Page: 989-1021 *URL: www.lijdlr.com*

LAWFOYER INTERNATIONAL JOURNAL OF DOCTRINAL LEGAL RESEARCH

ISSN: 2583-7753

LawFoyer International Journal of Doctrinal Legal Research

[ISSN: 2583-7753]

Volume 4 | Issue 2

2026

DOI: https://doi.org/10.70183/lijdlr.2026.v04.192

FROM VICTIM TO VILLAIN: A FORENSIC AND LINGUISTIC INQUIRY INTO THE CONSTRUCTION OF CRIMINAL IDENTITY IN THE AGE OF DEEPFAKES

Neha Goyal[1] Siddharth Sinha[2]

I. ABSTRACT

Deepfake technology has transformed the creation and dissemination of digital content by enabling the generation of highly realistic synthetic audio, video, and images. While these tools have legitimate applications in entertainment, education, and accessibility, their misuse has created significant challenges for criminal law, evidentiary standards, and the protection of personal identity. This paper examines how deepfakes alter the construction of criminal identity by transforming victims into perceived offenders through fabricated but persuasive digital media. Using a doctrinal, comparative, and case study methodology, the paper analyses the Indian legal framework, including the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023, the Bharatiya Sakshya Adhiniyam, 2023, the Protection of Children from Sexual Offences Act, 2012, and the Copyright Act, 1957. It further compares regulatory developments in the United Kingdom, the United States, and the European Union, with particular reference to the Online Safety Act 2023, the proposed NO FAKES Act, the DEEPFAKES Accountability Act, and the EU Artificial Intelligence Act. Drawing on case studies involving financial fraud, intimate image abuse, family court evidence, political manipulation, and reputational attacks, the study highlights the role of digital forensics and forensic linguistics in authenticating disputed content and exposing synthetic identities. The findings reveal significant gaps in Indian law, including the absence of deepfake-specific offences, limited platform accountability, and inadequate evidentiary protocols. The paper recommends targeted legislative amendments, specialised forensic-linguistic laboratories, AI-

[1] Assistant Prof. at Indore Institute of Law (India). Email: nehagoyal.8510@gmail.com
[2] Assistant Prof. at Indore Institute of Law (India). Email: sid84242@gmail.com

 (ISSN: 2583-7753)

based verification systems, judicial training, and public awareness initiatives to strengthen victim protection, preserve evidentiary integrity, and restore trust in digital media.

II. KEYWORDS

Deepfakes, Forensic Linguistics, Criminal Identity, Cybercrime, AI Manipulation.

III. INTRODUCTION

Deepfake technology is a type of artificial intelligence (AI) that produces incredibly realistic but fake audio, video, or images using deep learning methods, especially generative adversarial networks (GANs).[3]Combining the words "deep learning" and "fake" results in the term "deepfake." It makes it possible to digitally alter a person's voice, face, or behavior to create information that looks real but is completely fake or changed.[4]

Deepfakes, created for creative and entertainment purposes, are now a concern due to their potential for misinformation, identity theft, cybercrime, political manipulation, and nonconsensual explicit content. Detection involves visual artifacts, voice patterns, and metadata examination. While they are legitimately used in film production, education, and accessibility tools, misuse has raised several concerns regarding privacy, permission, and the need for regulation and detection systems.[5]

Deepfakes, which first appeared in 2017[6], quickly progressed from crude, easily detectable manipulations to incredibly convincing synthetic media capable of mimicking facial emotions, voices, and movements with amazing accuracy.

Several causes have fueled this expansion. The explosion of open-source tools and mobile applications has made deep-fake production accessible to non-experts, while

[3] Ian Goodfellow et al., *Generative Adversarial Nets*, 27 Advances in Neural Info. Processing Sys. 2672, 2674–76 (2014) (introducing GANs); Hany Farid, *Creating, Using, Misusing, and Detecting Deep Fakes*, 2 J. Online Tr. & Sec. 1, 2–4 (2020).

[4] Greene et al., Deepfake, EBSCO, (2024), https://www.ebsco.com/research-starters/computer-science/deepfake

[5] Sami Alanazi et al., Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses, Human-Intelligent Systems Integration, (February 20, 2025), https://link.springer.com/article/10.1007/s42454-025-00060-4

[6] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, Vice (Dec. 11, 2017), https://www.vice.com/en/article/gy8b7m/ai-fake-porn-on-reddit; Hao Li et al., *Emerging Threats: Deepfake Technology*, 58 Comm. ACM 36, 37–39 (2019).

massive amounts of online data from social media give ample training material for AI models.[7]

Commercial industries, including film, advertising, and gaming, have adopted deep-fake-like technologies for dubbing, visual effects, and tailored content. However, the same skills have been used to spread misinformation, manipulate politics, steal identities, and publish non-consensual content, posing serious ethical and security problems.[8]

Linguistic analysis investigates how language use (word choice, syntax, accent, discourse style) shapes and reflects a person's social, cultural, and individual identity. Experts can uncover variance between a person's claimed identity and their true linguistic patterns using linguistic profiling and authorship attribution. This is especially important in online interactions, legal conflicts, and cases involving impersonation or manipulation, including deepfakes.[9]

Forensic analysis supplements this by using scientific methods to verify or counter identity claims. Digital forensics explores metadata, device logs, and file sources; audio-visual forensics studies voiceprints, facial features, and traces of media manipulation; and biometric comparison assists in matching physical or vocal traits to recognized individuals.

The rise of deepfake technology has significant societal implications across multiple fields, including the spread of misinformation, complex legal and ethical concerns, and a growing erosion of trust in journalism and digital media.

1. **Spread Misinformation:** Deep fakes struck public confidence in the media, institutions, and public personalities by aiding in the dissemination of false

[7] Ángel Fernández Gambín et al., Deepfakes: Current and Future Trends, Artificial Intelligence Review, (February 19, 2024), https://link.springer.com/article/10.1007/s10462-023-10679-x
[8] Ishan Gupta, *Pros and Cons of Deepfake Technology in Digital Marketing*, IMark Infotech (July 12, 2023), https://www.imarkinfotech.com/pros-and-cons-of-deepfake-technology-in-digital-marketing/
[9] Dana Roemling & Jack Grieve, Forensic Authorship Analysis, Crest Research (Jan 14,2021), https://crestresearch.ac.uk/comment/forensic-authorship-analysis/

information. The functioning of society overall, public debate, and democratic procedures are all impacted severely.[10]

2. **Legal and Ethical Concerns:** Deepfakes bring up difficult moral and legal issues. Holding anyone responsible for the production and distribution of Deepfakes is difficult since current rules frequently can't keep up with the quickly changing technology. Legislators and legal regimes constantly struggle to strike a balance between the right to free speech and the need to safeguard people's rights and prevent harm.[11]
3. **Impact on Media:** The media landscape and journalism are affected by deepfakes. The public finds it more difficult to distinguish between trustworthy information and altered content as a result of the widespread occurrence of deepfakes, which also damages the credibility of news sources and fosters the dissemination of false information.[12]

A. Research Objectives

This study seeks to:

1. Examine how deepfake technology alters the construction of criminal identity by transforming victims into perceived offenders.
2. Analyse the role of digital forensics and forensic linguistics in detecting manipulated audio, video, and textual content.
3. Evaluate the adequacy of the current Indian legal framework, including the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023, the Bharatiya Sakshya Adhiniyam, 2023, the Protection of Children from Sexual Offences Act, 2012, and the Copyright Act, 1957, in addressing deepfake-related harms.

[10] Sami Alanazi et al., Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses, Human-Intelligent Systems Integration, (February 20, 2025), https://link.springer.com/article/10.1007/s42454-025-00060-4
[11] The New Face of Misinformation: Deepfakes and their Legal Challenges, PACE University, (October 17, 2024), https://socialmediablawg.blogs.pace.edu/2024/10/17/the-new-face-of-misinformation-deepfakes-and-their-legal-challenges/
[12]Ingrid de Chevigny, The Impact of Deepfakes on Information Credibility, Media Connect, (September 26, 2023), https://mediaconnect.com/the-impact-of-deepfakes-on-information-credibility

4. Compare India's regulatory approach with developments in the United Kingdom, the United States, and the European Union.
5. Propose legal, forensic, and institutional reforms to strengthen victim protection, evidentiary reliability, and accountability in deepfake-related offences.

B. Research Questions

This paper addresses the following research questions:

1. How does deepfake technology influence the construction of criminal identity by recasting victims as apparent perpetrators?
2. What evidentiary value does digital forensic and forensic linguistic techniques provide in detecting and authenticating deepfake content?
3. To what extent does the existing Indian legal framework adequately regulate the creation and misuse of harmful deepfakes?
4. What lessons can India draw from the legal responses adopted in the United Kingdom, the United States, and the European Union?
5. What legislative, institutional, and technological reforms are necessary to safeguard individual rights and preserve trust in digital evidence?

C. Research Methodology

This study adopts a doctrinal and analytical research methodology. Primary legal sources, including statutes, delegated legislation, judicial decisions, and official regulatory materials from India, the United Kingdom, the United States, and the European Union, are examined to assess the legal treatment of deepfake-related harms. Secondary sources, such as scholarly articles, policy reports, and technical literature on artificial intelligence, digital forensics, and forensic linguistics, are used to contextualise the technological and evidentiary issues. The study also employs a comparative approach to identify best practices in foreign jurisdictions and a case study method to analyse selected incidents involving fraud, intimate image abuse, family litigation, political manipulation, and reputational harm. Through this

combined methodology, the paper evaluates the intersection of technology, law, and identity and develops recommendations for strengthening India's regulatory and evidentiary response to deepfakes.

IV. LEGAL FRAMEWORK

A. Applicable Legal Framework in India

In the Indian context, deepfake-related offences are currently addressed through different provisions spread across multiple statutes rather than through a single, unified law. "Information Technology Act, 2000" (IT Act)[13] remains the primary legislation for cyber offences. Sections 66C and 66D deal with identity theft and cheating by personation through the use of computer resources. These provisions are particularly relevant when deepfakes involve the impersonation of an individual's likeness, voice, or identity to obtain a benefit or cause harm. Section 66E criminalises the violation of privacy by capturing, publishing, or transmitting images of private areas without consent this can extend to AI-generated images if they depict private body parts. Sections 67, 67A, and 67B address the creation, publication, and transmission of obscene or sexually explicit material, with Section 67B specifically targeting material involving children, whether real or computer-generated.

With the introduction of the "Bharatiya Nyaya Sanhita, 2023" (BNS)[14], replacing the "Indian Penal Code" from July 2024, there is an updated framework for general criminal provisions. Defamation under Section 356 of the BNS can be invoked when a deepfake damages the reputation of a person, whether the harm is inflicted through an apparently "humorous" alteration or malicious fabrication. Provisions on forgery of electronic records and cheating by personation also apply, particularly where a deepfake is used to fabricate evidence or mislead law enforcement. "Protection of Children from Sexual Offences Act, 2012" (POCSO)[15] reinforces these protections,

[13] Information Technology Act, No. 21 of 2000, §§ 66C–66D, Acts of Parliament, 2000 (India).
[14] Bharatiya Nyaya Sanhita, No. 45 of 2023, § 356, Acts of Parliament, 2023 (India).
[15] Protection of Children from Sexual Offences Act, No. 32 of 2012, §§ 13–15, Acts of Parliament, 2012 (India).

making it an offence under Sections 13 to 15 to create, possess, or distribute pornographic content involving minors, even if the material is purely synthetic.

"The Copyright Act, 1957"[16], provides a further layer of protection in cases where a deepfake involves copying a substantial part of a copyright-protected work, such as a photograph, video, or voice recording. Section 57's protection of "moral rights" is especially important in cases where a work is distorted or mutilated in a way that prejudices the honour or reputation of the author, as is often the case with deepfakes that insert an individual's likeness into false contexts. Additionally, the "IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021"[17], and subsequent MeitY advisories require platforms to remove unlawful content, including deepfakes, upon receiving notice.

From an evidentiary perspective, the "Bharatiya Sakshya Adhiniyam, 2023," which replaced the Indian Evidence Act, modernises the treatment of electronic records. Section 61 ensures that electronic evidence cannot be rejected solely for being digital in nature, while Section 63 introduces updated certification requirements for computer outputs, and Sections 86 to 87 create presumptions in favour of secure electronic records. These are vital for ensuring the admissibility of forensic evidence in deep-fake related cases.

B. Gaps in the Indian Legal Framework

Despite these provisions, significant gaps remain. There is no dedicated offence for the creation or possession of harmful deepfakes outside the domains of sexual content, child sexual abuse material, fraud, or defamation. This means that non-sexual but still harmful uses—such as political disinformation, fabricated emergency announcements, or reputational sabotage without clear defamatory imputations are not directly addressed.[18] Consent-based protections are also inconsistent: while transmitting or publishing a non-consensual sexual deepfake is criminalized, mere

[16] Copyright Act, No. 14 of 1957, § 57, Acts of Parliament, 1957 (India).
[17] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Gazette of India, Feb. 25, 2021 (India).
[18] Arghya Sengupta & Shreya Ganguly, *Regulating Deepfakes in India: Challenges and Opportunities*, Vidhi Centre for Legal Policy (Mar. 15, 2023), https://vidhilegalpolicy.in/blog/regulating-deepfakes-in-india

creation of such material without distribution is not explicitly punishable in most contexts. India also lacks a statutory "right of publicity" or likeness right, leaving victims to patch together remedies from privacy, passing-off, copyright, and moral rights law.[19]

Platform accountability is limited to takedown obligations and advisories, without a mandatory provenance or watermarking regime for synthetic media. This complicates attribution and makes it easier for perpetrators to disclaim authorship. Finally, the cross-border nature of many deepfake incidents where servers, AI models, and perpetrators are often outside India creates serious enforcement delays due to reliance on Mutual Legal Assistance Treaties (MLATs) and slow international cooperation.[20]

V. CASE STUDIES

A. The Arup "Deepfake CFO" $25 Million Scam (2024, Hong Kong)

1. **Facts:** In early 2024, a finance staffer at Arup (a British engineering firm) received a message purporting confidential instructions from the company's UK-based CFO. The employee's suspicions at first led him to believe the communication was a phishing attempt, but after joining a multi-participant video call, he saw what appeared to be his CFO and several other colleagues in real time. All the "attendees" were, in fact, deepfakes generated using advanced AI technology. Convinced of the call's authenticity, the employee executed 15 transfers totaling almost $26 million to accounts in Hong Kong.[21]

2. **Forensic & Linguistic Evidence:** Specialists analyzed the video feed and detected subtle inconsistencies in facial dynamics and audio-visual sync, alongside technical metadata mismatches. Financial forensics tracked the

[19] N.S. Nappinai, *Technology Laws Decoded* 297–302 (2d ed. 2022); Apar Gupta, *India's Missing Right of Publicity*, Internet Freedom Foundation (July 20, 2021), https://internetfreedom.in.
[20] Brookings Institution, *Deepfakes and Synthetic Media: Policy Implications* (Sept. 2020), https://www.brookings.edu; U.S. Dep't of Justice, *Mutual Legal Assistance Treaties (MLATs)*, https://www.justice.gov/criminal-oia/international (last visited Aug. 21, 2025).
[21] Alexandra Bacon, *Engineering Giant Arup's CFO Deepfake Scam Costs $25 Million*, Bus. Insider (May 17, 2024), https://www.businessinsider.com/engineering-giant-arup-target-of-25-million-deepfake-scam-2024-5

destination of transferred funds, correlating multiple instances globally of similar fraud patterns with deepfake signatures.

3. **Identity Construction:** The "villain" was constructed as both a trusted corporate superior and a faceless, foreign threat; the employee (nominal victim) faced internal scrutiny for his role. Linguistic cues via emails and instant messages were scrutinized: formal language and corporate jargon mimicked the real CFO, showing how attackers constructed trust through linguistic proximity.

4. **Legal Response:** Hong Kong law enforcement arrested several related individuals, uncovering identity card theft and a network of fraudulent applications using AI-augmented impersonations.[22]The case led to increased global alerts within the banking sector about "synthetic identity fraud" and consideration for stricter authentication protocols.

B. Kerala's First Deepfake Fraud (2023, India)

1. **Facts**: A 73-year-old man in Kerala, India (Radhakrishnan), received a WhatsApp video call supposedly from his friend and former colleague.[23] The caller appearing and sounding identical to the colleague requested an urgent loan of ₹40,000. Deepfake technology was used to mimic both voice and facial features, leveraging social media information. The man, trusting the caller, transferred the money, only to discover he had been scammed.[24]

2. **Forensic & Linguistic Evidence**: Digital forensics traced the monetary transfer to an external bank account. Linguistic analysis of the chat logs showed cultural and idiomatic familiarity designed to maximize trust. Police unmasked the scam by cross-referencing IP addresses, social media scraping, and calling log metadata.

[22] Grace Noto, *Scammers Siphon $25 M from Engineering Firm Arup via AI Deepfake 'CFO'*, CFO Dive (May 17, 2024), https://www.cfodive.com/news/scammers-siphon-25m-engineering-firm-arup-deepfake-cfo-ai/716501

[23] *Kerala's First Deepfake Fraud Case*, Kozhikode Cyber Police Station Crime No. 101/2023 (India).

[24] *Kerala man loses ₹40k to AI-enabled deep-fake fraud*, *Hindustan Times* (July 18, 2023); *"Sister in hospital, please help": Ex-Coal India exec loses Rs 40,000 in Kerala's first deepfake case*, *Indian Express* (Nov. 10, 2023); *Accused in deepfake fraud remanded in judicial custody*, *The Hindu* (Nov. 9, 2023).

3. **Identity Construction**: Here, the victim is literally recast as the potential enabler of a crime (e.g., if the transaction has been used for illegal purposes). The deepfake leveraged the inherent trust of personal relationships, turning the familiar into the fraudulent.
4. **Legal Response:** The Kerala police issued new warnings, updating cybercrime advisories, and investigated similar scam attempts. The victim's claim was substantiated, and efforts are ongoing to apprehend cross-state and possibly international fraudsters.

C. Deepfake Audio in a UK Family Court (2019, United Kingdom)

1. **Facts:** During a bitter child custody dispute in a UK family court, a mother presented an audio recording purportedly of the father making direct violent threats. The court initially accepted the evidence, leading to a suspension of the father's visitation rights.[25] Forensic audio experts, engaged by the father's legal team, identified doctored segments inserting words and phrases via AI deepfake voice synthesis.
2. **Forensic & Linguistic Evidence:** Experts performed waveform and spectrogram analysis to reveal irregular transitions and anomalous prosody (speaking rhythm/pitch). Linguistic stylometry identified segments of the recording inconsistent with the father's speech patterns—awkward syntax and non-native vocabulary usage were red flags. Metadata showed the audio file had been edited after the original call.[26]
3. **Identity Construction:** The father was quickly constructed as a "dangerous villain" in legal and familial discourse. Forensic reversal reframed him as a

[25] Kathryn Snowdon, *Deepfake Audio Used in Custody Battle, Lawyer Reveals Doctored Evidence*, *Telegraph* (Jan. 31, 2020), https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/.
[26] *Family Law Evidence and the Problem with Deep Fake Images*, Holland Family Law (2023), https://hollandfamilylaw.co.uk/family-law-evidence-and-the-problem-with-deep-fake-images/.

victim of malicious fabrication, illustrating how deepfake evidence can both destroy and restore legal reputations.[27]

4. **Legal Response:** The court accepted the forensic analysis, dismissing the deep-fake evidence and reinstating the father's visitation rights. This case set a precedent for greater scrutiny of audio/visual "evidence" in family disputes, highlighting emergent standards for AI-derived content.

D. Taylor Swift Explicit Deepfake Scandal (2024, USA and International)

1. **Facts:** In early 2024, AI-generated explicit images depicting singer Taylor Swift circulated widely on X (formerly Twitter), Reddit, and similar platforms. The images, entirely fake, were designed to look realistic and were created without Swift's consent. They rapidly went viral, triggering significant reputational and psychological harm and calls for urgent social media regulation.[28]

2. **Forensic and Linguistic Evidence:** Digital forensics identified telltale signs of deep learning artifact "blending," with minor anatomical inconsistencies and metadata mismatches. The images were reverse searched across the internet, mapping their rapid proliferation and the channels of distribution. Linguistic analysis of accompanying text posts revealed coordinated bot-like messaging patterns and hate-driven rhetoric intended to increase reach and impact.[29]

3. **Identity Construction:** Swift, the victim, was portrayed as the architect of scandal—a stark example of how deepfake technology can weaponize a public personality's image, turning audience perception against the victim, even temporarily.

---

[27] *From Photoshop to Deepfakes: The Evolution of Fake Evidence Creation in Family Law Cases*, Penningtons Manches Cooper (2024), https://www.penningtonslaw.com/news-publications/latest-news/2024/from-photoshop-to-deepfakes-the-evolution-of-fake-evidence-creation-in-family-law-cases

[28] Taylor Swift Explicit Deepfake Scandal Reported, BBC News (Jan. 25, 2024), https://www.bbc.com/news/technology-deepfake-scandal.

[29] Pranav Dixit, Taylor Swift Deepfake Images Spark Outrage, Highlighting AI's Dangers, Wired (Jan. 26, 2024), https://www.wired.com/story/taylor-swift-deepfake-ai-dangers.

4. **Legal Response:** Swift's legal team issued public statements, served takedown notices, and urged government intervention. U.S. lawmakers responded by fast-tracking legislation targeting non-consensual intimate AI content, while platforms faced renewed scrutiny for slow takedowns. Social platforms eventually acted, but gaps in rapid response remained evident.[30]

E. Deepfake Political Manipulation in Indian Elections (2020, India)

1. **Facts:** During a critical legislative assembly election, a video of a prominent Indian MP was circulated widely on messaging and social platforms. In the original, the MP spoke English and encouraged voting for his party. Deepfake technology was used by political consultants at The Ideaz Factory to algorithmically reanimate his face and generate two synthetic videos, one in Haryanvi and one in English, mimicking both facial movements and speech to make it appear as though the MP was personally addressing targeted voter groups ahead of the Delhi Legislative Assembly elections.
2. **Forensic and Linguistic Evidence:** Analysis revealed mismatches between mouth movements and audio, especially in the less-common language versions. Linguists analyzed the translated content, noting awkward syntax and non-native idioms. Social media metadata traced the spread to accounts coordinated by a political communication firm.
3. **Identity Construction:** The deepfakes blurred the lines between authenticity and manipulation, presenting the MP as multilingual and more in touch with diverse voter bases. Here, the "villain" narrative was inverted—the technological manipulation cast political teams as digital deceivers, exploiting voters' trust in the MP's supposed direct engagement.

[30] Statement from Taylor Swift's Representative on Deepfake Images, Rolling Stone (Jan. 26, 2024), https://www.rollingstone.com/music/music-news/taylor-swift-deepfake-statement; Cat Zakrzewski & Taylor Lorenz, Lawmakers Fast-Track Bill to Criminalize Non-Consensual Deepfake Porn, Wash. Post (Jan. 30, 2024), https://www.washingtonpost.com/technology/2024/01/30/deepfake-law-bill-taylor-swift; Kalhan Rosenblatt, X and Reddit Struggle to Tackle Taylor Swift Deepfake Scandal, NBC News (Jan. 27, 2024), https://www.nbcnews.com/tech/social-media/x-reddit-taylor-swift-deepfakes

4. **Legal Response:** While no criminal case was filed due to the non-defamatory, campaign-oriented context, the case raised concerns about fair campaigning and led to public statements from election authorities regarding deepfake regulation. It spotlighted the legal grey zone surrounding deepfakes in political advertising and incited debate about the need for clear legislation on synthetic media in elections.[31]

F. Analytical Insights

1. **Forensics**: Each case relied on advanced audio-visual forensics, digital trace analysis, and stylometric linguistic profiling, but faced limitations as deepfake quality improved.[32]
2. **Villain/Victim Narrative:** Deepfakes consistently exploited public trust—casting both global celebrities and democratic processes as "villains" or "frauds" in the eyes of the unwary, until expert intervention reversed the narrative.[33]
3. **Legal Response:** Existing laws are often slow to adapt, but new orders for swift takedowns, targeted criminal codes, and cross-agency investigations are increasingly common. Civil suits and enacted regulatory measures, including the U.S. TAKE IT DOWN Act, are being used to deter malicious creators and compel platforms to remove non-consensual intimate deepfakes within 48 hours of receiving a valid complaint.[34]
4. **Forensics, Linguistics, and the Modern Villain:** Digital Forensics now regularly incorporates deep learning tools, artifact comparison, and

[31] Legal and Regulatory Gaps in Deepfake Oversight During Indian Elections (explaining reliance on IT Act 2000, IPC, and Model Code), *Political Marketer* (2024), https://politicalmarketer.com/deepfakes-and-misinformation-in-indian-democracy.
[32]Hany Farid, Creating, Using, Misusing, and Detecting Deep Fakes, 2 J. Online Tr. & Sec. 1, 4–7 (2020); Siwei Lyu, Deepfake Detection: Current Challenges and Next Steps, 2 Frontiers in Computer. Sci. 1, 2–5 (2020).
[33]Robert Chesney & Danielle Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1789–91 (2019); Britt Paris & Joan Donovan, Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence, Data & Soc'y Research Inst. 4–6 (2019), https://datasociety.net/library/deepfakes-and-cheap-fakes.
[34] Danielle Keats Citron, Sexual Privacy, 128 Yale L.J. 1870, 1923–25 (2019); TAKE IT DOWN Act, S. 146, 119th Cong. (2025) (enacted May 19, 2025).

provenance analysis. Real-time cross-comparison with known authentic data (video, audio, metadata, written style) is crucial.[35] Linguistic Evidence syntactic anomalies, semantic mismatches, and non-idiomatic speech quickly reveal fakes, yet the more a deepfake "learns" from real samples, the subtler these cues become.[36]

5. **AI Detection Tools and Limitations:** While deepfake detectors flag inconsistencies in audio-visual streams and biometric cues, new "zero-day" fakes often evade detection. Detectors are less robust across languages, accents, and lighting conditions and adversarial attacks constantly adapt to bypass security. Tools excel at triaging, but courts and investigators must supplement detection with traditional forensic and linguistic expertise.

G. Legal and Social Consequences

Victims are regularly "villainized," even facing legal sanction or lasting reputational damage until forensic experts intervene.[37] Courts worldwide are now increasingly accepting digital and linguistic forensic testimony, but legal frameworks are only beginning to adapt, with injunctions, John Doe orders, and evidence admissibility rules slowly evolving.[38]

[35] Hany Farid, Digital Forensics in a Post-Truth Age, 36 Forensic Sci. Int'l: Dig. Investigation 301, 303–05 (2021); Matthew Stamm et al., Information Forensics: An Overview of the First Decade, 105 IEEE Trans. on Info. Forensics & Sec. 1, 7–10 (2020).

[36] Gerald Friedland & Robin Sommer, Cybercasting the Joint: On the Privacy Implications of Geo-Tagging, 5 Int'l J. Multimedia Intelligence & Sec. 111, 118–20 (2019) (noting forensic linguistic vulnerabilities in synthetic media); Chengqing Zong et al., Linguistic Characteristics of Machine-Generated Text, 2 J. Computer. Linguistics & Lang. Tech. 45, 52–54 (2020).

[37] Danielle Keats Citron, Deep Fakes: The Looming Crisis for National Security, Privacy, and Democracy, 107 Calif. L. Rev. 1753, 1792–95 (2019)

[38] Aparna Chandra, Anuj Bhuwania & Sital Kalantry, Law Courts and Deepfakes: Emerging Evidentiary Challenges, 14 Indian J.L. & Tech. 1, 23–28 (2021) (on admissibility of deepfake evidence in India); Lorraine Hope & Aldert Vrij, Expert Testimony and Emerging Technologies: Forensic Evidence in Court, 28 Psych. Pub. Pol'y & L. 105, 110–13 (2022); Shreya Singhal v. Union of India, (2015) 5 S.C.C. 1 (India) (illustrating courts' use of injunctions/John Doe orders in digital speech cases).

The construction of criminal identity in the deepfake era is thus a dynamic negotiation between technology, expert witnesses, and the media where the line between villain and victim is more fragile than ever.[39]

VI. COMPARATIVE PERSPECTIVE

A. UK

The United Kingdom has moved more decisively toward deepfake-specific criminalisation, particularly in the context of intimate image abuse. Amendments linked to the Online Safety Act 2023 updated the Sexual Offences Act 2003, making the sharing of non-consensual intimate images, including deepfakes, a criminal offence. These changes came into force on 31 January 2024.[40] The UK government has also announced plans to go further by criminalising the very creation of sexually explicit deepfakes, even where they are not distributed. These measures reflect recognition that harm often arises at the moment of creation, particularly in coercive or blackmail situations.[41]

In addition, the UK framework includes strong regulatory obligations on online platforms. Under the Online Safety Act, Ofcom has the power to enforce duties on platforms to detect and swiftly remove priority illegal content, including intimate image abuse. The law emphasises victim protection through confidentiality and rapid takedown procedures, creating a more streamlined path to relief compared to India's current reliance on general cybercrime and obscenity provisions.[42]

B. USA

[39] Britt Paris & Joan Donovan, Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence, Data & Soc'y Research Inst. 10–12 (2019), https://datasociety.net/library/deepfakes-and-cheap-fakes; Robert Chesney & Danielle Citron, Deep Fakes and the New Disinformation War, 2018 U. Md. Legal Stud. Rsch. Paper No. 2018-21, at 18–22.

[40] Sexual Offences Act 2003, c. 42, § 66A (as amended 2023) (U.K.); Online Safety Act 2023, c. 49, § 187 (U.K.); Commencement Regulations, The Sexual Offences Act 2003 (Commencement No. 22) Regulations 2023, SI 2023/1293 (U.K.) (in force Jan. 31, 2024).

[41] HM Gov't, Dep't for Sci., Innovation & Tech., *Press Release: Government to Strengthen the Law Against Deepfake Intimate Images* (Nov. 2023), https://www.gov.uk/government/news/government-to-strengthen-the-law-against-deepfake-intimate-images

[42] Online Safety Act 2023, c. 49, §§ 65–70 (U.K.) (platform duties and Ofcom enforcement powers); Ofcom, *Protecting People from Harmful Online Content Under the Online Safety Act 2023: Guidance on Illegal Content Duties* (Apr. 2024), https://www.ofcom.org.uk

The United States lacks a comprehensive federal deepfake law, but legislative activity is increasing. The proposed DEEPFAKES Accountability Act would require watermarking and disclosure for synthetic media and create civil remedies for violations.[43] Another pending proposal, the NO FAKES Act, seeks to establish a federal right to protect an individual's voice and likeness from unauthorized AI-generated replicas—this addresses a significant gap in Indian law.[44]

At the state level, laws are emerging in a piecemeal fashion. California allows civil actions for sexual deepfakes and has enacted time-limited bans on election-related deepfakes[45]. Texas criminalises deceptive deepfakes intended to influence elections[46], and Virginia has extended its "revenge porn" laws to cover deepfakes[47]. Many other states have introduced similar provisions, particularly targeting sexually explicit deepfakes and political disinformation.

The Deepfake Accountability Act (proposed in the US) would require clear labeling of synthetic media, penalties for misuse, and mechanisms for victims to seek redress.

The UK is moving towards stricter laws on intimate image abuse, proposed to expand to harmful deepfakes.

C. Forensic and Linguistic Implications

From a forensic standpoint, Indian law's current patchwork approach requires careful mapping of the alleged conduct to an existing provision, often combining IT Act offences, BNS provisions, and special laws like POCSO or the Copyright Act[48]. The evidentiary regime under the Bharatiya Sakshya Adhiniyam is capable of supporting complex AI forensics, provided investigators capture and preserve originals, generate hash values, and obtain Section 63(4) certificates[49]. Linguistic analysis can play a role

[43] DEEPFAKES Accountability Act, H.R. 3230, 116th Cong. (2019).
[44] Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2023 ("NO FAKES Act"), Discussion Draft, 118th Cong. (2023).
[45]Cal. Civ. Code § 1708.86 (West 2023) (civil actions for sexually explicit deepfakes); Cal. Elec. Code § 20010 (West 2023) (temporary ban on election-related deepfakes).
[46]Tex. Elec. Code Ann. § 255.004 (West 2023).
[47]Va. Code Ann. § 18.2-386.2 (West 2023).
[48]Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67–67B, INDIA CODE (2000); Bharatiya Nyaya Sanhita, No. 45 of 2023, INDIA CODE (2023); Protection of Children from Sexual Offences Act, No. 32 of 2012, INDIA CODE (2012); Copyright Act, No. 14 of 1957, INDIA CODE (1957).
[49]Bharatiya Sakshya Adhiniyam, No. 46 of 2023, § 63(4), INDIA CODE (2023).

in detecting synthetic voices or speech patterns inconsistent with a purported speaker, which may be crucial in personation cases under Sections 66C and 66D of the IT Act[50].

The "victim to villain" dynamic emerges when manipulated content is misinterpreted or maliciously framed as genuine, leading the victim to face investigation or social sanction[51]. This can be compounded by the difficulty of proving a deepfake's artificiality in the eyes of non-technical decision-makers, making forensic-linguistic clarity essential.[52]

D. Policy Reform Directions

A deepfake-specific offence, at least for non-consensual intimate material and election interference, would fill a core gap in Indian law[53]. This could be accompanied by a statutory right of publicity to protect likeness and voice, drawing inspiration from the NO FAKES Act in the US[54]. Watermarking or provenance requirements, similar to those in the DEEPFAKES Accountability Act proposal[55], could be paired with safe-harbour provisions for platforms that act swiftly upon notice. Procedurally, India could adopt fast-track takedown and evidence preservation measures akin to the UK's Online Safety Act obligations, ensuring that victims receive timely relief[56]. These reforms would not only address current enforcement challenges but also align India's response with emerging global standards for deepfake regulation.[57]

---

[50]Information Technology Act, No. 21 of 2000, §§ 66C–66D, INDIA CODE (2000).
[51]Aparna Chandra, Anuj Bhuwania & Sital Kalantry, *Law Courts and Deepfakes: Emerging Evidentiary Challenges*, 14 Indian J.L. & Tech. 1, 20–23 (2021) (discussing misinterpretation of manipulated content).
[52]Raghav Aggarwal, *Deepfakes and the Indian Legal System: Evidentiary and Investigative Challenges*, 4 NALSAR Student L. Rev. 45, 58–60 (2022).
[53]Cf. Sexual Offences Act 2003, c. 42, § 66A (as amended 2023) (U.K.); Tex. Elec. Code Ann. § 255.004 (West 2023)
[54]Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2023 ("NO FAKES Act"), Discussion Draft, 118th Cong. (2023).
[55]DEEPFAKES Accountability Act, H.R. 3230, 116th Cong. (2019).
[56]Online Safety Act 2023, c. 49, §§ 65–70 (U.K.); Ofcom, *Protecting People from Harmful Online Content Under the Online Safety Act 2023: Guidance on Illegal Content Duties* (Apr. 2024), https://www.ofcom.org.uk
[57]Nat'l Conf. of State Legislatures (NCSL), *State Legislation Addressing Deepfakes* (updated Apr. 17, 2024), https://www.ncsl.org/technology-and-communication/state-legislation-addressing-deepfakes; HM Gov't, Dep't for Sci., Innovation & Tech., *Press Release: Government to Strengthen the Law Against Deepfake Intimate Images* (Nov. 2023), https://www.gov.uk/government/news/government-to-strengthen-the-law-against-deepfake-intimate-images

VII. FINDINGS

1. **Forensic limitations in identifying manipulated identities:** *S*everal obstacles make it difficult to forensically identify modified identities in the era of deep fakes. Advanced AI generated content can easily change voices, face traits, and language patterns, and frequently evade traditional identification techniques. Verification is further hampered by compromised chain-of-custody, metadata loss, and the simplicity of re-encoding files. Furthermore, artistic imitation might alter language attribution, resulting in misidentification. These limitations weaken confidence in forensic and legal procedures by making it more difficult to accurately attribute modified content and increasing the possibility of false allegations or acquittals.[58]
2. **How Linguistic pattern acts as warnings:** When it comes to identifying modified identities in deepfake content, language patterns can be vital warning signs. Minor discrepancies in syntax, grammar & vocabulary selection could be signs of artificial text production or authorship incompatibilities. Changes in register, tone, or cultural reference may indicate attempts to incorrectly impersonate a victim's writing or speech patterns. Chronological discrepancies, such as the usage of language that predates the claimed tape, are also indicators of fabrication. Forensic linguists use these indications, along with metadata and contextual information, to detect manipulation which uncovers false narratives, and disputes the authenticity of criminal attributions.[59]
3. **Deepfake content often lacks verifiable metadata, but leaves forensic traces:** *R*e-encoding, compression, or intentional manipulation removes creation dates, device identifiers, and geolocation data, which are needed for identification. However, it frequently leaves trace forensic evidence that aids in detection. Inconsistent lighting, shadows, and reflections; unusual facial

[58] Denis de Montigny, The Evolving Landscape of Deepfake Detention: Current Challenges and Strategic Imperiatives, Ongota, (June 12, 2025), https://www.ongota.com/evolving-landscape-deepfake-detection-current-challenges-strategic-imperiatives/
[59] Alicja Martinek & Ewelina Bartuzi Trokielewicz, Detecting deepfakes and false ads through analysis of text and social engineering techniques, ACL Anthology, (January,2025), https://aclanthology.org/2025.coling-main.564/

microexpressions or eye-blinking patterns; incorrect lip-sync with audio; and pixel-level noise or compression issues. Linguistic or vocal irregularity, such as unnatural pauses or inconsistent accent patterns, may suggest a fake origin. While these traces are less certain than original metadata, competent forensic analysis can combine them to question the legitimacy of deepfakes.[60]

4. **Discrepancy between public perception and actual culpability:** The realism of deepfakes and their quick online propagation aggravate the disparity between public perception and actual blame in From Victim to Villain. Manipulated photos, videos, or audio can quickly frame an innocent individual, resulting in a presumption of guilt even before investigations are completed. Social media extension, twisted reporting, and confirmation bias all contribute to public judgment, which frequently ignores forensic doubts and due procedure.When a misleading narrative spreads, retractions or corrections seldom reach the same audience or have the same impact. This gap between perception and reality not only permanently ruins reputations but also causes legal institutions to act on incorrect or insufficient evidence.[61]
5. **Rapid spread Causes Irreversible Harm:** Deepfake information travels quickly through social media and messaging channels, frequently reaching millions before its validity is confirmed. Individuals who are unjustly depicted as criminals or wrongdoers suffer significant reputational damage as a result of this rapid transmission. Victims may suffer job loss, online harassment, and social isolation, and this reputational harm often occurs long before corrections are made public. To mitigate dangers, experts recommend quick reaction measures, verifiable rebuttals, tighter platform regulations, and public

[60]Achhardeep Kaur et al., Deepfake video detection: challenges and opportunities, Artificial Intelligence Review, (May 29,2024), https://link.springer.com/article/10.1007/s10462-024-10810-6
[61] Saifuddin Ahmed & Hui Wen Chua, Perception and deception: Exploring Individual responses to deepfakes across different modalities, National Library of Medicine, (September 21, 2023), https://pmc.ncbi.nlm.nih.gov/articles/PMC10556585/

awareness campaigns. Without such protection, Deepfakes cause widespread, severe reputational harm.[62]

VIII. DISCUSSION

The research topic "From Victim to Villain: A Forensic and Linguistic Inquiry into the Construction of Criminal Identity in the Age of Deepfakes" is significant because deepfake technology is changing the way we understand truth, identity, and responsibility in today's digital world. Deepfakes are created using artificial intelligence (AI) and advanced techniques like generative adversarial networks (GANs), which make it possible to change a person's face, voice, and actions in a way that looks real. While deepfakes were originally developed for positive uses like entertainment, education, and accessibility, they are now often misused for spreading false information, damaging reputations, stealing identities, committing fraud, influencing politics, and creating explicit content without consent. Because of this, deepfakes have become not only a technological development but also a social and legal problem that affects privacy, dignity, justice, and public trust.[63]

Forensic and linguistic studies are key to tackling these issues. From a forensic point of view, experts analyze things like metadata, device records, digital signatures, and facial or voice patterns to detect whether a video or audio file is real or fake. However, as deep-fake technology becomes more advanced, even the best forensic tools find it difficult to spot manipulated content. This leads to situations where innocent people can be wrongly seen as criminals, shifting them from victims to villains. From a linguistic point of view, specialists study speech patterns, word choices, accents, and communication styles to find differences between a person's natural speech and manipulated deepfake material. This kind of analysis not only helps verify the

[62] Ben Colman, The Impact of Deepfakes on Brand and Reputation, Reality Defender, (August 4, 2024), https://www.realitydefender.com/insights/deepfakes-harm-brand-reputation
[63] Sami Alanazi et al., Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses, Human-Intelligent Systems Integration, (February 20, 2025), https://link.springer.com/article/10.1007/s42454-025-00060-4

authenticity of evidence but can also identify the creators or distributors of fake content.[64]

In India, there is still no specific law focused only on deepfakes, which makes ensuring accountability challenging. Existing laws like the "Information Technology Act, 2000"[65] (Sections 66C, 66D, 66E, 67, 67A, and 67B), the "Bharatiya Nyaya Sanhita, 2023"[66], the "POCSO Act, 2012"[67], and the "Copyright Act, 1957"[68] address some issues, such as identity theft, defamation, and obscenity, but there are serious gaps. For example, deepfakes used for political propaganda or reputation damage often don't fit neatly into these laws. The "IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021"[69] require social media platforms to take down harmful content, but do not force them to use watermarking, tracking tools, or AI-based detection systems to prevent fake media from spreading.

Cross-border challenges make things even harder, as many deepfake creators operate from outside India, making legal enforcement complicated. Without strong laws and reliable forensic tools, criminal identity in the digital age becomes blurry. Deepfakes can frame innocent people as guilty, while real offenders take advantage of legal loopholes and deny responsibility—this is called the "liar's dividend." Such situations weaken public trust in digital evidence, make it harder for courts to deliver justice, and change the way society views guilt and innocence.[70]

This topic is highly relevant for research because it brings together technology, forensics, linguistics, and law to study how deepfakes affect personal identity, legal systems, and social perception. It highlights the urgent need for strong regulations, better detection methods, and international cooperation to protect individuals and

[64] Dana Roemling & Jack Grieve, Forensic Authorship Analysis, Crest Research (Jan 14,2021), https://crestresearch.ac.uk/comment/forensic-authorship-analysis/
[65]Information Technology Act, No. 21 of 2000, Acts of Parliament, 2000 (India).
[66]See Bharatiya Nyaya Sanhita, No. 45 of 2023, Acts of Parliament, 2023 (India)
[67]See Protection of Children from Sexual Offences Act, No. 32 of 2012,
[68]Copyright Act, No. 14 of 1957, INDIA CODE (1957).
[69]See Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Gazette of India, Feb. 25, 2021 (India).
[70] Nikita Agarwal, Legal Challenges of Deepfake Technology and AI- Generated Content in India, Jus Corpus, (April 21,2025), https://www.juscorpus.com/legal-challenges-of-deepfake-technology-and-ai-generated-content-in-india/

maintain trust in digital content. Without these measures, deepfakes will continue to blur the line between truth and lies, turning real victims into perceived villains and weakening the foundations of justice in the modern world.[71]

The selected case studies collectively demonstrate how deep-fake technology blurs the boundaries between victim and villain, shaping criminal identity in complex, dynamic, and sometimes contradictory ways. Through forensic and linguistic analysis, these cases reveal how advanced artificial intelligence manipulates appearances, voices, and linguistic cues to exploit human trust, influence perceptions, and construct narratives of guilt or innocence. Across all cases, the analytical insights deepen our understanding of how deepfakes reshape narratives of crime and morality.

Forensic techniques ranging from facial dynamics analysis, audio spectrograms, metadata tracing, and stylometric profiling are central to uncovering manipulations. Yet, as deepfake quality advances, forensic challenges grow, demanding more sophisticated detection methods and cross-border collaboration. Linguistic forensics emerge as equally vital: attackers replicate authentic speech patterns, dialects, and idioms to build trust, but inconsistencies in syntax, semantics, and vocabulary often reveal fakes. However, as AI systems learn from vast datasets, even linguistic cues are becoming harder to detect, raising the stakes for expert intervention.[72]

Crucially, these cases illuminate the shifting villain/victim narrative in the age of deepfakes. Victims often face skepticism, reputational harm, or even legal scrutiny before forensic intervention clarifies the truth. Perpetrators, meanwhile, hide behind

[71] Sami Alanazi et al., Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses, Human-Intelligent Systems Integration, (February 20, 2025), https://link.springer.com/article/10.1007/s42454-025-00060-4

[72]Alexandra Bacon, *Engineering Giant Arup's CFO Deepfake Scam Costs $25 Million*, Bus. Insider (May 17, 2024), https://www.businessinsider.com/engineering-giant-arup-target-of-25-million-deepfake-scam-2024-5; *Kerala's First Deepfake Fraud Case*, Kozhikode Cyber Police Station Crime No. 101/2023 (India).; Kathryn Snowdon, *Deepfake Audio Used in Custody Battle, Lawyer Reveals Doctored Evidence*, *Telegraph* (Jan. 31, 2020), https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/.; Taylor Swift Explicit Deepfake Scandal Reported, BBC News (Jan. 25, 2024), https://www.bbc.com/news/technology-deepfake-scandal.; The Election Commission directed political parties to remove manipulated or misleading electoral content within three hours, flagging its potential to sway voter opinions, *Indian Express* (May 6, 2024)

synthetic identities, remaining faceless villains shielded by technology. In other contexts, deepfakes are used to manufacture credibility, as seen in political manipulation, showing how identities can be artificially elevated or diminished based on strategic goals. These case studies collectively justify the research focus on "From Victim to Villain."

Deepfakes do more than create fabricated content they reshape social, legal, and psychological realities**,** challenging how identities are constructed, perceived, and adjudicated. Forensic and linguistic tools serve as the primary defenses against these manipulations, yet their limitations reveal the fragility of truth in the digital age. Understanding this interplay is essential for building stronger legal protections, advancing detection technologies, and safeguarding public trust in an era where seeing is no longer believing**.**[73]

The legal responses of the United Kingdom and the United States reflect a growing recognition that deepfakes reshape criminal identity by blurring the distinction between victims and perpetrators. In the United States, the TAKE IT DOWN Act, enacted in 2025, marks a significant federal development by criminalising the knowing publication of non-consensual intimate deepfakes and imposing a 48-hour takedown obligation on covered online platforms. Individuals in deepfake-related crimes are frequently unintentional victims—for example, when their private photos or likenesses are manipulated—but they can also be portrayed as villains through fraudulently attributed content, especially in politically or sexually sensitive circumstances.

In the United Kingdom, revisions to the "Sexual Offences Act 2003" through the "Online Safety Act 2023" reflect a knowledge that harm occurs not only at the time of

[73]Alexandra Bacon, *Engineering Giant Arup's CFO Deepfake Scam Costs $25 Million*, Bus. Insider (May 17, 2024), https://www.businessinsider.com/engineering-giant-arup-target-of-25-million-deepfake-scam-2024-5; *Kerala's First Deepfake Fraud Case*, Kozhikode Cyber Police Station Crime No. 101/2023 (India).; Kathryn Snowdon, *Deepfake Audio Used in Custody Battle, Lawyer Reveals Doctored Evidence*, *Telegraph* (Jan. 31, 2020), https://www.telegraph.co.uk/news/2020/01/31/deepfake-audio-used-custody-battle-lawyer-reveals-doctored-evidence/.; Taylor Swift Explicit Deepfake Scandal Reported, BBC News (Jan. 25, 2024), https://www.bbc.com/news/technology-deepfake-scandal.; The Election Commission directed political parties to remove manipulated or misleading electoral content within three hours, flagging its potential to sway voter opinions, *Indian Express* (May 6, 2024)

distribution, but also throughout the act of creation. This is consistent with the book's notion that deep-fake crimes necessitate a forensic lens that evaluates purpose and procedure, rather than just outcome. The UK approach recognizes the language and visual manipulation involved in establishing false identities and reputations by criminalizing both distribution and soon, the creation of sexually explicit deepfakes.[74]

In contrast, the United States lacks a comprehensive federal framework, but state-level legislation—such as California's limitations on sexual deepfakes, Texas's criminalization of election-related deepfakes, and expansion of revenge porn laws—reflects the study's emphasis on context-driven damages. The planned "DEEPFAKES Accountability Act" and "NO FAKES Act" seek disclosure and watermarking procedures, which are consistent with the book's linguistic study. These approaches directly address the authenticity dilemma that deepfakes cause, in which language and visuals combine to create misleading narratives that influence public perception and, as a result, form criminal identity.[75]

Both countries' answers show how deepfakes undermine standard notions of victims and blame. As emphasized, forensic and linguistic tools are becoming increasingly important in resolving disputes over authorship, permission, and validity. The UK's positive strategy contrasts with the USA's uneven approach, but both reflect a larger trend: the law is gradually responding to modern reality by redefining criminal identity through regulation, evidence preservation, and victim protection.

The lack of deepfake-specific regulation in India creates substantial hurdles in combating crimes, including modified intimate images, political propaganda, and AI-generated impersonation. Currently, India relies on broad laws outlined in the "Information Technology Act of 2000" (IT Act)[76], the "Bharatiya Nyaya

[74]Sexual Offences Act 2003, c. 42, § 66A (as amended 2023) (U.K.); Online Safety Act 2023, c. 49, § 187 (U.K.); Commencement Regulations, The Sexual Offences Act 2003 (Commencement No. 22) Regulations 2023, SI 2023/1293 (U.K.) (in force Jan. 31, 2024).; Online Safety Act 2023, c. 49, §§ 65–70 (U.K.) (platform duties and Ofcom enforcement powers); Ofcom, *Protecting People from Harmful Online Content Under the Online Safety Act 2023: Guidance on Illegal Content Duties* (Apr. 2024), https://www.ofcom.org.uk

[75] DEEPFAKES Accountability Act, H.R. 3230, 116th Cong. (2019); Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2023 ("NO FAKES Act"), Discussion Draft, 118th Cong. (2023).

[76]Information Technology Act, No. 21 of 2000, Acts of Parliament, 2000 (India).

Sanhita 2023",[77] which include offences such as obscenity, cyber harassment, and identity theft. However, these rules fall short in addressing crucial issues like non-consensual creation, platform accountability, and digital transparency. To overcome these gaps, India can take features from the legal frameworks of the UK and the USA.[78]

The UK's "Online Safety Act 2023" and revisions to the "Sexual Offences Act 2003" provide strong victim-centric safeguards by criminalizing both the dissemination and fabrication of non-consensual deepfakes, recognizing that harm begins at the point of production[79]. Furthermore, the UK authorizes Ofcom to impose obligations on online platforms to detect and quickly remove illegal content. India's "IT Rules 2021" require intermediary obligations but lack strong enforcement tools.[80]

Similarly, the US's layered approach provides additional information. Proposed laws include the "NO FAKES Act", which protects persons' voices and likenesses from illegal AI use, and the "DEEPFAKES Accountability Act", which requires watermarking and disclosure labels to improve digital transparency. U.S. states such as Texas and California have also implemented election-related deepfake controls, which India might replicate to protect democratic integrity.[81]
By combining these techniques, India can create a comprehensive deepfake regulatory environment that criminalizes both production and misuse, enforces platform responsibility, and includes forensic detection. Such amendments would modernize

[77]See Bharatiya Nyaya Sanhita, No. 45 of 2023, Acts of Parliament, 2023 (India)
[78] Archita Bhargava, Deepfake Technology and its Legal Regulation in India: A Doctrinal & Comparative Study, Vintage Legal, (last visited August 27,2025), https://www.vintagelegalvl.com/post/deepfake-technology-and-it-s-legal-regulation-in-india-a-doctrinal-and-comparative-study
[79]Sexual Offences Act 2003, c. 42, § 66A (as amended 2023) (U.K.); Online Safety Act 2023, c. 49, § 187 (U.K.); Commencement Regulations, The Sexual Offences Act 2003 (Commencement No. 22) Regulations 2023, SI 2023/1293 (U.K.) (in force Jan. 31, 2024).
[80]See Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Gazette of India, Feb. 25, 2021 (India).
[81]DEEPFAKES Accountability Act, H.R. 3230, 116th Cong. (2019);Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2023 ("NO FAKES Act"), Discussion Draft, 118th Cong. (2023); Cal. Civ. Code § 1708.86 (West 2023; Cal. Elec. Code § 20010 (West 2023); Tex. Elec. Code Ann. § 255.004 (West 2023).

India's legal system, better protect victims, and meet the emerging issues faced by deepfake-related crimes.[82]

In April 2021, the European Commission proposed the first EU artificial intelligence law, which would establish a risk-oriented AI categorization system. Artificial intelligence (AI) algorithms that can be used in a range of applications are evaluated and classified according to the risk they pose to users. AI compliance obligations vary depending on the risk level. The primary purpose was to make sure that AI systems used in the EU were secure, accessible, accessible, nondiscriminatory, and ecologically sound. To prevent unwanted consequences, AI systems need to be monitored through human beings instead of autonomous systems. Another goal was to offer a technologically consistent concept of AI that may be used by upcoming artificial intelligence systems.[83]

In June 2024, the European Union formally signed the Artificial Intelligence Act, the world's first comprehensive regulation governing artificial intelligence. The Artificial Intelligence Act will be completely effective 24 months after its enactment, however some provisions will take effect sooner. The prohibition on AI systems that pose unacceptable threats went into effect on February 2, 2025. Codes of practice will become effective nine months after they are implemented. Rules for general-purpose AI systems that must meet transparency standards will take effect 12 months after they are enacted. Highly dangerous platforms are given more time to comply with the requirements because their responsibilities come into effect 36 months after the entry into force.[84]

[82]Archita Bhargava, Deepfake Technology and its Legal Regulation in India: A Doctrinal & Comparative Study, Vintage Legal, (last visited August 27,2025), https://www.vintagelegalvl.com/post/deepfake-technology-and-it-s-legal-regulation-in-india-a-doctrinal-and-comparative-study

[83] EU AI Act: First Regulation on Artificial Intelligence, European Parliament, (February 19,2025), https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[84]EU AI Act: First Regulation on Artificial Intelligence, European Parliament, (February 19,2025), https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

"The European Union Artificial Intelligence Act" (EU AI Act)[85], passed in March 2024, is the world's first complete legal framework governing artificial intelligence, and it has important ramifications for deepfake technology. Because deepfakes contain AI-generated synthetic media capable of manipulating voice, video, and images, the Act specifically addresses their potential hazards, principally through transparency, accountability, and risk management requirements.[86]
First, the Act requires deepfake content to be transparent. Any AI-generated or modified media, such as films, photos, or audio samples, must be clearly labeled to tell consumers that the content is artificially created. This provision aims to prevent misinformation, identity fraud, and reputational harm by informing the audience about synthetic content. Second, the Act establishes a risk-based categorization of AI systems, classifying certain uses of deepfakes as high-risk or even unacceptable. Deepfakes employed in election manipulation, nonconsensual intimate photos, or identity fraud without disclosure, for example, may be subject to harsh restrictions or outright ban.

Deepfakes for entertainment or educational reasons are acceptable, but they must still be watermarked and labeled to ensure ethical use.[87] In summary, the "EU AI Act" seeks a compromise between innovation and regulation, allowing for creative uses of deepfakes while protecting individuals and democratic processes from harmful manipulation. It establishes a global standard for responsible AI governance, with the potential to impact comparable legislative systems around the world.[88]

IX. CONCLUSION AND RECOMMENDATIONS

This study has examined the complex interplay between technology, law, and identity in the age of deepfakes, focusing on the phenomenon of victims being reimagined or

[85] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act) [2024] OJ L 1689.
[86] AI Act, European Commission (last visited Aug. 27, 2025), https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[87] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (Artificial Intelligence Act) arts 50 and 52(3) [2024] OJ L 1689.
[88] The AI Act: EU's Giant Leap Towards Responsible AI Governance, Policy Circle Bureau (Mar. 14, 2024), https://www.policycircle.org/policy/eus-ai-act-model-legislation/

misrepresented as villains. The findings reveal that deepfake incidents often blur traditional legal categories of perpetrator and victim, producing situations in which individuals whose likeness or voice has been manipulated face social stigma, reputational damage, and even criminal investigation. These cases demonstrate that harm is not limited to direct perpetrators' actions; it also arises from the interpretive processes of law enforcement, the media, and the public when confronted with fabricated yet convincing digital artefacts.[89]

The research underscores that existing Indian legal provisions spread across the "IT Act", "BNS", "POCSO", and the "Copyright Act" provide partial coverage but lack a unified framework that fully captures the harms caused by deepfake manipulation, especially in non-sexual contexts. This gap is further compounded by forensic and linguistic challenges, such as proving the artificiality of content and demonstrating intent. Without consistent protocols for evidence preservation, analysis, and admissibility, victims may remain vulnerable to both initial harm and secondary victimisation through flawed investigative processes.[90]

Ultimately, the blurred lines between victim and villain in deepfake cases call for comprehensive reform: integrating digital forensic protocols into standard practice, formally recognising linguistic analysis as admissible and reliable evidence, and embedding a legal understanding of deepfake manipulation into the broader doctrine of criminal identity construction.[91]

[89]Danielle Keats Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Calif. L. Rev. 1753, 1788–94 (2019); Britt Paris & Joan Donovan, *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*, Data & Soc'y Research Inst. 8–10 (2019), https://datasociety.net/library/deepfakes-and-cheap-fakes

[90] Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67–67B, INDIA CODE (2000); Bharatiya Nyaya Sanhita, No. 45 of 2023, INDIA CODE (2023); Protection of Children from Sexual Offences Act, No. 32 of 2012, INDIA CODE (2012); Copyright Act, No. 14 of 1957, INDIA CODE (1957); Bharatiya Sakshya Adhiniyam, No. 46 of 2023, § 63(4), INDIA CODE (2023); Aparna Chandra, Anuj Bhuwania & Sital Kalantry, *Law Courts and Deepfakes: Emerging Evidentiary Challenges*, 14 Indian J.L. & Tech. 1, 21–27 (2021).

[91]Raghav Aggarwal, *Deepfakes and the Indian Legal System: Evidentiary and Investigative Challenges*, 4 NALSAR Student L. Rev. 45, 59–62 (2022); Lorraine Hope & Aldert Vrij, *Expert Testimony and Emerging Technologies: Forensic Evidence in Court*, 28 Psych. Pub. Pol'y & L. 105, 112–15 (2022); Evelyn Douek, *The Rise of Content Cartels*, 134 Harv. L. Rev. 1409, 1432–35 (2021).

To address the identified gaps and emerging risks, several targeted reforms are proposed:

1. **Legislative Amendments:** Indian law should be updated to explicitly criminalise the creation, possession, and dissemination of harmful deepfakes, including non-consensual intimate content and election-related disinformation. Provisions should also be introduced to establish a statutory right of publicity, covering likeness and voice, thereby aligning India with evolving global norms.[92]
2. **Specialised Forensic-Linguistic Laboratories:** Dedicated centres equipped to conduct integrated digital forensics and linguistic analysis should be established at the national and state levels. These labs would develop and standardise methodologies for detecting and authenticating deep-fake content, ensuring evidentiary reliability.[93]
3. **AI-Driven Real-Time Verification:** Digital platforms operating in India should be mandated or incentivised to deploy AI-based real-time verification systems capable of flagging and labelling synthetic media. Such systems could integrate watermark detection, metadata analysis, and cross-referencing with verified content repositories to ensure accurate and reliable results.[94]
4. **Public Awareness Campaigns:** National and regional awareness drives are essential to educate citizens about deepfake risks, detection tools, and reporting mechanisms. Public literacy in recognising potential

[92] Sexual Offences Act 2003, c. 42, § 66A (as amended 2023) (U.K.); Tex. Elec. Code Ann. § 255.004 (West 2023); Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2023 ("NO FAKES Act"), Discussion Draft, 118th Cong. (2023); Rakesh Kumar Singh, *The Case for a Statutory Right of Publicity in India*, 9 Indian J.L. & Tech. 122, 130–34 (2019).
[93] Lorraine Hope & Aldert Vrij, *Expert Testimony and Emerging Technologies: Forensic Evidence in Court*, 28 Psych. Pub. Pol'y & L. 105, 111–15 (2022); Siwei Lyu, *Deepfake Detection: Current Challenges and Next Steps*, 2 Frontiers Comput. Sci. 1, 3–7 (2020).
[94] DEEPFAKES Accountability Act, H.R. 3230, 116th Cong. (2019); Online Safety Act 2023, c. 49, §§ 65–70 (U.K.); Ofcom, *Protecting People from Harmful Online Content Under the Online Safety Act 2023: Guidance on Illegal Content Duties* (Apr. 2024), https://www.ofcom.org.uk

manipulations will reduce susceptibility to misinformation and protect individuals from reputational harm.[95]

5. **Capacity-Building for the Legal System:** Judges, police officers, and legal practitioners should receive specialised training in digital forensics, AI-based evidence assessment, and the socio-legal implications of deepfakes. This would ensure informed adjudication and more effective investigative responses, reducing the likelihood of victims being mischaracterized as offenders.[96]

Collectively, these measures would move India toward a more coherent, victim-centered approach to deepfake regulation. The enactment of the U.S. TAKE IT DOWN Act demonstrates the practical value of combining criminal liability with expedited platform takedown obligations, an approach that India could adapt to safeguard both individual rights and the integrity of the justice system.

X. REFERENCES

A. Books

1. Citron, Danielle Keats. *Sexual Privacy*. Yale Law Journal, vol. 128, 2019.
2. Nappinai, N. S. *Technology Laws Decoded*. 2nd ed., OakBridge Publishing, 2022.

B. Journal Articles

1. Aggarwal, Raghav. "Deepfakes and the Indian Legal System: Evidentiary and Investigative Challenges." *NALSAR Student Law Review*, vol. 4, 2022, pp. 45–62.

[95] U.N. Educ., Sci. & Cultural Org. (UNESCO), *Balancing Freedom of Expression and Addressing Disinformation: A Guide to Policy Making* 22–25 (2022), https://unesdoc.unesco.org; Press Release, HM Gov't, Dep't for Sci., Innovation & Tech., *Government to Strengthen the Law Against Deepfake Intimate Images* (Nov. 2023), https://www.gov.uk/government/news/government-to-strengthen-the-law-against-deepfake-intimate-images

[96] Nat'l Judicial Acad., *Annual Report 2022–23* 44–47 (2023), https://nja.gov.in (noting the importance of judicial training in emerging technologies); Danielle Keats Citron, *Sexual Privacy* 128 Yale L.J. 1870, 1935–39 (2019) (highlighting risks of mischaracterisation of victims in digital abuse cases); Aparna Chandra, Anuj Bhuwania & Sital Kalantry, *Law Courts and Deepfakes: Emerging Evidentiary Challenges*, 14 Indian J.L. & Tech. 1, 28–32 (2021).

2. Alanazi, Sami, et al. "Unmasking Deepfakes: A Multidisciplinary Examination of Social Impacts and Regulatory Responses." *Human-Intelligent Systems Integration*, 2025.
3. Chandra, Aparna, Anuj Bhuwania, and Sital Kalantry. "Law Courts and Deepfakes: Emerging Evidentiary Challenges." *Indian Journal of Law and Technology*, vol. 14, 2021, pp. 1–32.
4. Chesney, Robert, and Danielle Keats Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review*, vol. 107, no. 6, 2019, pp. 1753–1820.
5. Douek, Evelyn. "The Rise of Content Cartels." *Harvard Law Review*, vol. 134, 2021, pp. 1409–1435.
6. Farid, Hany. "Creating, Using, Misusing, and Detecting Deep Fakes." *Journal of Online Trust and Safety*, vol. 2, no. 1, 2020.
7. Farid, Hany. "Digital Forensics in a Post-Truth Age." *Forensic Science International: Digital Investigation*, vol. 36, 2021.
8. Goodfellow, Ian, et al. "Generative Adversarial Nets." In *Advances in Neural Information Processing Systems 27*, 2014, pp. 2672–2680.
9. Hope, Lorraine, and Aldert Vrij. "Expert Testimony and Emerging Technologies: Forensic Evidence in Court." *Psychology, Public Policy, and Law*, vol. 28, no. 1, 2022, pp. 105–115.
10. Lyu, Siwei. "Deepfake Detection: Current Challenges and Next Steps." *Frontiers in Computer Science*, vol. 2, 2020.
11. Paris, Britt, and Joan Donovan. *Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence*. Data & Society Research Institute, 2019.
12. Roemling, Dana, and Jack Grieve. "Forensic Authorship Analysis." CREST Research, 2021.
13. Stamm, Matthew, et al. "Information Forensics: An Overview of the First Decade." *IEEE Transactions on Information Forensics and Security*, 2020.

14. Zong, Chengqing, et al. "Linguistic Characteristics of Machine-Generated Text." *Journal of Computational Linguistics and Language Technology*, vol. 2, 2020.

C. Reports and Policy Papers

1. Brookings Institution. *Deepfakes and Synthetic Media: Policy Implications*. 2020.
2. National Conference of State Legislatures. *State Legislation Addressing Deepfakes*. Updated Apr. 17, 2024.
3. UNESCO. *Balancing Freedom of Expression and Addressing Disinformation: A Guide to Policy Making*. 2022.

D. Statutes and Regulations (India)

1. Bharatiya Nyaya Sanhita, No. 45 of 2023.
2. Bharatiya Sakshya Adhiniyam, No. 46 of 2023.
3. Copyright Act, No. 14 of 1957.
4. Information Technology Act, No. 21 of 2000.
5. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021.
6. Protection of Children from Sexual Offences Act, No. 32 of 2012.

E. Statutes and Regulations (United Kingdom)

1. Online Safety Act 2023, c. 49.
2. Sexual Offences Act 2003, c. 42.

F. Statutes and Regulations (United States)

1. Cal. Civ. Code § 1708.86.
2. Cal. Elec. Code § 20010.
3. DEEPFAKES Accountability Act, H.R. 3230, 116th Cong. (2019).
4. NO FAKES Act, Discussion Draft, 118th Cong. (2023).
5. TAKE IT DOWN Act, S. 146, 119th Cong. (2025).

6. Tex. Elec. Code Ann. § 255.004.

7. Va. Code Ann. § 18.2-386.2.

G. European Union Legislation

1. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).

H. Cases

1. *Shreya Singhal v. Union of India*, (2015) 5 SCC 1 (India).

I. Newspaper and Magazine Articles

1. Bacon, Alexandra. "Engineering Giant Arup's CFO Deepfake Scam Costs $25 Million." *Business Insider*, May 17, 2024.

2. Dixit, Pranav. "Taylor Swift Deepfake Images Spark Outrage, Highlighting AI's Dangers." *Wired*, Jan. 26, 2024.

3. Snowdon, Kathryn. "Deepfake Audio Used in Custody Battle, Lawyer Reveals Doctored Evidence." *The Telegraph*, Jan. 31, 2020.

4. Zakrzewski, Cat, and Taylor Lorenz. "Lawmakers Fast-Track Bill to Criminalize Non-Consensual Deepfake Porn." *The Washington Post*, Jan. 30, 2024.