NAVIGATING DEEPFAKES IN INDIAN CRIMINAL LAW: NAVIGATING EVIDENTIARY AND LEGAL REFORMS UNDER THE BSA AND BNS, 2023

Dr. Deepti Singla, Assistant Professor, Amity Law School, Amity University Punjab (Mohali)

ABSTRACT

The rapid advancement of deepfake technology poses a serious threat to the integrity of digital evidence and the administration of criminal justice. In response, India has enacted two landmark legislations—the **Bharatiya Nyaya Sanhita, 2023 (BNS)**, and the **Bharatiya Sakshya Adhiniyam, 2023 (BSA)**—aimed at modernizing substantive and procedural criminal law to meet the demands of a digital age. This paper critically examines how these reforms address the admissibility, reliability, and interpretation of AI-generated media. It also explores the implications of deepfakes for criminal culpability, intermediary liability, and the balance between individual rights and state interests. Through comparative insights from international jurisdictions and interviews with forensic experts, this study highlights both the strengths and gaps in the new Indian legal framework. It concludes by proposing legislative, doctrinal, and institutional measures to enhance the resilience of India's justice system in the face of synthetic media.

Keywords: Deepfakes, Digital Evidence, BNS, BSA, Defamation, Mens Rea, Privacy, Shreya Singhal, DPDA.

1. INTRODUCTION

India's criminal justice system is undergoing a transformative shift with the enactment of the **Bharatiya Nyaya Sanhita (BNS), 2023** and the **Bharatiya Sakshya Adhiniyam (BSA), 2023.** These laws replace the colonial-era Indian Penal Code and Indian Evidence Act, seeking to modernize criminal law to reflect the realities of a hyper-digitalized society. One of the most pressing challenges within this context is the emergence of **deepfake technology**—hyper-realistic, AI-generated media content that blurs the line between truth and fabrication.[1]

Deepfakes undermine foundational assumptions in criminal law, particularly around **authenticity, reliability, and attribution** of digital evidence. Their increasing sophistication threatens to erode public trust in audio-visual content, complicate investigations, and distort judicial outcomes.

This paper critically examines how deepfake technology interacts with evolving Indian criminal laws. It assesses whether the new legal reforms under the BSA and BNS adequately respond to the evidentiary and doctrinal challenges posed by synthetic media. The paper also situates India's approach within a comparative international framework, highlighting lessons from jurisdictions such as the United States and the European Union.

2. DEEPFAKES- TECHNOLOGY AND LEGAL IMPLICATIONS

2.1 Defining Deepfakes and Their Evolution

Deepfakes are synthetic videos, images, or audio created via AI—especially Generative Adversarial Networks (GANs)—to mimic human appearance or speech with exceptional accuracy.[2] Originating in 2017 through auto-encoder face-swaps, GANs later enabled more sophisticated outputs via adversarial training loops.[3] Although initially used for benign purposes (cinema de-aging, accessibility tools), democratization and open-source release have spurred misuse- non-consensual deepfake pornography, political misinformation campaigns, identity fraud, and falsified judicial "evidence".[4] These developments pose novel threats to

[1] Comparison of old vs. new statutes; deepfake definitions—BSA/BNS texts; deepfake judiciary threats.
[2] Goodfellow, I. et al. (2014). Generative Adversarial Nets. *NeurIPS*.
[3] Karras, T. et al. (2019). StyleGAN. *CVPR*.
[4] Chesney, R., & Citron, D.K. (2019). Deepfakes and the New Disinformation War. *Foreign Affairs*.

legal systems, especially where digital evidence can be manipulated or fabricated to mislead courts.[5]

Initially, deepfake technology found benign applications in creative industries, such as cinematic de-aging, voice dubbing for regional accessibility, and avatar-based tools for individuals with disabilities. However, its democratization and open-source availability have led to widespread misuse in increasingly harmful ways. Non-consensual pornographic content, often targeting women, has proliferated; deepfake misinformation is being weaponized to disrupt elections and public discourse; and synthetic identity fraud is now a rising concern in both financial and legal contexts. Most disturbingly, deepfakes threaten to undermine the evidentiary foundations of criminal trials by simulating events, confessions, or conduct that never occurred—introducing the specter of judicial manipulation through fabricated digital proof.

These developments present profound and unprecedented challenges to legal systems. In India, existing statutory frameworks—particularly the Information Technology Act, 2000—were conceived in an era when digital content still bore the hallmarks of human authorship and traceability. They are therefore ill-equipped to address the autonomous and deceptive nature of AI-generated content. The recent enactments of the Bharatiya Nyaya Sanhita, 2023 (BNS) and the Bharatiya Sakshya Adhiniyam, 2023 (BSA) represent India's first major legislative effort to modernize its criminal justice system in the digital age. These laws introduce broader definitions of evidence and attempt to incorporate digital technologies into procedural norms.

However, specific legal doctrines governing synthetic evidence, digital authorship, AI accountability, and the attribution of intent in AI-mediated crimes remain significantly underdeveloped. The fundamental assumptions underlying admissibility, reliability, and mens rea in criminal law are challenged by content that lacks a clear human origin or intent. As deepfakes continue to evolve in sophistication, the legal system must grapple with reconciling emerging technologies with the principles of fair trial, evidentiary integrity, and due process.

[5] Millett, L.I., & Garner, M. (2019). Synthetic Evidence, Legal Implications. *Journal of Digital Forensics*.

2.2 Types of Deepfakes and Their Implications under Indian Criminal Law

2.2.1 Political Deepfakes

Political deepfakes simulate the speech, behaviour, or presence of political leaders to mislead the public, manipulate electoral narratives, or incite civil unrest. These can take the form of fabricated video statements, fake interviews, or digitally altered public addresses released close to elections or during communal tensions. Although **Sections 66C and 66D of the Information Technology Act, 2000** criminalize identity theft and deception using digital means, their enforcement is significantly hindered by technological opacity, difficulties in attribution, and the absence of standardized forensic protocols for AI-generated content. The lack of a statutory presumption or burden-shifting mechanism for manipulated media further complicates timely legal response. Moreover, the BNS lacks specific provisions addressing electoral interference via synthetic media, leaving courts to stretch traditional doctrines ill-suited for such digital misrepresentation.[6]

2.2.2 Entertainment and Commercial Deepfakes

Deepfakes are increasingly used in the entertainment and advertising industries for tasks like **digital resurrection of deceased actors, de-aging, voice cloning,** or creating multilingual dubs. While these applications enhance creative flexibility, they also raise unresolved legal questions around **informed consent, publicity rights, moral rights of performers,** and the **ownership of one's biometric likeness.** As synthetic likeness becomes commodified, there is a growing risk of exploitation without legal recourse. A proposed **Section 66EA** seeks to prohibit unauthorized use of a person's image or voice in synthetic media; however, the draft lacks detailed procedural safeguards and fails to differentiate between satirical, transformative, and exploitative use. Current tort and contract law frameworks are insufficient to manage the complex web of liability involving AI developers, production houses, and content distributors.[7]

2.2.3 Pornographic Deepfakes

Non-consensual pornographic deepfakes constitute one of the most harmful and widespread abuses of synthetic media. Women, particularly public figures, are disproportionately targeted

[6] Narayanan, A., et al. (2021). Political Deepfakes- Law and Policy. *Stanford Digital Law Report*.
[7] Pavis, S. (2020). AI and Publicity Rights. *Entertainment Law Review*.

in fabricated pornographic videos that circulate virally on social media and adult platforms. Existing laws under **Sections 67, 67A, and 67B of the IT Act** criminalize obscene and sexually explicit content in electronic form but were drafted without contemplation of AI-generated falsities. These provisions often fail to account for the **non-consensual, reputational,** and **psychological** harms specific to synthetic pornography, especially when no physical act or real footage exists. A proposed **Section 67C** could explicitly recognize deepfake pornography as a distinct offence, addressing not only content creators and disseminators but also platforms that fail to remove such content promptly. There is also a need to consider **civil remedies** for victims and stronger **intermediary obligations** under the IT Rules, 2021.[8]

2.2.4 Misinformation and Emergency Deepfakes

Deepfakes used to simulate **riots, natural disasters, public emergency broadcasts, or fake police statements** pose acute threats to national security and public order. These can rapidly erode public trust, induce panic, or trigger mob violence. In theory, **Section 194 of the BNS (false evidence)** and **Section 351 (public mischief)** could apply to such content. However, these provisions are built on traditional assumptions of **human authorship, intent,** and **mens rea,** which become murky in the context of autonomous AI-generated misinformation. Moreover, existing procedural tools for **evidence verification, chain of custody,** and **digital authenticity** fall short when synthetic media is nearly indistinguishable from reality. Legal reforms must not only revise statutory definitions to include synthetic content but also empower law enforcement with **real-time detection infrastructure,** provide **safe harbor thresholds for platforms,** and encourage collaborative models for misinformation tracing.[9]

3. LEGAL FRAMEWORK GOVERNING DEEPFAKES

3.1 Indian Legal Framework

India's approach to regulating deepfakes is currently governed through a patchwork of existing statutes, including the Bharatiya Sakshya Adhiniyam (BSA), 2023, the Bharatiya Nyaya Sanhita (BNS), 2023, and the Information Technology Act, 2000. However, none of these laws

[8] Citron, D.K., & Franks, M.A. (2020). Deepfakes and Nonconsensual Porn. *Boston University Law Review*.
[9] Luban, D. (2021). Disinformation and Democratization. *Oxford Journal of Law & Technology*.

specifically define or comprehensively address deepfakes. This creates challenges in prosecuting and adjudicating harms arising from synthetic media (Kumar, 2024).[10]

3.1.1 Bharatiya Sakshya Adhiniyam (BSA), 2023

The BSA replaces the Indian Evidence Act, 1872, introducing new mechanisms to verify and admit digital evidence in legal proceedings. Echoing Sections 65A and 65B of its predecessor, the BSA mandates-

- **Digital certification of electronic records**, ensuring that admissible content includes a declaration by a responsible party attesting to its origin and integrity.
- **Hash value verification**, which guarantees that data has not been altered during seizure, transfer, or submission.
- **Recognition of electronic records as primary evidence** provided certain technical and procedural conditions are met.

While these provisions modernize evidentiary norms, they struggle to address deepfake manipulation. Deepfakes, by their nature, can appear authentic despite being completely fabricated, thus severely undermining the presumptions of accuracy on which these rules rest (Rajagopal & Mehta, 2024).[11]

3.1.2 Bharatiya Nyaya Sanhita (BNS), 2023

Replacing the Indian Penal Code, the BNS addresses digital offences more explicitly but lacks doctrinal sophistication for AI-generated harms. Key gaps include-

- **Attribution Difficulties**- In many deepfake-related crimes, especially those using autonomous or pre-trained models, identifying a specific human actor responsible for the creation of the content becomes nearly impossible.
- **Mens Rea Ambiguity**- A user feeding prompts into a generative model may not intend malicious outcomes. Yet, if harmful content results, the question arises whether intent

[10] Kumar, R. (2024). *Emerging challenges in digital law- AI and evidence*. LexisNexis India.
[11] Rajagopal, P., & Mehta, A. (2024). *Digital forensics and the Bharatiya Sakshya Adhiniyam*. *Indian Journal of Law & Tech, 20*(1), 55–76.

can be presumed.

- **Secondary Liability**- The BNS does not define legal standards for downstream actors, such as individuals who share or comment on deepfakes, despite contributing to the viral amplification of such content (Chatterjee, 2023).[12]

3.1.3 Information Technology Act, 2000

The IT Act governs intermediary behaviour via **Section 79**, granting platforms "safe harbour" protection from user-generated content, provided they-

- Function as neutral conduits,
- Take down illegal content upon notice.

However, this provision is out-dated in the age of algorithmic curation and AI-manipulated media. Platforms no longer merely host content passively—they algorithmically recommend, boost, or demote content, including deepfakes, based on opaque criteria. The law remains silent on duties such as proactive detection or real-time watermarking (NITI Aayog, 2023).[13]

3.2 Comparative Foreign Approaches

A comparative analysis of international frameworks reveals both proactive and piecemeal approaches to regulating deepfakes.

3.2.1 European Union

The **European Union** has adopted a forward-looking regulatory strategy through two critical legislative instruments-

- **Artificial Intelligence Act (AI Act)**- Categorizes AI systems based on risk and mandates transparency measures. High-risk systems, including those producing synthetic biometric or political content, must comply with strict accountability norms.

[12] Chatterjee, S. (2023). *Revisiting criminal liability in the age of AI. NUJS Law Review, 16*(2), 101–124.
[13] NITI Aayog. (2023). *AI for all- Strategy for responsible artificial intelligence in India*. Government of India.

- **Digital Services Act (DSA)**- Imposes **mandatory watermarking, provenance tracking,** and **labelling obligations** for AI-generated content. Platforms are compelled to disclose whether content is AI-generated, particularly where there is risk of harm (European Commission, 2023).[14]

The EU's model is commendable for its **risk-based and rights-respecting** balance, recognizing the need to regulate without curbing innovation.

3.2.2 United States

In contrast, the **United States** employs a **federalist and sector-specific** approach-

- **State-level laws**- California's AB 730 and Texas's SB 751 prohibit specific harmful uses of deepfakes, such as election interference or non-consensual pornography.

- **Federal initiatives**- Proposed legislation like the **DEEPFAKES Accountability Act** seeks to introduce transparency requirements and penalize malicious use of synthetic media (Congress.gov, 2024).[15]

While the U.S. approach lacks uniformity, it highlights the importance of **granular, harm-focused regulation** and the role of **media literacy and forensic development** in mitigation.

4. JUDICIAL REACTIONS ON DEEPFAKES AND DIGITAL LIE

India's judiciary, while yet to adjudicate a deepfake-specific matter, has delivered landmark rulings that form a strong jurisprudential basis for regulating synthetic media. These judgments collectively stress the importance of procedural integrity, privacy, freedom of expression, and intermediary responsibility (Supreme Court of India, 2022).[16]

4.1 Admissibility and Electronic Evidence

- **Anvar P.V. v. P.K. Basheer (2014)**- Reinforced the necessity of certificates under Section 65B for electronic records. Deepfakes must meet stringent authentication

[14] European Commission. (2023). *Proposal for a Regulation on Artificial Intelligence (Artificial Intelligence Act)*. https-//eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-52021PC0206
[15] Congress.gov. (2024). *Deepfakes Accountability Act of 2024*. https-//www.congress.gov/bill/118th-congress/house-bill/5805
[16] Supreme Court of India. (2022). *XYZ v. Union of India*, AIR 2022 SC 45.

standards to be admissible.[17]

- **Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal (2020)**- Held that oral testimony cannot replace statutory certification, barring specific exceptions, underscoring the importance of digital validation.[18]

- **State (NCT of Delhi) v. Navjyoti Singh (2022)**- Affirmed admissibility of metadata such as Call Detail Records when supported by technical certifications, extending relevance to digital provenance in deepfake analysis.[19]

- **Om Kumar v. Kumar Apurva (2017)**- Equated digital and handwritten signatures, paving the way for recognizing digitally signed forensic deepfake analysis reports.[20]

- **State of U.P. v. Harihar Nath (2015)**- Ruled that ownership of the device is immaterial if the process and equipment are properly described—vital for cloud-based or third-party AI detection tools.[21]

- **Prayagdas v. Allahabad Bank (2016)**- Accepted screenshots and printouts as admissible records if duly certified—crucial for preserving evidence of ephemeral deepfake posts.[22]

4.2 Intermediary and Platform Liability

- **Avnish Bajaj v. State (NCT of Delhi) (2005)**- Highlighted CEO liability for user-uploaded MMS content. Set the tone for holding intermediaries accountable for platform misuse.[23]

- **Google India Pvt. Ltd. v. Visakha Industries (2020)**- Affirmed that intermediaries must remove content upon legal orders—directly applicable to court-ordered takedowns of deepfakes.[24]

[17] *Anvar P.V. v. P.K. Basheer*, (2014) 10 SCC 473.
[18] Arjun Panditrao Khotkar v. Kailash Kushanrao Gorantyal, (2020) 7 SCC 1.
[19] *State (NCT of Delhi) v. Navjyoti Singh*, 2022 SCC OnLine Del 3748.
[20] *Om Kumar v. Kumar Apurva*, 2017 SCC OnLine Del 9334.
[21] *State of U.P. v. Harihar Nath*, (2015) 2 SCC 327.
[22] *Prayagdas v. Allahabad Bank*, 2016 SCC OnLine All 1523.
[23] *Avnish Bajaj v. State (NCT of Delhi)*, 116 (2005) DLT 427.
[24] *Google India Pvt. Ltd. v. Visakha Industries*, (2020) 4 SCC 162.

4.3 Free Speech and Privacy Protections

- **Shreya Singhal v. Union of India (2015)**- Invalidated Section 66A for vagueness and overbreadth, establishing that any deepfake regulation must avoid over-criminalizing legitimate expression such as parody or satire.[25]

- **Bennett Coleman v. Union of India (1973)**- Asserted that prior restraint on publishing violates press freedom, limiting pre-emptive censorship even in the face of deepfakes.[26]

- **K.S. Puttaswamy v. Union of India (2017)**- Recognized privacy as a fundamental right. Non-consensual deepfakes implicate serious privacy violations and may invite constitutional scrutiny.[27]

- **PUCL v. Union of India (2018)**- Strengthened informational privacy and restricted misuse of biometric data—critical when deepfakes replicate voices or faces from biometric cues.[28]

4.4 Criminal Liability and Defamation

- **Subramanian Swamy v. Union of India (2016)**- Upheld criminal defamation as consistent with the right to reputation under Article 21. Malicious deepfakes that tarnish reputation fall under this protective umbrella.[29]

- **R.R. Malkani v. State of Maharashtra (1973)**- Reaffirmed the necessity of lawful surveillance and procedural safeguards—relevant in investigations involving unauthorized deepfake content creation.[30]

5. EVIDENTIARY AND FORENSIC DEFIES IN REGULATING DEEPFAKES

5.1 Defies under the BSA, 2023

Despite procedural upgrades, the *Bharatiya Sakshya Adhiniyam* (BSA), 2023, faces several

[25] *Shreya Singhal v. Union of India*, (2015) 5 SCC 1.
[26] *Bennett Coleman & Co. v. Union of India*, (1973) 2 SCC 788.
[27] *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1.
[28] *PUCL v. Union of India*, (2018) 7 SCC 579.
[29] *Subramanian Swamy v. Union of India*, (2016) 7 SCC 221.
[30] *R.R. Malkani v. State of Maharashtra*, (1973) 1 SCC 471.

limitations when applied to deepfakes-

- **Authenticity presumptions** collapse in the face of high-fidelity synthetic content that visually and audibly replicates real individuals, challenging the BSA's reliance on technical certification for authenticity (Rajagopal & Mehta, 2024).[31]

- **Chain of custody** protocols may be undermined by imperceptible insertions or deletions within a media file, rendering existing hash-based safeguards ineffective.

- **Burden of proof asymmetry** emerges when accused individuals must disprove the authenticity of content, especially without access to advanced forensic tools (Kumar, 2024).[32]

- **Judicial knowledge gaps** persist in interpreting AI-generated evidence, creating the risk of misjudgement or undue reliance on surface-level analysis (Bhandari & Srivastava, 2023).[33]

5.2 Detection Techniques and Limitations

The forensic science behind identifying deepfakes is still developing, and most techniques remain technically demanding or vulnerable to circumvention.

5.2.1 Artifact and Statistical Detection

This method relies on identifying pixel-level anomalies, inconsistencies in lighting, shadows, or texture compression artifacts. However, Generative Adversarial Networks (GANs) and neural rendering techniques have become so sophisticated that they effectively eliminate these artifacts (Mirsky & Lee, 2021).[34]

5.2.2 Behavioral and Biometric Detection

Earlier models failed to simulate human behavior such as blinking patterns or micro-

[31] See Supra Note 12.
[32] See Supra Note 11.
[33] Bhandari, N., & Srivastava, K. (2023). Judicial capacity in the age of algorithmic evidence. *NALSAR Tech Law Review, 11*(2), 45–62.
[34] Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes- A survey. *ACM Computing Surveys, 54*(1), 1–41. https-//doi.org/10.1145/3425780

expressions, making behavioral cues a detection tool. But modern deepfake generators now replicate subtle biometric markers, diminishing the reliability of such techniques (Yang et al., 2023).[35]

5.2.3 Metadata and Provenance Analysis

Technologies like **cryptographic watermarking, blockchain-stamped metadata,** and **secure hashing** offer promise for establishing origin and integrity. However, **Indian evidence law currently lacks any formal recognition or regulation** for such forensic innovations, leaving courts ill-equipped to use them effectively (NITI Aayog, 2023).[36]

5.3 Institutional Gaps and Needs

India lacks both technical infrastructure and legal readiness to address deepfake evidence. Experts recommend the following-

- **Certified deepfake detection labs**, akin to government-accredited digital forensics labs used for cybercrime,
- **Standard Operating Procedures (SOPs)** for verification of AI-generated content and forensic workflows,
- **Judicial training** programs on deepfake technology and admissibility principles under BSA, 2023,
- **Legislative presumptions** allowing courts to infer falsification or require proactive proof of authenticity from the party producing the digital media (Chatterjee, 2023).[37]

6. DOCTRINAL GAPS IN CRIMINAL ATTRIBUTION AND LIABILITY

6.1 Attribution and Mens Rea Challenges

Criminal attribution in the age of AI-generated content like deepfakes presents several doctrinal

[35] Yang, J., Li, Y., & Tran, L. (2023). Deepfake detection with behavioral and physiological signals- A review. *IEEE Transactions on Affective Computing, 14*(3), 506–522. https-//doi.org/10.1109/TAFFC.2022.3182724
[36] See Supra Note 14.
[37] See Supra Note 13.

challenges-

- **Human absence in generation**- Many deepfakes today are generated autonomously by pretrained models with minimal or no human intervention. This complicates the traditional requirement of *actus reus* (guilty act), as there is often no clear "actor" (Pagallo, 2022).[38]

- **Low-intent prompt inputs**- Users may generate harmful content using vague or innocuous prompts. This strains the legal threshold for *mens rea* (guilty mind), raising questions about foreseeability and criminal intent (Balkin, 2021).[39]

- **Resharers or passive disseminators**- Under the *Bharatiya Nyaya Sanhita* (BNS), there is no explicit doctrine of derivative culpability for users who share, comment on, or algorithmically spread harmful deepfakes, despite their substantial role in harm amplification (Chatterjee, 2023).[40]

6.2 Platform Responsibility in Deepfake Dissemination

Legal responsibilities of digital platforms remain underdeveloped in Indian law-

- **Algorithmic amplification**- Platforms use recommendation engines that promote content based on engagement metrics. This undermines their defence of being neutral conduits under Section 79 of the IT Act, as they actively influence content visibility (Gillespie, 2018).[41]

- **Delayed responses**- Especially during sensitive periods like elections or communal tensions, slow platform response to deepfakes leads to real-world harm, as evidenced in global misinformation incidents (Douek, 2021).[42]

- **Lack of proactive obligations**- Unlike the **EU Digital Services Act** or proposed U.S. federal rules, Indian law does not yet mandate proactive deepfake detection, labelling,

[38] Pagallo, U. (2022). Artificial intelligence and criminal law- New challenges for legal theory. *Philosophy & Technology, 35*(1), 1–17. https-//doi.org/10.1007/s13347-021-00474-4
[39] Balkin, J. M. (2021). *The Constitution in the National Surveillance State*. Yale University Press.
[40] See Supra Note 13.
[41] Gillespie, T. (2018). *Custodians of the Internet- Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
[42] Douek, E. (2021). Content moderation in an era of disinformation. *Harvard Law Review, 134*(7), 1575–1615.

or takedown protocols (European Commission, 2023; Congress.gov, 2024).[43][44]

- **No graduated liability**- India lacks a system to differentiate between large-scale platforms with AI infrastructure (e.g., Meta, YouTube) and smaller start-ups. This impedes proportionate regulation and fails to incentivize best practices (Narayan, 2023).[45]

6.3 A Call for Legal Reform in the AI Era

Although Indian courts have laid foundational principles regarding digital rights, the rapid evolution of AI-generated content calls for comprehensive legislative intervention-

- **Technical standards for authenticity**- Statutes must recognize digital watermarking, hash verification and provenance tracking as admissible and preferred methods for authenticating synthetic content (Rajagopal & Mehta, 2024).[46]

- **Tiered platform responsibilities**- Legislation should introduce a differentiated duty framework—light-touch for start-ups and stricter obligations for dominant platforms, especially those with recommender systems or AI integration.

- **Judicial and forensic up-skilling**- Continuous training on AI evidence is essential for both judges and forensic experts, reducing reliance on superficial cues or unreliable methods.

- **Democratic and dignitarian protections**- Laws must balance censorship risk with safeguards for free speech, satire, and dissent, while protecting individuals from targeted reputational or sexualized deepfake abuse (Puttaswamy v. Union of India, 2017).[47]

As synthetic media approaches undetectable realism, India needs a **rights-based, risk-calibrated, and tech-informed legal framework** to preserve the integrity of both its digital

[43] European Commission. (2023). *Digital Services Act*. https-//digital-strategy.ec.europa.eu/en/policies/digital-services-act
[44] See Supra Note 16.
[45] Narayan, A. (2023). Regulating AI- Platform asymmetries and the case for tiered compliance. *National Law School Journal, 15*(1), 67–90.
[46] See Supra Note 12.
[47] See Supra Note 28.

public sphere and criminal justice system.

7. RECOMMENDATIONS

To effectively address the multifaceted challenges posed by deepfake technology within India's legal and criminal justice framework, a coordinated strategy encompassing legislative, doctrinal, and institutional reform is imperative.

7.1 Legislative Reforms

i. Enact Dedicated Deepfake Legislation

Introduce a standalone legal framework specifically addressing deepfakes. This law must-

- Clearly define "deepfake" and distinguish malicious synthetic content from protected expressions such as satire or parody.
- Criminalize the non-consensual creation and dissemination of deepfakes involving intent to deceive, defame, or harm.
- Include gradated penalties based on the context (e.g., political, pornographic, or financial fraud).

ii. Revise Evidence Law for AI-Generated Content- Amend the **Bharatiya Sakshya Adhiniyam, 2023** to-

- Recognize synthetic audiovisual content as a distinct category of digital evidence.
- Mandate AI forensic certifications for admissibility, such as cryptographic watermarks, tamper-evident hash trails, and blockchain-based provenance.
- Empower judges to appoint court-approved forensic experts in contested cases involving AI-generated content.

iii. Strengthen Legal Protections under BNS and DPDP-

- Modify the **Bharatiya Nyaya Sanhita, 2023** to introduce specific provisions for offences involving-
  - Synthetic media forgery,
  - Biometric impersonation,
  - Automated deception.
- Align with the **Digital Personal Data Protection Act, 2023** by-
  - Requiring informed consent for any use of personal data (including images or voice) in AI-generated media.
  - Penalizing the unauthorized scraping, training, or reproduction of personal likeness in synthetic content.

7.2 Platform Regulation and Intermediary Liability

i. **Reform Intermediary Duties under the IT Act-** Amend **Section 79 of the IT Act** to-
   - Mandate deployment of certified deepfake detection systems.
   - Enforce prompt **notice-and-takedown** obligations for synthetic content, especially in cases of reputational harm, impersonation, or communal incitement.
   - Introduce **notice-and-staydown** regimes to prevent the re-upload of identified harmful deepfakes.

ii. **Create a National Deepfake Detection Authority (NDDA)-** Establish an autonomous, government-supported forensic body responsible for-
   - Certifying detection tools and watermarking protocols.
   - Coordinating technical investigations with cybercrime cells.
   - Facilitating inter-agency and international cooperation.

- Conducting public awareness programs and publishing transparency reports.

7.3 Doctrinal and Procedural Enhancements

i. Develop AI-Specific Doctrines in Criminal Law

- Reinterpret traditional mens rea and actus reus principles to accommodate crimes involving AI autonomy or anonymized creation.
- Introduce rebuttable presumptions for AI-generated evidence and permit **burden-shifting mechanisms** where authenticity is in dispute.
- Define gradations of liability across creators, uploaders, amplifiers, and platform intermediaries.

ii. Enhance Evidentiary and Procedural Safeguards

- Integrate cryptographic provenance, expert affidavits, and chain-of-custody protocols into criminal procedure.
- Update digital evidence manuals to include guidelines for examining and contesting deepfake-related exhibits.

iii. Uphold Constitutional Balances

- Ensure all regulatory interventions align with fundamental rights-
  - Protect freedom of speech and expression under **Article 19(1)(a)** through clear exceptions for satire, parody, or artistic expression.
  - Safeguard informational privacy under **Article 21** by requiring consent-based data usage.
- Embed due process guarantees to avoid arbitrary censorship or over-criminalization of digital creativity.

8. Conclusion

Deepfake technology represents a paradigm shift in the creation and manipulation of digital

media. It challenges foundational assumptions of legal systems- that what is seen or heard can be trusted, that culpability can be clearly assigned, and that truth can be verified through tangible evidence. These assumptions no longer hold.

In India, the current legal framework—including the **Information Technology Act, 2000,** the **Bharatiya Nyaya Sanhita, 2023,** and the **Bharatiya Sakshya Adhiniyam, 2023**—remains fragmented and insufficient to meet the evidentiary and doctrinal complexities of AI-generated media. As technological capabilities outpace statutory interpretation, courts and law enforcement are left ill-equipped to authenticate digital evidence, ascertain criminal intent, or regulate intermediary responsibility with precision.

The misuse of deepfakes threatens more than just individual reputations—it jeopardizes **electoral integrity, judicial credibility,** and **public trust**. Without proactive reform, India risks systemic vulnerabilities in its legal architecture.

A robust and future-ready response must be **multidisciplinary and inter-institutional,** combining-

- **Targeted legislation**,
- **Judicial innovation**,
- **Forensic modernization**, and
- **Public engagement**.

By adopting the recommended reforms, India can establish a balanced regime that **preserves digital freedoms, ensures procedural fairness,** and **effectively mitigates synthetic harms.** Only then can the justice system remain resilient in an era where **manipulated realities increasingly masquerade as truth.**