DEEPFAKES AND CRIMINAL LIABILITY: A DOCTRINAL, COMPARATIVE AND EVIDENTIARY ANALYSIS

Sampada Singh, MIT World Peace University, Pune

Ananya Agarwal, MIT World Peace University, Pune

ABSTRACT

The augmentation of artificial intelligence has validated the creation of deepfakes - synthetic media where artificial intelligence is used to realistically replace or alter a person's likeness, voice, or actions in images, audio, or video. The term is a portmanteau of "deep learning" (a type of AI) and "fake," and is commonly used to describe both the technology and the manipulated content. This have given rise to a unique and difficult question of criminal liability. This paper scrutinizes the doctrinal and substantive criminal law response to deepfake- enabled offences through four interlinked lines of inquiry. It first examines the doctrinal foundations of criminal liability; it examines how actus reus is fragmented across the chain of events, how mens rea must be evaluated across various actors with divergent knowledge and intent, and how causation, mode of participation, and corporate liability principles apply to AI-mediated harm. It also examines whether AI systems, lacking legal personhood, can meaningfully be a locus of criminal fault, or whether the doctrine of innocent agency better explains liability for AI-facilitated offences. The paper then undertakes a comparative survey of statutory frameworks across India, the United States, the European Union, the United Kingdom, evaluating each jurisdiction's legislative approach to deepfake-related harm. It further analyses judicial responses, focusing on Indian judiciary's extension to personal rights and privacy jurisprudence to synthetic media in the absence of a dedicated legislation, while noting a marked absence of criminal – as opposed to civil-injunction – precedent in India. Finally, the paper examines various challenges of prosecuting deepfake offences, including the fact that MLAT requests to foreign platforms/servers can take months to years, during which evidence may be deleted or altered, section 334-336 and section 356 of Bharatiya Nyaya Sanhita, 2023 exist as remedies but fail to adequately address the unique challenges of deepfakes — they were designed for physical/textual forgery and human-authored defamation, not AI-synthesised likeness. This paper concludes that while courts have demonstrated considerable doctrinal ingenuity in adapting existing law to synthetic media nonetheless persistent gaps in mens rea calibration and evidentiary standard continue to weaken the criminal law's capacity to respond effectively to crimes facilitated by deepfake technology.

Introduction

The application of substantive criminal law to deepfake- enabled offences revels a fundamental doctrinal strain: criminal law was formulated around a unitary, identifiable wrongdoer whose act and intent could be traced within a continuous and single transaction. Deepfakes on the other hand, unfold across a distributed technological chain – content generation, initial upload, algorithmic amplification, and continued resharing – in which no single actor owns complete knowledge or control over the resulting harm. The analysis begins at the level of doctrinal study examining how the foundational building blocks of criminal liability which translate to conduct that unfolds not through a single actor but across a distributed technological chain of creators and hosting platforms. The doctrinal groundwork supplies the conceptual vocabulary necessary to evaluate everything that follows.

From doctrine, the study turns to positive law, undertaking a comparative evaluation of how different jurisdictions have translated these doctrinal principles into enforceable statute. India's fragmented reliance on the Information Technology Act,2000, the Bharatiya Nyaya Sanhita,2023, the Digital Personal Data Protection Act,2023, and successive amendments to the IT Rules is set against various emerging and existing regulatory models of United Nations, China and United Kingdom. This comparative study is not merely descriptive, it is designed to expose the enforcement and jurisdictional loopholes that persist within comparatively advanced statutory framework.

The paper accordingly turns to how courts particularly in India have exercised ingenuity to address deepfake- related harm in the absence of a dedicated legislation. The extension of personality rights and privacy jurisprudence reveals both the adaptive capacity of the existing legal doctrine and its limits – most notably, the near-total absence of criminal, as opposed to civil-injunctive, precedent addressing deepfake harm within the Indian legal system.

The final stage of analysis interrogates the adequacy of existing standards for authenticating electronic records in an environment of synthetic content. It examines the corrosive effect whereby the mere existence to the deepfakes enables genius evidence to be plausibly disavowed and considers the practical difficulties of forensic detection and cross-border evidence gathering that will ultimately determine whether the doctrinal and statutory framework discussed earlier can be given a practical effect.

CHAPTER 1: Doctrinal Foundation of Criminal Liability

Actus Reus in a Distributed Technological Chain

The actus reus of a deepfake offence is rarely a single, discrete act. It is more accurately perceived as a sequence of contributory acts performed by different people at different point of time. It includes multiple actors like the creator who created the synthetic content using a trained model, the first disseminator who uploads or transmits it, and its hosting platform whose infrastructure enables its continued circulation. The traditional offences like forgery, obscenity and criminal defamation assume a single act of making or publishing something. This makes it poorly suited to a pipeline where creation, distribution, and hosting are carried out by different people, in different place, often without any coordination between them. This fragmentation raises a threshold doctrinal question of whether liability should be attached to the entire chain as a continuing transaction, or whether each contributing act be assessed as an independent actus reus requiring its own mental element.

Mens Rea Across the Deepfake Chain

Images and video manipulation has legitimate uses in entertainment, education, and satire, so most deepfake-related offences cannot rest on the act of creation alone – they require proof of a heightened mental element. This calls for a calibrated approach to mens rea across the distributed chain described above. A creator who deliberately fabricates non-consensual intimate imagery or a fraudulent voice-cloned solicitation typically meets the threshold of intention or purposive knowledge. A first disseminator, who shares content without verifying its provenance, may only reach the level of constructive knowledge or wilful blindness. Platforms and downstream resharers present the weakest case, where liability, if any, rests more on recklessness or negligence than actual knowledge. The calibration of mens rea along this gradient offers a doctrinally coherent method of distributing criminal responsibility without collapsing into either impunity for knowing wrongdoers or disproportionate liability for innocent intermediaries.

Modes of Participation

When liability cannot be grounded in direct authorship of the act, criminal law traditionally relies on doctrines of secondary participation to reach the actors of the crime. Under Indian law

the doctrines of abetment, criminal conspiracy and common intention permit the prosecution of persons who instigate the creation of deepfake content, who engage another to produce it, or who act in furtherance of a shared criminal design, even absent direct authorship. These modes of participation are particularly notable in the deepfake context because it helps the prosecutors to reach the commissioning parties without requiring proof that the accused personally operated the generative tool. The challenge is applying these frameworks to loosely coordinated or anonymous online networks, where proving the "meeting of minds" required for conspiracy, or the knowledge required for abetment, is difficult through conventional evidence.

Corporate Liability and the Strict Liability Debate

Cooperate entities sit at multiple points along the deepfake pipeline – AI developers who build the generative models, deplores who repackage those models into consumer-facing applications, and platforms that host and circulate the resulting content. Attributing liability to any of these actors first requires resolving how liability attaches to a corporation at all — either by identifying a directing mind whose intent is imputed to it, or through a statutory vicarious liability regime that dispenses with such identification. The choice between these two models is not merely technical – it is precisely what determines whether platforms and toolmakers face liability only for what they knowingly permitted, or for what occurred on or through their systems regardless of fault. Most jurisdictions apply fault-based liability to platforms, trying safe-harbour loss to failure to act on known unlawful content, while developer liability remains contested, generally reserved for tools purposively designed or marketed for unlawful ends. This approach has costs: it offers weak deterrence against platforms operating at scale and burdens prosecutors with proving knowledge in anonymised chains, yet strict liability risks chilling legitimate activity. Comparative practice increasingly favours a hybrid: fault-based liability for creators and knowing disseminators, and a modified strict liability standard for platforms — triggered not by knowledge of specific content but by their failure to implement reasonable detection and takedown measures.

The "AI as Offender" Problem and the Doctrine of Innocent Agency

Gabriel Hallevy's influential taxonomy of criminal liability models for AI system identifies several conceptual frameworks by which liability might be attributed in AI-mediated offences, ranging from treating AI system as a mere instrument of a human principal to more expansive

models that contemplate direct or probable-consequence liability for autonomous systems. The overwhelming weight of doctrinal opinion, favours the former position: because an AI system lacks the capacity for legal personhood, intention, or moral culpability, it cannot itself be prosecuted, and criminal responsibility must instead be traced back to a human. This position finds support in the doctrine of innocent agency, used to address offences committed through the medium of a person lacking criminal capacity i.e., unsound mind or a child, whereby the law attributes the act of the innocent agent to the culpable principal who directed or procured it. Applied to generative AI, the doctrine treats the tool as an innocent instrumentality, with criminal responsibility vesting in the human actor who deploys it with guilty intent.

CHAPTER 2: Comparative Statutory Frameworks

The emergence of hyper-realistic "deepfakes"—synthetically generated information (SGI) created using artificial intelligence—has created a global legal crisis. These technologies are increasingly used for non-consensual intimate imagery (NCII), financial fraud, and political disinformation. Consequently, major jurisdictions are transitioning from applying fragmented, existing laws to enacting technology-specific criminal statutes.

India: The IT Rules 2026 Shift India has significantly modernized its regulatory landscape with the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, effective February 20, 2026.Statutory Integration: The framework integrates the IT Act, 2000, the Digital Personal Data Protection (DPDP) Act, 2023, and the Bharatiya Nyaya Sanhita (BNS), 2023.Rapid Takedown Mandates: Platforms must now remove non-consensual deepfakes within two to three hours of a complaint or government/court order. Transparency and Traceability: The 2026 amendments mandate that all AI-generated audio, video, and images carry visible labels, watermarks, or permanent metadata to ensure traceability. Criminal Liability: While the IT Act (Sections 66E, 67) addresses privacy and obscenity, the BNS provides updated criminal provisions for fraud and impersonation committed through synthetic media. A Shift Toward Strict Platform Accountability India's regulatory landscape for synthetic media underwent a major transformation with the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026. These rules formally recognize "Synthetically Generated Information (SGI)"—defined as digitally created or modified content that is indistinguishable from real life—as a critical regulatory category. Criminal Liability &

Penalties: Criminal liability for deepfakes is increasingly channelled through the Bharatiya Nyaya Sanhita (BNS), 2023, which replaced the IPC. The 2026 IT amendments align platform references with the BNS to address deepfake- driven fraud, impersonation, and obscenity. Key Mandates: The 2026 rules impose a three-hour takedown window for certain AI-generated content following government or court orders—one of the fastest global standards. Intermediaries must also implement mandatory labelling and traceability for synthetic content to prevent misuse.

United States: Federal and State Patchwork In 2025, the U.S. established its first comprehensive federal response to deepfake abuse. Federal Action: The TAKE IT DOWN Act, signed into law on May 19, 2025, criminalizes the non-consensual publication of intimate "digital forgeries" (deepfakes). It mandates that "covered platforms" establish notice-and-removal processes within 48 hours. Civil Remedies: The DEFIANCE Act (introduced Jan 2026) aims to build on this by providing victims with a direct civil path to sue platforms and creators of sexually explicit deepfakes. State Level: States like Florida, Tennessee (notably the ELVIS Act protecting voice/likeness), Washington, and California continue to lead with local statutes addressing non-consensual pornography and election-related deepfakes. Emerging Federal Protections and State Patchwork The U.S. has transitioned from pure state-level regulation to a burgeoning federal framework focused on non-consensual sexual content. Federal Action: The TAKE IT DOWN Act, signed May 19, 2025, criminalizes the publication of non-consensual intimate imagery (NCII), including AI-generated deepfakes. It mandates a 48-hour removal process for platforms. Complementing this, the DEFIANCE Act provides victims with civil relief for "intimate digital forgeries". State-Level Innovations: Tennessee's ELVIS Act (Ensuring Likeness, Voice, and Image Security) of 2024 is ground-breaking for protecting an individual's voice as a property right. Violations are a Class A misdemeanour, punishable by nearly a year of incarceration. Other states like California and Washington continue to refine their own protections against deceptive deepfakes in elections and entertainment.

European Union: The AI Act and DSA The EU utilizes a risk-based approach, focusing on systemic transparency rather than purely criminal penalties.AI Act Article 50: Mandates that providers of AI systems ensure machine-readable marking and detectability of synthetic content. Deplorers must clearly inform users when they are interacting with deepfakes. Regulatory Overlap: Transparency obligations under the AI Act facilitate compliance with the

Digital Services Act (DSA), which holds platforms accountable for content moderation. The GDPR provides victims with legal recourse through the "Right to Erasure" and "Right to Object". Transparency and Risk-Based Regulation The EU utilizes a comprehensive, horizontal approach through the AI Act, which interacts with the GDPR and the Digital Services Act (DSA). Transparency Obligations: Under Article 50 of the AI Act, providers and deployers must clearly label deepfakes or AI-generated text on public interest matters in a machine-readable format. Legal Interplay: The Act necessitates that AI documentation overlaps with GDPR requirements, such as risk assessments. The DSA further imposes intermediary obligations to proactively address "systemic risks," including the spread of harmful synthetic media.

United Kingdom: Online Safety Act 2023The UK has aggressively criminalized the creation and distribution of harmful synthetic media. Criminalization: Amendments to the Online Safety Act (OSA) 2023 and the Criminal Justice Bill (April 2024) made it a criminal offense to create sexually explicit deepfakes without consent, even if they are not shared. Liability: Creators can face an unlimited fine and a criminal record, while those who circulate such images face potential imprisonment. Targeting Intimate Image Abuse The UK has specifically weaponized the Online Safety Act 2023 against deepfake-related sex crimes. Criminalization: New offenses introduced in early 2024 explicitly include deepfakes as a form of intimate image abuse. Creating non-consensual sexual deepfakes is now a criminal offense. Enforcement: Platforms face extra duties to proactively prevent such content. In early 2026, the government signalled further tightening, requiring tech firms to remove abusive images within 48 hours.

Other Key Jurisdictions China: Enforces some of the world's strictest "Deep Synthesis" regulations, requiring active government oversight and mandatory watermarking of AI-generated media. The Deep Synthesis Regulations (2023) mandate explicit labelling of doctored content and require service providers to obtain separate consent before editing biometric information like faces or voices. South Korea: Following a widespread deepfake crisis, laws were amended in late 2024 and 2026 to criminalize not only the production but also the possession and viewing of sexually explicit deepfakes, carrying penalties of up to seven years in prison. Following a national crisis over Telegram-based deepfake sex crimes, South Korea passed legislation in September 2024 criminalizing not just the production, but also the watching or possessing of sexually explicit deepfakes, carrying penalties of up to 3 years in prison.

CHAPTER 3: Judicial Responses and Case Law Analysis

Indian High Court Personality-Rights Jurisprudence

The Delhi High Court has emerged as a rudiment forum through which Indian court have addressed deepfake-related harm, doing so primarily through the vehicle of personality rights rather than a deepfake-specific statutory provision. In Anil Kapoor v. Simply Life India & Ors. (2023), the court ordered broad injunctive relief restraining the unauthorised commercial and AI-driven use of the actor's name, image, voice, and likeness, holding that such misappropriation implicated both personality rights and the constitutional guarantee of privacy under Article 21. The significance of this decision lies not merely in its outcome but in its reasoning: the court treated AI-driven manipulation of likeness as an extension of familiar misappropriation rather than something categorically new, avoiding the need to wait for dedicated legislation. This approach was extended and refined in Ankur Warikoo v. John Doe & Ors., where the Delhi High Court granted interim relief against unidentified parties who had circulated AI-generated deepfake videos of a finance educator to lure victims into fraudulent investment schemes — demonstrating both a willingness to combine personality-rights doctrine with cyber-fraud concerns, and a procedural mechanism, through relief against unidentified "John Doe" defendants, for addressing the anonymity that so often insulates deepfake creators from accountability. The trajectory reached its most expansive point in Sadhguru Jagadish Vasudev & Anr. v. Igor Isakov & Ors. (2025), where the Delhi High Court extended personality-rights protection explicitly to AI-generated reproductions and chatbot impersonations, confirming that the judicially recognised right extends beyond visual and vocal likeness to encompass conversational and interactive synthetic representations. This suggests Indian courts are prepared to construe personality rights broadly enough to keep pace with evolving generative technology, rather than confining protection to the specific manifestations of misuse that happened to arise in earlier cases. Taken together, these three decisions establish a reasonably settled judicial principle: non-consensual synthetic replication of a person's identity, in whatever modality, constitutes a legally cognisable harm remediable through injunctive relief, even absent a dedicated statutory basis.

A Brief Comparative Glimpse

Developments in the United Nations offer a useful point of contrast. The TAKE IT DOWN Act secured its first criminal conviction in April 2026, involving an individual who used AI to

generate and distribute non-consensual intimate imagery of both adults and minors — demonstrating that criminal, rather than merely civil, enforcement is achievable where clear statutory language exists. At the same time, American courts have shown far less deference toward deepfake-specific regulation in the politically sensitive domain. These two developments suggest a jurisdiction-specific pattern that recurs across many legal systems: criminalisation of non-consensual intimate imagery attracts comparatively little constitutional resistance, whereas restrictions on political and expressive synthetic content face substantially higher judicial scrutiny.

Evidentiary Jurisprudence Applied to Synthetic Media

The personality-rights decisions above assume courts can reliably determine, as a threshold matter, that the content is synthetic and falsely attributed to the claimant. This places heavy reliance on India's existing electronic evidence framework — developed long before generative AI existed. The Supreme Court's decision in Anvar P.V. v. P.K. Basheer (2014), which mandated strict compliance with certification requirements for the admissibility of electronic records, and the earlier position in State of Maharashtra v. Praful Desai (2003) on the reception of electronic and remote evidence, together constitute the governing evidentiary architecture that trial courts must now apply to synthetic media. Neither decision contemplated content that might itself be an AI-generated fabrication, rather than an authentic but improperly certified recording — leaving open the harder question of how courts are to adjudicate contested claims of synthetic manipulation once procedural certification alone can no longer guarantee authenticity.

The Constitutional Backdrop

The judicial extension of personality rights to synthetic media does not occur in a doctrinal vacuum; it draws directly on the constitutional foundations laid by the Supreme Court in two decisions of broader significance. Justice K.S. Puttaswamy (Retd.) v. Union of India (2017), in recognising informational privacy as an incident of the right to life and personal liberty under Article 21, supplies the constitutional basis on which High Courts have grounded the individual's interest in controlling the use of their own image, voice, and likeness. Shreya Singhal v. Union of India (2015), while striking down the overbroad Section 66A of the IT Act, simultaneously affirmed that reasonable and narrowly tailored restrictions on online speech remain constitutionally permissible. These two decisions together frame the outer

boundaries within which any future deepfake-specific criminal provision must operate: broad enough to protect individuals against non-consensual synthetic misappropriation, yet precise enough to withstand the vagueness-based constitutional challenge that proved fatal to Section 66A.

The Absence of Criminal Precedent

Notwithstanding the doctrinal richness of the personality-rights line of cases examined above, a critical gap becomes apparent on closer inspection: every significant judicial intervention discussed in this chapter has been civil and injunctive in character, arising from suits for permanent or interim injunction rather than from criminal prosecution. While First Information Reports have been registered under Sections 66C, 66E, and 67 of the IT Act in numerous deepfake pornography cases involving public figures and private individuals alike, reported appellate criminal jurisprudence interpreting these provisions specifically in the deepfake context remains conspicuously absent. Civil injunctive relief, however sophisticated, offers only content removal and, at most, damages — it lacks the deterrent weight of criminal sanction and the broader public-interest vindication that prosecution serves. This shows that judicial engagement has stayed concentrated in civil cases at the High Courts, not criminal trials. Either criminal cases aren't being appealed far enough to create precedent, or they're being decided at the trial stage without written judgments that clarify how Sections 66C, 66D, and 67A should apply to AI-generated content. This gap connects directly to the doctrinal uncertainties around mens rea and causation, and the evidentiary difficulties in authenticating synthetic media: without a developed body of criminal appellate precedent, prosecutors, trial courts, and victims are left to navigate these questions with only civil law analogies for guidance.

CHAPTER 4: Evidentiary and procedural Challenges

This chapter focuses on the complex evidentiary and procedural hurdles that deepfakes and AI-generated content pose within the criminal justice system. As deepfake technology becomes increasingly sophisticated, traditional assumptions about digital evidence—historically summed up as "seeing is believing"—are being fundamentally undermined.

Authentication of Electronic Evidence (BSA ss.65A/65B)

The legal foundation for admitting digital evidence in India is governed by the Bharatiya

Sakshya Adhiniyam (BSA), 2023, which carries over provisions from Sections 65A and 65B of the Indian Evidence Act. Self-Contained Code: These sections form a complete code for the admissibility of electronic records. They require specific conditions to be met, including a certification from the person responsible for the computer system. Primary vs. Secondary Evidence: Under the BSA, electronic records are treated as primary evidence if the original device is produced. For secondary evidence, such as copies, a certificate with a hash value (a digital fingerprint) is mandatory to ensure the evidence has not been tampered with. Post-Truth Challenges: Deepfakes magnify the existing concern that electronic records are highly prone to manipulation, making the authentication process exponentially more difficult. The transition to a "post-truth" evidentiary environment has severely complicated the authentication of digital records. Traditionally, Indian courts relied on Sections 65A and 65B of the Indian Evidence Act (now replaced by the Bharatiya Sakshya Adhiniyam, 2023) to admit electronic evidence. Certification Requirements: Admissibility for secondary evidence (copies or printouts) necessitates a certificate of authenticity. Under the new BSA framework, these certificates now require a "hash value"—a digital fingerprint—to ensure the evidence has not been tampered with. The Deepfake Dilemma: Current laws were not fully prepared for AI-generated manipulations that can create convincing but entirely fabricated audio-visual content. Proving that an individual in a fake image is not human is a strenuous challenge, as deepfakes can be virtually indistinguishable from reality without advanced forensic tools.

The "Liar’s Dividend" Problem

A critical social and legal side effect of deepfakes is the "liar's dividend." This phenomenon occurs when the mere existence of deepfakes allows malicious actors to cast doubt on genuine, incriminating evidence by falsely claiming it is AI-generated. Plausible Deniability: Defendants can now invoke the "deepfake defence" even against legitimate audio-visual proof. This shifts the evidentiary burden onto the complainant and can lead to prolonged forensic inquiries and delayed justice. Erosion of Trust: The constant threat of AI manipulation leads to a post-truth environment where people are more likely to dismiss objective facts in favour of emotional or biased narratives. A significant procedural challenge is the "Liar’s Dividend," a phenomenon where the mere existence of deepfakes allows malicious actors to cast doubt on genuine evidence. Plausible Deniability: Defendants can escape culpability by falsely claiming that authentic incriminating evidence is AI-generated. Shifting Burden: This skepticism shifts the evidentiary burden onto the complainant, who must now prove that real evidence is not a

deepfake. Over time, this erodes the presumptive credibility of all audio-visual evidence in court.

Forensic Detection and Expert Testimony

To combat digital fabrications, courts must increasingly rely on expert testimony and forensic analysis. Forensic Techniques: Experts examine pixel-level inconsistencies, abnormal lighting patterns, compression artifacts, and biometric anomalies (such as irregular facial movements or heart rate). Limitations: Advanced AI systems can generate synthetic content that lacks traditional forensic markers, like editing traces or metadata discrepancies, rendering conventional forensic methods less effective. As deepfakes enter the courtroom, the role of forensic experts becomes paramount. However, several gaps remain: Detection Methods: Experts utilize advanced tools like metadata analysis, biometric inconsistencies (e.g., micro-expression or heartbeat fluctuation analysis), and AI-based detection algorithms. Lack of Standardization: There is currently a lack of universal technical standards across forensic laboratories. Establishing these minimum technical requirements is essential to prevent the wrongful admission or rejection of evidence.

Chain of Custody for AI-Generated Content

Maintaining a rigorous chain of custody is vital for ensuring the integrity of digital evidence. Documentation: Every step—from collection and storage to transfer—must be documented to show the evidence has not been altered or contaminated. Forensic Tools: The use of certified forensic software is necessary to prevent accidental alterations during the investigation process. Maintaining a rigorous chain of custody is vital for the integrity of digital evidence. Documentation: Documentation must be complete from the moment data exits the original device through all cloud storage and messaging platforms. Vulnerability: Any break in this chain can be used by the defence to claim the material was replaced by a deepfake or underwent unauthorized changes. Forensic analysts warn that even they cannot always trace the chain of custody for deepfakes reliably.

Cross-Border and Jurisdictional Challenges

Investigating deepfake- related crimes frequently involves cross-border jurisdictional issues. Server Locations: Because many digital platforms operate internationally, investigative

authorities often face difficulties when servers or platform operators are located outside their domestic jurisdiction. Deepfake crimes often transcend national boundaries, creating significant jurisdictional conflicts: Server Locations: Digital evidence is often stored on servers in multiple foreign countries, requiring external access that is hindered by varying regulatory frameworks. Sovereignty and Legality: Direct implementation of cross-border forensics without judicial assistance can infringe on the sovereignty of other nations, potentially rendering the evidence inadmissible. Procedural Delays: Current mutual legal assistance treaties (MLATs) are often too slow to handle the volatile nature of electronic evidence, which can be deleted or moved instantly.

Enforcement Issues

A lack of international treaties and standardization of forensic procedures makes it difficult to enforce laws against foreign platforms that may refuse to cooperate. In summary, while frameworks like the BSA provide a basis for electronic evidence, they are currently strained by the rapid evolution of deepfake technology. A multidisciplinary approach—combining legal reform, standardized forensic protocols, and international cooperation—is essential to protect the integrity of the judicial process in a synthetic media environment.

Conclusion

The analysis undertaken in this paper set out to examine whether existing criminal law doctrine, comparative statutory frameworks, judicial interpretation, and evidentiary procedure are collectively capable of responding to the harms generated by deepfake technology. The findings suggest a qualified answer: the criminal law has proved remarkably adaptable at the margins, yet remains structurally unequipped to address deepfake-enabled harm as a coherent category of offending. At the doctrinal level, concepts such as actus reus, mens rea, causation, and modes of participation can be stretched to reach deepfake conduct, but only through considerable interpretive strain the fragmentation of criminal conduct across creators, disseminators, and hosting platforms resists the criminal law's traditional assumption of a unitary wrongdoer, forcing courts and prosecutors to calibrate fault along a gradient without clear statutory guidance on where these thresholds should lie. Liability for platforms remains anchored in fault-based, safe-harbour-conditional standards, while liability for AI developers and deployers is governed by analogy to dual-use technology doctrine rather than any settled principle specific to generative AI; and the theoretical inquiry into the "AI as offender" problem

confirmed that criminal responsibility must, for the foreseeable future, remain tethered to human and corporate actors through doctrines such as innocent agency, since no jurisdiction currently recognises the AI system itself as a bearer of criminal fault.

The comparative survey of statutory frameworks revealed a global legal order in transition, moving unevenly from reliance on fragmented, pre-existing law toward technology-specific criminal statutes. India's 2026 IT Rules amendments, the United States' TAKE IT DOWN and DEFIANCE Acts, the European Union's AI Act transparency mandates, the United Kingdom's Online Safety Act provisions, and the more aggressive criminalisation models adopted in China and South Korea together demonstrate that no single regulatory philosophy has yet emerged as dominant, though a common asymmetry persists: non-consensual intimate imagery attracts comparatively swift and uncontroversial criminalisation, while political and expressive synthetic content continues to generate constitutional friction. This pattern was reinforced by the examination of judicial responses, which showed genuine ingenuity — Indian courts have extended personality-rights and privacy jurisprudence to synthetic media and fashioned procedural innovations such as relief against unidentified "John Doe" defendants — but almost entirely within civil, injunctive proceedings. The near-total absence of reported criminal appellate precedent interpreting Sections 66C, 66D, and 67A specifically in the deepfake context represents perhaps the most significant gap identified in this study, leaving prosecutors, trial courts, and victims without authoritative guidance on how the doctrines discussed earlier actually operate in practice. Compounding this, the evidentiary analysis demonstrated that even a well-designed liability framework cannot function without a parallel evidentiary infrastructure: the authentication requirements under the Bharatiya Sakshya Adhiniyam were not designed to adjudicate contested claims of AI-generated fabrication, the "liar's dividend" threatens to undermine confidence in genuine evidence, and the absence of standardised forensic protocols, unresolved chain-of-custody questions, and slow-moving cross-border cooperation mechanisms compound the difficulty of prosecuting these offences.

Taken together, these findings support the central argument of this paper: criminal law's capacity to respond to deepfake-enabled harm cannot be assessed at any single level in isolation. A jurisdiction may possess doctrinally sound liability principles, a modern statutory framework, and an active judiciary, and still fail to deliver effective criminal justice if its evidentiary infrastructure cannot reliably distinguish the authentic from the synthetic — just as the most sophisticated evidentiary standards will prove meaningless without settled doctrine to

determine who, along the chain of creation and dissemination, ought to bear criminal responsibility in the first place. What emerges is a case not for a single, sweeping deepfake statute, but for a coordinated reform agenda: graded, actor-specific liability provisions that expressly contemplate AI-generated content, harmonised international cooperation to close jurisdictional gaps, and sustained investment in forensic and judicial capacity to ensure that doctrine and statute can be given genuine effect in the courtroom. Until these three strands doctrine, law, and proof are developed in tandem, the criminal law's response to deepfakes will remain, as this paper has shown, adaptive but incomplete.