anonymity in intimate cases, and court-ordered de-indexing—would provide practical relief. At the investigative layer, standardized preservation notices, clear timelines for log retention and disclosure, and capacity building in digital forensics for police and prosecutors would strengthen attribution and evidentiary reliability.[24]

Public education and media literacy are essential complements. Clear guidance on recognizing and reporting suspected deepfakes, combined with widely trusted official communication channels for rapid rebuttal, reduces both the spread and the staying power of harmful fabrications. Finally, because deepfakes are a cross-border phenomenon, India's framework should be paired with pragmatic cooperation instruments, standardized templates for preservation and data access requests, participation in international work on provenance and labelling standards, and bilateral pathways for expedited lawful accessto prevent jurisdictional barriers from undermining timely remedies.[25]

In its current form, India's legal toolkit can address many deepfake harms through a combination of cybercrime provisions, platform due diligence, copyright enforcement, and personality-rights injunctions. But to meet the momentwhere convincing falsehoods can be generated and propagated in minutes, the system needs dedicated definitions, calibrated offenses and remedies, modern evidentiary standards, and enforceable platform duties designed for synthetic media. A carefully crafted, rights-respecting framework that prioritizes consent, transparency, proportionality, and rapid redress will allow India to protect dignity, reputation, democratic processes, and innovation in the synthetic media era.

10. Proposals for Legal Reforms and Policy Recommendations

A credible, future-ready response to deepfake harms in India must be precise in scope, swift in remedy, and balanced in protecting both rights and innovation. Consent and identity autonomy must be central pillars. A statutory, purpose-bound consent standard for using a person's likeness or voice in synthetic media should be explicit, specific, informed, revocable, and time-limited. Consent for one use (e.g., a headshot) must not be stretched to cover unrelated synthetic exploitation (e.g., intimate fakes or political speech). For minors

[24] National Research Council. (2009). Strengthening forensic science in the United States: A path forward. National Academies Press. https://doi.org/10.17226/12589

[25] Citron, D. K., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1819. https://doi.org/10.2139/ssrn.3213954

and intimate contexts, the law should mandate heightened safeguards, including parental or guardian oversight, mandatory age and identity verification for uploaders of sensitive content, and per se offenses for intimate deepfakes without express consent. Alongside, personality rights should be codified for all personsnot just celebritiesto restrain unauthorized commercial and harmful uses of identity, with swift interim injunctions and damages geared to viral reach and persistence.[26]

Intellectual property alignment requires clarity at both input and output stages. Input protection should confirm that ingesting proprietary photos, videos, or audio into a synthetic workflow without authorization can constitute infringement, particularly for commercial or harmful uses. Output guidance should define when a synthetic work is a derivative, an unauthorized reproduction, or a transformative use, with fair dealing preserved for criticism, review, scholarship, and clearly labelled parody. Moral rightsattribution and integrityshould extend to prohibit synthetic mutilation or misattribution of creators' works. Performers' rights and neighbouring rights also need explicit recognition where a performer's voice or performance is cloned without consent.[27]

Platform accountability should be risk-tiered and time-bound. Intermediaries that host and distribute content at scale ought to meet graduated obligations: fast-lane review for intimate deepfakes and impersonation of public authorities; strict service-level timelines for removal; trusted-flagger channels for verified victims, election authorities, and law enforcement; user-facing transparency on labelling decisions and appeals; persistent stay-down mechanisms using cryptographic hashes and perceptual matching for adjudicated violative clips; and regular reporting on detection efficacy, error rates, and response times. For election periods and public-order emergencies, narrowly tailored surge protocolstemporary acceleration of review, geofenced demotion/removal under judicial or independent oversight, and mandatory provenance labelscan prevent last-minute manipulation while preserving due process.

Institutional capacity-building should proceed on three tracks. First, specialized cyber-forensics units in police and prosecution services with tools and training for detection, provenance tracing, and platform cooperation. Second, judicial training on evaluating technical evidence, distinguishing malicious deepfakes from protected expression, and

[26] European Data Protection Board. (2020). Guidelines 05/2020 on consent under Regulation 2016/679 (Version 1.1). European Union.

[27] World Intellectual Property Organization. (2023). WIPO issues paper on intellectual property policy and artificial intelligence (2nd ed.). World Intellectual Property Organization. https://www.wipo.int/edocs/pubdocs/en/wipo-pub-1055.pdf

crafting proportionate orders. Third, integrated coordination cells that link law enforcement, platform trust-and-safety teams, and public information officers to accelerate takedowns, counter-messaging, and evidence preservation.[28]

Technology and standards must complement the law. Encourage provenance-by-design in official communications (cryptographic signing/watermarking of government releases) so the public can verify authenticity. Promote interoperable watermarking and labelling norms for major generation tools to aid downstream detection. Support open evaluation benchmarks for detectors, with transparent accuracy and bias reporting, to guide judicial reliance and procurement by agencies. [29]Where feasible, incentivize or mandate adoption of provenance standards in sensitive domains-elections, public safety, and financial communications.

Public education is the social backbone of resilience. Media literacy campaigns should help citizens recognize manipulation cues, understand labels and provenance signals, and know how to report and seek redress. Survivor-cantered support-confidential reporting channels, anonymity protections in intimate-image cases, counselling, and step-by-step takedown assistance-must be embedded in institutional practice. Universities, NGOs, and industry associations can partner on curricula, outreach, and independent audits of platform safeguards.

Because deepfakes are transnational, international cooperation is indispensable. Establish standardized preservation requests and rapid data-freeze templates for platforms; negotiate bilateral and multilateral pathways for expedited lawful access to logs and subscriber information; and participate in global dialogues on election-period safeguards, provenance standards, and cross-border enforcement. Aligning select provisions with mature regimes abroad-while honouring India's constitutional commitments-reduces regulatory arbitrage and accelerates practical remedies.Taken together, these reforms create a layered, proportionate architecture: clear definitions; robust consent and identity rights; calibrated IP protection; time-bound, risk-tiered platform duties; modern evidentiary standards; institutional capacity; technical provenance; public education; and cross-border cooperation. Such an approach protects dignity, reputation, and democratic processes while preserving space for journalism, satire, research, and responsible innovation.

[28] National Institute of Standards and Technology. (2006). NIST SP 800-86: Guide to integrating forensic techniques into incident response. U.S. Department of Commerce. https://doi.org/10.6028/NIST.SP.800-86

[29] Wang, S.-Y., Wang, O., Zhang, R., Owens, A., & Efros, A. A. (2020). CNN-generated images are surprisingly easy to spot… for now. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8695–8704. https://doi.org/10.1109/CVPR42600.2020.00872

Conclusion

Deepfake technology compresses cost and time to manufacture convincing falsehoods, directly challenging legal systems that depend on authentic records and traceable accountability. In India, the harms are multidimensionalprivacy violations, non-consensual intimate imagery, defamation, fraud, and threats to electoral integrity and public order yet the legal apparatus remains largely technology-neutral and reactive. Existing cybercrime provisions, platform due diligence rules, copyright and performers' rights, and judge-made personality rights provide meaningful footholds, but they do not supply the precision, speed, and clarity needed for synthetic media's scale and velocity.The core difficulty lies at three pressure points: reliable detection and admissible proof in court; rapid, durable remedies against viral spread and re-uploads; and credible attribution when creators hide behind anonymity and cross-border infrastructure.Withoutstatutory definitions and graded offenses tailored to deepfakes, authorities must stretch general provisions to novel facts, risking inconsistent outcomes. Without codified consent standards and identity rights, individualsespecially non-celebritieslack clear, swift recourse when their likeness and voice are exploited. And without risk-tiered platform obligations, takedowns and stay-downs may arrive too late to prevent reputational and public-order damage.

A purpose-built, rights-respecting framework addresses these gaps. By defining synthetic media precisely and calibrating offenses by harm and intent, the law can target truly dangerous conduct while safeguarding satire, scholarship, and labelled artistic uses. By centring consent and identity autonomy, it empowers individuals against exploitation. By clarifying IP treatment at both input and output stages, it protects creators and performers without suffocating fair dealing and transformation. By imposing time-bound, risk-tiered duties on platforms, fast lanes for high-harm cases, persistent stay-down for adjudicated content, transparency, and appeal rights, it aligns incentives with public safety and due process. And by modernizing evidentiary protocols, training courts and investigators, and institutionalizing rapid preservation and cooperation with platforms, it turns technical capability into courtroom-ready proof.Technology and governance must move in lockstep. Provenance and watermarking standards, cryptographic signing of official communications, detector evaluation benchmarks, and cross-platform hash-sharing can reduce systemic risk and residual harm. Public literacy initiatives build societal antibodies against deception and blunt the *liar's dividend*, where wrongdoers dismiss authentic evidence as fake. Finally, cross-border cooperation is non-negotiable in a world where creation, hosting, and

consumption span jurisdictions; streamlined lawful access and shared norms on labelling and election safeguards are essential.India can meet this challenge without compromising constitutional freedoms. The path forward is not maximal censorship but precise, proportionate rules that focus on malicious deception and high-harm contexts, paired with transparent labels, due process, and robust remedies. With carefully drafted legislation, empowered institutions, accountable platforms, and an informed public, India can protect citizens' dignity and democratic resilience while cultivating a responsible, innovative ecosystem for synthetic media.