Journal of International Research and Multidisciplinary Innovation

Synthetic Deception and the Erosion of Evidentiary Trust: Deepfake Technologies, Identity Crimes, and the Crisis of Legal and Democratic Institutions

Author: Dr. Praveen rai
Affiliation: Allahabad University, Prayagraj
Email: praveenrai0059@gmail.com
Submitted to: Journal of International Research & Multidisciplinary Innovation (JIRMI)
Date: 21 March 2026

Abstract

The rapid maturation of generative adversarial network (GAN) architectures and diffusion-based synthesis models has produced a class of synthetic media—commonly designated as deepfakes—capable of fabricating audiovisual representations of real individuals with a degree of perceptual fidelity that increasingly exceeds unaided human detection. This paper examines deepfake technologies as a systemic threat to the legal, political, and social infrastructure of identity verification, situating the phenomenon within the broader scholarly frameworks of criminology, political science, and information law. Drawing upon forensic data, legislative records, judicial proceedings, and peer-reviewed technical literature, the paper argues that deepfakes do not merely constitute a novel instrument of individual harm—encompassing non-consensual intimate image abuse, financial fraud, and targeted defamation—but represent a structural challenge to the evidentiary foundations of criminal justice, the epistemic integrity of democratic deliberation, and the philosophical coherence of legal identity as a stable category. The paper reviews the existing literature on synthetic media crime, interrogates the adequacy of current legal responses in major jurisdictions, and proposes an analytical framework for understanding the governance challenges

posed by identity crimes in the synthetic media era. The analysis concludes that effective governance of deepfake-enabled identity crimes requires a multi-layered regulatory architecture integrating technical detection standards, platform liability reform, international evidentiary law harmonization, and the development of robust digital identity infrastructure—none of which, as currently constituted, is adequate to the scale or velocity of the technological threat.

Keywords: deepfakes, synthetic media, identity crimes, generative adversarial networks, non-consensual intimate imagery, electoral disinformation, evidentiary law, digital identity, platform governance

1. Introduction

On 4 February 2024, robocalls impersonating the voice of United States President Joseph Biden were distributed to Democratic primary voters in New Hampshire, instructing them not to vote in the upcoming primary election. The calls, subsequently traced to a political operative using an AI voice-cloning service, were sufficiently realistic to prompt the Federal Communications Commission to issue an emergency ruling within weeks, declaring AI-generated voice content in political communications subject to existing robocall prohibitions (Federal Communications Commission, 2024). The incident, while ultimately limited in its electoral impact, crystallized a set of concerns about synthetic media that scholars, policymakers, and legal theorists had been advancing with increasing urgency since approximately 2017—the year in which the term deepfake first entered public discourse following the publication of synthetic celebrity videos on the platform Reddit.

The technological substrate of the deepfake phenomenon has evolved with extraordinary rapidity. What began as a computationally intensive process requiring weeks of training on consumer graphics processing units has, within a period of approximately seven years, become accessible through commercial services—including Synthesia, HeyGen, ElevenLabs, and numerous less reputable alternatives—that enable the production of convincing synthetic video or

audio content within minutes, without specialized technical expertise, and at minimal cost. A 2023 report by the cybersecurity firm Sumsub documented a 704 percent annual increase in deepfake incidents detected on its identity verification platform between 2022 and 2023, with particular concentrations in the financial services and cryptocurrency sectors (Sumsub, 2023). The deeptrace laboratory, subsequently acquired by Sensity AI, had earlier estimated that 96 percent of deepfake videos circulating online in 2019 constituted non-consensual synthetic pornography, of which 99 percent depicted women—a distribution that foregrounds the gendered dimensions of the harm and situates deepfake crimes within the broader political economy of technology-facilitated gender-based violence (Sensity AI, 2019).

The political science and criminological literatures have been slower than the technical literature to engage systematically with the deepfake phenomenon, and the present paper is intended as a contribution to closing that gap. The analysis proceeds through four substantive sections: a review of the existing literature; a methodological account; an empirical and theoretical analysis organized around three primary dimensions of harm—individual identity crimes, institutional and evidentiary harms, and democratic and political harms; and a conclusion that assesses the adequacy of existing governance frameworks and identifies the structural reforms required for an effective regulatory response. Throughout, the paper resists the temptation to treat deepfake crimes as an exceptional or unprecedented category of harm—they are better understood as an intensification and technological amplification of long-standing patterns of identity-based fraud, sexual image abuse, and political disinformation—while insisting on the structural novelty of the synthetic media environment as a governance challenge.

2. Literature Review

The scholarly literature on deepfakes and identity crimes is relatively recent but has expanded rapidly across multiple disciplines. The earliest systematic academic engagement with the phenomenon emerged from computer science and media studies. The foundational technical paper by Goodfellow et al. (2014), which introduced the generative adversarial network architecture, did not anticipate the misuse applications that would follow, but subsequent technical

literature has extensively documented the capability trajectory of synthetic media generation (Goodfellow et al., 2014). The work of Tolosana et al. (2020) provides a comprehensive taxonomy of deepfake generation techniques, distinguishing between face-swap methods, face reenactment, face synthesis, and voice conversion, and surveys the detection methodologies developed in response—a literature that has itself become a subject of an arms-race dynamic as detection methods are incorporated into the training processes of new generation models.

The legal and criminological literature has organized itself around several distinct dimensions of harm. The non-consensual intimate imagery (NCII) dimension has been most extensively documented. Franks (2017), whose work on revenge pornography law significantly shaped legislative responses in the United States and United Kingdom, has argued that the extension of synthetic image generation to intimate contexts constitutes a qualitative as well as quantitative escalation of NCII harms, because it severs the connection between the existence of genuine intimate images and the capacity to victimize—meaning that individuals with no history of intimate image sharing are now equally vulnerable to synthetic image-based abuse. This observation has significant implications for the legal frameworks governing NCII, most of which were drafted on the assumption that the images in question were authentic records of actual events.

The electoral and political disinformation dimension has attracted attention from political scientists working in the traditions of media studies, democratic theory, and security studies. Chesney and Citron (2019), in what remains the most cited legal-academic treatment of the subject, introduced the concept of the liar's dividend—the observation that the mere existence of deepfake technology enables bad actors to plausibly deny authentic evidence by claiming it has been fabricated, thereby undermining the evidentiary value of genuine audiovisual documentation even in cases where no synthetic media is actually deployed. This paradox—that the harm of deepfakes extends beyond instances of actual fabrication to contaminate the entire evidentiary environment—represents the most analytically challenging dimension of the phenomenon and has prompted significant engagement from evidence law scholars, including Mnookin (2017) and Delfino (2021).

The financial crime dimensions of synthetic identity fraud have been examined primarily in the criminological and financial regulation literature. Riek (2023) documents the use of AI-generated synthetic identities—combining fabricated biographical data with synthetic facial images to create entirely fictitious persons—to perpetrate large-scale financial fraud, including the opening of fraudulent bank accounts, the filing of false tax returns, and the exploitation of government benefit programs. The United States Government Accountability Office estimated losses from synthetic identity fraud in the American financial system at approximately $6 billion annually as of 2022, making it the fastest-growing category of financial crime in that jurisdiction (United States Government Accountability Office, 2022). The literature on financial deepfake crime intersects with a broader body of work on the political economy of cybercrime (Lusthaus, 2018), which situates individual criminal acts within organized networks of criminal enterprise that increasingly operate across jurisdictional boundaries, complicating both detection and prosecution.

A notable gap in the existing literature concerns the differential vulnerability of populations to deepfake-enabled identity crimes. With the partial exception of the gender analysis pioneered by Franks and others in the NCII context, the literature has been relatively inattentive to the ways in which race, class, geopolitical location, and prior marginalization by state institutions shape both exposure to synthetic media harm and access to remediation through legal and technical means. This gap is addressed, at least partially, in the present analysis.

3. Methodology

This paper employs a multi-method qualitative research design drawing upon legal analysis, critical discourse analysis, and interpretive political science. The epistemological foundation is broadly critical realist: the paper treats deepfake-enabled identity crimes as real phenomena with material effects on individuals and institutions, while recognizing that the categories through which these phenomena are understood, named, and governed are socially and politically constructed and subject to contestation. This orientation is particularly important in a

domain where the boundary between authentic and synthetic media is itself technologically unstable and contested.

Primary data sources include legislative texts and committee reports from the United States Congress, the United Kingdom Parliament, the European Parliament, and the Indian Ministry of Electronics and Information Technology; judicial decisions from American, British, and European courts addressing synthetic media-related claims; enforcement actions by financial regulators, including the United States Federal Trade Commission and the United Kingdom Financial Conduct Authority; and technical documentation from cybersecurity firms including Sensity AI, Sumsub, and the Stanford Internet Observatory. Quantitative data on deepfake incident rates, financial fraud losses, and detection accuracy are drawn from these primary sources and supplemented by peer-reviewed technical literature.

The analytical approach combines doctrinal legal analysis—assessing the fit between existing legal categories and the novel harms presented by deepfake technologies—with a political economy analysis that situates the deepfake phenomenon within the broader structural conditions of platform capitalism, the commercial incentive structures of technology development, and the international governance deficits that allow synthetic media harms to proliferate across jurisdictional boundaries. The paper does not attempt to predict specific future developments in either deepfake generation technology or its governance, given the pace of change in both domains, but rather to identify structural patterns and governance challenges that are likely to persist across a range of technological trajectories.

4. Analysis

4.1 Individual Identity Crimes: Fraud, Sexual Abuse, and Targeted Defamation

The most immediately documented and statistically measurable harms of deepfake technologies are those directed against specific individuals. These fall into three principal categories: financial fraud and synthetic identity crime; non-consensual intimate imagery; and targeted defamation and reputational harm. Each category raises distinct legal questions,

implicates different regulatory frameworks, and affects different demographic populations with different degrees of severity.

In the domain of financial fraud, the combination of deepfake video and voice synthesis with synthetic identity documents has enabled a new generation of identity crimes that defeat traditional verification procedures. In February 2024, a multinational company based in Hong Kong lost approximately $25 million after an employee was deceived into authorizing wire transfers during a video conference call in which all other participants—including the company's chief financial officer—were AI-generated deepfakes, convincingly mimicking real colleagues whose faces and voices had been synthesized from publicly available video material (CNN, 2024). This case, widely reported, is illustrative of a broader pattern: the 2023 Sumsub report documented that deepfake fraud attempts in financial services increased by 3,000 percent globally between 2022 and 2023, with the highest rates recorded in North America, Western Europe, and the Asia-Pacific region. The United States Federal Trade Commission reported that impersonation fraud—a category that increasingly encompasses AI-voice and synthetic image fraud—resulted in losses of $2.7 billion to American consumers in 2023 alone (Federal Trade Commission, 2024).

The non-consensual intimate imagery dimension represents perhaps the most pervasive and deeply harmful individual application of deepfake technology. As documented by Sensity AI (2019), the earliest and numerically dominant use of deepfake technology was the non-consensual production and distribution of synthetic pornographic images and videos depicting real individuals—overwhelmingly women—without their knowledge or consent. The passage of the Online Safety Act in the United Kingdom in 2023 and the SHIELD Act provisions in various American states have created criminal liability for the production and distribution of NCII, but the application of these frameworks to synthetic images has proven legally complex. Key issues include the question of whether synthetic NCII, which does not involve the unauthorized disclosure of genuinely captured images, constitutes the same category of harm as authentic NCII; whether criminal liability requires proof of intent to cause distress; and whether platform operators bear criminal or civil liability for hosting synthetic NCII generated by their users. In 2023, the image of a prominent American sportscaster was the subject of a widely publicized synthetic NCII

incident; the episode prompted Senate hearings and renewed legislative attention to the adequacy of existing frameworks, which most witnesses described as insufficient (United States Senate Committee on the Judiciary, 2024).

Targeted defamation through deepfakes occupies a legally complex position, because the existing law of defamation in most common law jurisdictions was developed to address verbal and written false statements rather than audiovisual fabrications. The specific challenge deepfakes present to defamation law is twofold. First, the perceptual realism of synthetic video creates a qualitatively different epistemic situation from false verbal assertions: human cognition is neurologically predisposed to assign greater credibility to audiovisual evidence than to verbal claims, a disposition that synthetic media exploits directly. Second, the speed of viral distribution on social media platforms means that deepfake defamation typically achieves mass reach before detection or rebuttal is possible—a temporal asymmetry that undermines the tort law assumption that the harm of defamation can be remediated by correction and retraction.

4.2 Institutional and Evidentiary Harms: The Crisis of Authenticated Truth

Beyond the harms to specific individuals, deepfake technologies present a systemic challenge to the institutional frameworks through which modern societies establish and verify facts. The most consequential of these challenges operates in the domain of evidentiary law—the rules governing what counts as proof in judicial proceedings. The use of audiovisual recordings as evidence in criminal and civil proceedings is relatively recent in historical terms but has become foundational to the functioning of investigative and prosecutorial processes: surveillance camera footage, bodycam recordings, digital communications, and audiovisual witness records all depend for their evidentiary authority on the assumption that digital media is a reliable index of real events.

The liar's dividend—Chesney and Citron's (2019) term for the defensive use of deepfake possibility to discredit authentic evidence—has already appeared in documented legal proceedings. In a 2022 child custody dispute in the state of Arkansas, a litigant's attorney argued that a video recording submitted as evidence of the client's misconduct might be a deepfake, successfully introducing sufficient doubt to complicate the proceeding despite the absence of any

technical evidence supporting the fabrication claim (Delfino, 2021). While this particular gambit was ultimately unsuccessful, it illustrates the broader structural vulnerability: as awareness of deepfake capabilities diffuses through the population, courts, juries, and administrative adjudicators will face increasing uncertainty about the authenticity of audiovisual evidence, a development with potentially far-reaching consequences for the administration of justice.

The forensic response to this challenge has developed along two primary lines: technical deepfake detection systems, and cryptographic content provenance frameworks. Technical detection systems—employing convolutional neural networks trained to identify the characteristic artifacts of GAN-generated or diffusion-generated imagery—have achieved high accuracy rates in controlled laboratory conditions but perform significantly less reliably on real-world content, particularly material that has been compressed or re-encoded through social media platforms. A 2022 benchmark study published in IEEE Transactions on Information Forensics and Security found that the leading detection algorithms achieved accuracy rates of 91 percent on benchmark test sets but fell to between 65 and 73 percent accuracy on content that had undergone standard social media compression (Rossler et al., as cited in Tolosana et al., 2020). The content provenance approach, championed by the Coalition for Content Provenance and Authenticity (C2PA)—a consortium including Adobe, Microsoft, the BBC, and others—proposes cryptographic signing of media at the point of capture to establish a verifiable chain of custody. While technically promising, this approach requires adoption across the full ecosystem of capture devices, editing tools, and distribution platforms, and provides no retrospective protection for the vast archive of existing media.

4.3 Democratic and Political Harms: Synthetic Media and Electoral Integrity

The implications of deepfake technologies for democratic politics represent the dimension of the phenomenon that has attracted the greatest legislative attention, though not necessarily the most sophisticated analytical treatment. The concern, at its most basic, is that synthetic audiovisual fabrications of political figures could be deployed to manipulate electoral outcomes—by falsely depicting candidates making disqualifying statements, by fabricating evidence of criminal

conduct, or by manufacturing crises that alter the dynamics of electoral competition. The New Hampshire Biden robocall incident described in this paper's introduction represents one documented instance of this class of harm, though at a relatively limited scale.

The more structurally significant democratic harm may operate through the mechanism of epistemic corrosion rather than specific fabrication events. Research in political psychology has demonstrated that exposure to false information produces persistent belief distortion even after explicit correction—a phenomenon known as the continued influence effect (Ecker et al., 2022). If synthetic media events—whether or not they involve actual deepfakes in specific instances—generate generalized doubt about the authenticity of political communication, the aggregate effect on democratic deliberation could be substantial, eroding the shared epistemic ground upon which political argument and electoral choice depend. This concern connects to a broader body of scholarship on disinformation and democratic resilience, including the work of Wardle and Derakhshan (2017) on information disorder and Runciman (2018) on the vulnerability of democratic institutions to technological disruption.

Legislative responses to the electoral deepfake problem have emerged in several jurisdictions but remain fragmented and inconsistent. California enacted AB 602 and AB 730 in 2019, prohibiting the distribution of materially deceptive synthetic media depicting candidates within 60 days of an election. Texas, Georgia, and Virginia enacted analogous provisions in subsequent legislative sessions. At the federal level in the United States, the Deepfake Task Force Act and the DEEPFAKES Accountability Act were introduced in successive Congresses but had not been enacted as of early 2024 (United States Congress, 2023). The European Union's Artificial Intelligence Act, finalized in 2024, includes provisions requiring disclosure of AI-generated content but does not establish a specific prohibition on electoral deepfakes. India, which conducted the world's largest democratic election in 2024, issued emergency advisory guidelines on AI-generated electoral content but lacked a comprehensive statutory framework. The common limitation of existing legislative approaches is that they address the supply of deepfake content at the point of production or distribution without adequately addressing the demand dynamics,

platform amplification mechanisms, and international production and distribution routes through which synthetic electoral disinformation actually reaches voters.

4.4 Governance Frameworks and Their Adequacy

The governance challenges presented by deepfake-enabled identity crimes can be organized around three structural deficits: the pace gap between technological capability and regulatory response; the jurisdictional gap between the international distribution of synthetic media harm and the national scope of regulatory authority; and the evidentiary gap between the technical capabilities required for deepfake detection and the resources available to law enforcement and judicial institutions.

The pace gap is illustrated by the trajectory of NCII legislation. The first state-level non-consensual pornography statutes were enacted in the United States in 2013—more than a decade after the problem emerged and several years after academic researchers had documented its harms. Synthetic NCII—which is technically distinct from non-consensual disclosure of authentic intimate images and may not fall within the scope of existing statutes—emerged as a significant harm around 2017, yet comprehensive legislative responses addressing the synthetic dimension were still being developed in most jurisdictions as of 2024. Given that deepfake generation capabilities are improving at a rate that makes even currently available commercial services dramatically more capable than those available two years ago, legislative processes operating on electoral cycles are structurally ill-suited to provide timely regulatory responses.

The jurisdictional gap is equally acute. The production of synthetic media content is geographically unbounded: a deepfake video targeting a political candidate in one country can be produced in another, distributed through servers in a third, and amplified by platforms incorporated in a fourth. Existing frameworks for mutual legal assistance in criminal matters—the primary mechanism for cross-border law enforcement cooperation—are slow, cumbersome, and were not designed for the velocity and volume of digital evidence. The Budapest Convention on Cybercrime, which provides the most developed international framework for cooperation in digital

crime, has been signed by 68 states but not ratified by several major sources of synthetic media harm, and its provisions do not specifically address deepfake content.

Platform liability frameworks represent the most contested dimension of deepfake governance. In the United States, Section 230 of the Communications Decency Act immunizes platform operators from civil liability for user-generated content, including synthetic media produced and distributed by users—a protection that critics argue removes the primary economic incentive for platforms to invest in detection and removal infrastructure. The European Union's Digital Services Act (DSA), which came into force for large platforms in August 2023, establishes a more demanding regime requiring designated Very Large Online Platforms to conduct systemic risk assessments addressing synthetic media and to implement proportionate mitigation measures—a framework that, if effectively enforced, could significantly alter the incentive structure for platform content governance. The DSA's extraterritorial application to non-European platforms serving European users means that its provisions could function as a de facto global standard, analogous to the role played by the GDPR in data protection—though early enforcement experience suggests that the DSA's implementation will be contested by major platform operators.

5. Conclusion

This paper has argued that deepfake technologies present a set of governance challenges that are structural rather than merely technical in character—challenges that arise not from the particular capabilities of current synthesis systems but from the fundamental incompatibility between the pace and scale of synthetic media generation and the institutional frameworks—legal, political, epistemic—through which modern societies organize the verification of identity, the authentication of evidence, and the conduct of democratic deliberation. The individual harms of deepfake-enabled identity crimes are real, measurable, and disproportionately borne by women, marginalized communities, and political actors in high-information environments; the institutional and democratic harms, while less immediately quantifiable, are potentially more consequential in the long term.

The existing regulatory landscape—comprising a patchwork of state-level criminal statutes, nascent federal proposals, platform self-regulatory commitments, and international frameworks of limited reach—is inadequate to these challenges. Effective governance would require, at minimum, the development of mandatory technical standards for synthetic media disclosure at the point of production; a reformed platform liability framework that creates meaningful incentives for investment in detection and removal without producing overbroad censorship; harmonized international legal instruments addressing the cross-border production and distribution of harmful synthetic content; and the development of public digital identity infrastructure that provides individuals with cryptographically verifiable means of authenticating their communications.

The deeper challenge is epistemic rather than regulatory. Deepfake technologies are not merely a new instrument of crime; they constitute a systematic attack on the social epistemology through which individuals, institutions, and polities establish shared understandings of what is real and what has occurred. Addressing this challenge requires not only legal and technical innovation but sustained investment in the media literacy, institutional trust, and epistemic culture upon which democratic societies ultimately depend. The history of information technology suggests that these cultural and institutional adaptations typically lag far behind the technological developments that necessitate them. Whether the pace of that adaptation can be accelerated to match the velocity of synthetic media development is the central political question of the deepfake era.

References

1. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review,* 107(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

2. Delfino, R. A. (2021). Pornographic deepfakes: The case for federal criminalization of revenge porn's next tragic chapter. *Fordham Law Review,* 88(3), 887–938.

3. Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology,* 1(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

4. Federal Communications Commission. (2024). *FCC rules AI-generated voices in robocalls illegal (News Release DA 24-137).* https://www.fcc.gov/document/fcc-rules-ai-generated-voices-robocalls-illegal

5. Federal Trade Commission. (2024). *Consumer sentinel network data book 2023.* FTC. https://www.ftc.gov/reports/consumer-sentinel-network-data-book-2023

6. Franks, M. A. (2017). Revenge porn reform: A view from the front lines. *Florida Law Review,* 69(5), 1251–1337.

7. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems,* 27, 2672–2680.

8. Lusthaus, J. (2018). *Industry of anonymity: Inside the business of cybercrime.* Harvard University Press.

9. Sensity AI. (2019). *The state of deepfakes: Landscape, threats, and impact.* Sensity Research Report. https://sensity.ai/reports

10. Sumsub. (2023). *Identity fraud report 2023: The state of fraud in the digital world.* Sumsub Research. https://sumsub.com/research/identity-fraud-report-2023

11. Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion,* 64, 131–148. https://doi.org/10.1016/j.inffus.2020.06.014

12. United States Government Accountability Office. (2022). *Synthetic identity fraud: FinCEN should enhance its support for law enforcement and industry partners (GAO-22-104467).* GAO. https://www.gao.gov/products/gao-22-104467