International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Deepfake Technology as a New Tool for Criminal Offenses: Legal Challenges and Its Way Forward in Criminal Law

Akanksha Wadhawan

Amity Institute of Advanced Legal Studies (AIALS), Amity University, Noida, Uttar Pradesh

Abstract:
Deepfake technology, driven by advances in artificial intelligence and deep learning, enables the creation of hyper-realistic audio, video, and image content that can convincingly mimic real individuals. Originally developed for entertainment, art, and satire, deepfakes have rapidly evolved into a potent tool for malicious actors. Their increasing accessibility and sophistication have led to a surge in criminal activities, including identity theft, fraud, revenge porn, political misinformation, and extortion. This paper explores the growing threat posed by deepfake technology in the realm of cybercrime and real-world offenses. It aims to analyze the technical foundation of deepfakes, their application in various forms of criminal behavior, and the legal, ethical, and societal challenges they present. The study also includes case analyses that demonstrate the real-life impact of deepfake misuse and evaluates the effectiveness of current detection tools and regulatory frameworks. The findings highlight the urgent need for a multi-layered response involving robust legal reforms, enhanced detection technologies, and increased public awareness. Recommendations are provided for policymakers, law enforcement, and digital platforms to counteract the malicious use of this evolving technology while preserving its legitimate applications.

Keywords: deepfake, cybercrime, misinformation, identity theft, legal regulation.

Introduction:

Deepfake technology refers to synthetic media—primarily video, audio, and images—that have been digitally altered or entirely generated using artificial intelligence to make it appear as though someone is doing or saying something they never did. The term "deepfake" is derived from "deep learning," a subset of machine learning in artificial intelligence, which enables computers to process data in a way that mimics the neural networks of the human brain. The development of Generative Adversarial Networks (GANs) in 2014 marked a significant breakthrough in the creation of deepfakes. GANs function through two neural networks—the generator and the discriminator—that work against each other to improve the realism of the synthetic content produced[1] (Goodfellow et al., 2014).

Initially, deepfake technology found application in benign fields such as film production, satire, gaming, and even voice reconstruction for medical purposes. However, as the technology became more refined and accessible through open-source tools, it began to be exploited for malicious purposes. Today, deepfakes are increasingly used to commit crimes ranging from identity theft, financial fraud, and revenge

[1] Ian Goodfellow et al, "Generative Adversarial Networks" (2014) https://arxiv.org/abs/1406.2661

porn to political disinformation and social manipulation. The threat is no longer speculative—it is real, immediate, and growing at an alarming rate[2] (Chesney & Citron, 2019).

In the digital era, where the boundaries between real and artificial content are increasingly blurred, the proliferation of deepfake technology raises serious concerns. It not only endangers individuals' privacy and reputations but also threatens democratic processes, judicial integrity, and national security. With fake content becoming indistinguishable from reality, even legitimate media can be dismissed as fake, leading to a phenomenon known as the "liar's dividend."

This paper seeks to address the following research questions: How is deepfake technology currently being misused in criminal contexts? What are the legal, ethical, and societal implications of this misuse? And finally, what frameworks—both technological and regulatory—can be developed to detect, prevent, and respond to the criminal use of deepfakes? By exploring these questions, this study aims to contribute to the growing body of research on digital harm and to propose actionable solutions that balance innovation with accountability.

Understanding Deepfakes:

Deepfakes are created through a sophisticated process involving artificial intelligence, particularly a machine learning technique known as Generative Adversarial Networks (GANs). Introduced by Ian Goodfellow in 2014, GANs consist of two neural networks—a generator and a discriminator—that operate in tandem. The generator creates synthetic data (such as a fake video or image), while the discriminator evaluates its authenticity. Through constant feedback and iteration, the generator improves its outputs until they become nearly indistinguishable from real data (Goodfellow et al., 2014).

The scope of deepfakes extends beyond video manipulation. They can be classified into four major categories: video**,** audio**,** image, and text. Video deepfakes are perhaps the most well-known, where a person's face is altered to say or do things they never did. Audio deepfakes can clone voices with stunning accuracy, often used in scams and impersonation. Image-based deepfakes involve the insertion or replacement of faces in photographs, commonly used in non-consensual pornography or identity theft. Meanwhile, text-based deepfakes—though still developing—use natural language models to produce fake conversations, emails, or even academic work.

The widespread availability of deepfake tools has contributed significantly to their rise. Open-source software like DeepFaceLab, Faceswap, and apps such as Zao and Reface allow users with minimal technical expertise to create convincing synthetic media. These tools, once limited to computer scientists and digital artists, are now freely available on GitHub and mobile app stores, making the creation of deepfakes highly democratized.

The increasing sophistication of deepfake technology is cause for concern. With improvements in computational power and AI algorithms, deepfakes are becoming harder to detect even with advanced forensic tools. This growing realism amplifies the potential for misuse in criminal, political, and social domains. While the technology itself is neutral, its implications depend heavily on the intent and context of its use. Understanding its technical foundation and diverse applications is critical for framing appropriate legal and regulatory responses.

[2]Bobby Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security" (2019) 107(1) California Law Review 175.

Criminal Applications of Deepfake Technology

Deepfake technology, while innovative in design, has rapidly become a tool for serious criminal offenses. Its ability to convincingly mimic human faces, voices, and behaviors gives wrongdoers the power to commit acts that bypass traditional security measures and exploit both individuals and institutions. Among its most concerning uses is identity theft and fraud. Criminals have used AI-generated voices and faces to impersonate executives and trick employees into transferring large sums of money. In a 2020 incident, fraudsters used deepfake audio to imitate a CEO and deceived an employee into wiring $243,000 (BBC News, 2019).[3] Such cases illustrate the growing vulnerability of voice and facial recognition systems in financial institutions.[4]

Equally disturbing is the use of deepfakes in revenge porn and sexual exploitation. Offenders superimpose an individual's face onto explicit material, creating videos and images that can destroy reputations and cause lasting trauma. Women are especially targeted—according to Deeptrace, 96% of deepfake videos online in 2019 were pornographic, and nearly all featured women (Ajder et al., 2019). Victims often face immense difficulty in removing such content or obtaining justice, due to jurisdictional and technological limitations.

Political disinformation represents another grave threat. Deepfakes are now used to spread misinformation and manipulate public opinion by showing politicians making false statements or behaving controversially. These can provoke unrest and destabilize democratic institutions. A widely shared 2018 deepfake of former U.S. President Barack Obama demonstrated how even fabricated content can blur the line between satire and misinformation, shaking public trust in audiovisual evidence.

In the corporate world, deepfakes are deployed for espionage and fraud. Impersonating key figures such as CEOs or CFOs, cybercriminals gain access to confidential data or authorize fraudulent financial transactions. There have also been reports of fake video calls used to mislead employees. Moreover, deepfakes can be used to damage a company's reputation or manipulate markets, presenting a new frontier in corporate sabotage.

Cyberbullying and extortion are further extensions of deepfake abuse. Offenders fabricate compromising media and threaten to release it unless their victims meet specific demands. While victims may prove the content is false, the reputational and psychological damage is often irreversible. Young individuals, particularly teenagers and public figures, are especially vulnerable to such emotional abuse.

Perhaps most dangerously, deepfakes have begun to infiltrate the legal domain. Manipulated audio and video are being passed off as authentic evidence in courtrooms, risking miscarriages of justice. Forensic experts often struggle to differentiate real from fake, placing immense pressure on judicial systems to develop better verification tools. These developments pose a threat not just to individuals, but to institutions upholding justice and truth.

In conclusion, deepfakes' adaptability across a range of criminal activities—identity theft, sexual exploitation, disinformation, corporate fraud, cyberbullying, and legal tampering—makes them uniquely threatening. Their realism and ease of creation make them not just tools of deception, but catalysts for far-reaching personal and societal harm.

[3] BBC News, "Fraudsters use AI to mimic CEO's voice in unusual cybercrime case" (30 August 2019) https://www.bbc.com/news/technology-48908736

[4] Deeptrace, Henry Ajder et al, "The State of Deepfakes: Landscape, Threats, and Impact" (2019) https://regmedia.co.uk/2019/10/07/deepfake_report.pdf

Case Studies: Real-Life Incidents of Deepfake Misuse

Deepfake technology has already moved beyond hypothetical threat to real-world exploitation, with multiple high-profile cases demonstrating its criminal potential. These case studies reveal the depth of damage that synthetic media can cause—economically, socially, and politically—and underscore the urgent need for regulatory and technological safeguards.

1. CEO Voice Spoofing Scam in the United Kingdom: In a landmark case in 2019, criminals used deepfake audio technology to impersonate the voice of a German CEO, instructing a UK-based energy firm's executive to transfer €220,000 (approx. $243,000) to a Hungarian supplier. The executive believed he was speaking with his boss, as the AI-generated voice mimicked the tone, accent, and even subtle inflections of the real CEO. The money was sent, only to be siphoned off across multiple countries and ultimately lost. Despite law enforcement involvement, the perpetrators remain unidentified, highlighting the cross-border complexity of deepfake crimes (BBC News, 2019). The case triggered widespread concern in the business world, prompting a reevaluation of voice-verification systems and internal security protocols.[5]

2. Fake Political Speech in Gabon Leading to Unrest: In 2019, political instability in Gabon escalated after the circulation of a suspicious video featuring President Ali Bongo, who had been absent from the public eye due to illness. The video was allegedly manipulated to show the president in better health than he actually was. The unnatural appearance of the footage sparked speculation that it was a deepfake, fueling rumors that the president was incapacitated or dead. This perceived deception led to a failed military coup, revealing how synthetic media can directly influence national security and governance (The Guardian, 2019). Though never confirmed to be a deepfake, the ambiguity alone was enough to provoke unrest, demonstrating the power of deepfake suspicion in politically fragile environments.[6]

3. Indian Actresses Targeted with Fake Pornographic Content: India has witnessed an alarming rise in deepfake pornography involving public figures. Bollywood actresses such as Rashmika Mandanna and Kajol have been targeted by manipulated videos where their faces were superimposed onto the bodies of adult film actors. These videos went viral on social media, causing widespread outrage and distress to the victims. Despite public statements by the actresses and action by law enforcement, the deepfakes continue to resurface across different platforms. The legal framework in India, including sections under the IT Act and IPC, has proven inadequate in tackling the virality and cross-platform proliferation of such content. These cases have intensified public debate about digital consent, cyber laws, and gendered harassment online (NDTV, 2023).[7]

4. Ukraine War Propaganda Videos: During the ongoing Russia-Ukraine conflict, deepfake videos have emerged as instruments of wartime propaganda. In one instance, a fake video of Ukrainian President Volodymyr Zelenskyy appeared online, falsely declaring Ukraine's surrender. Though quickly debunked, the video circulated widely before removal, with potential to confuse soldiers and citizens alike. It was reportedly spread through hacked news websites and social media platforms, amplifying its reach. The

[5] BBC News, "Fraudsters use AI to mimic CEO's voice in unusual cybercrime case" (30 August 2019) https://www.bbc.com/news/technology-48908736

[6] The Guardian, "Ali Bongo video deepfake allegations in Gabon" (9 January 2019) https://www.theguardian.com/world/2019/jan/09/gabon-coup-ali-bongo-video-speech-deepfake

[7] NDTV, "Rashmika Mandanna's Deepfake Video Sparks Concern Over AI Misuse" (10 November 2023) https://www.ndtv.com/india-news/rashmika-mandanna-deepfake-viral-video-prompts-call-for-stricter-laws-4538493

deepfake aimed to destabilize morale and influence international opinion. This incident showed how synthetic media could be deployed as psychological weapons in modern warfare, raising questions about information integrity during armed conflicts (Reuters, 2022).[8]

In each of these cases, the aftermath revealed severe gaps in legal preparedness, digital literacy, and crisis response. Victims—whether individuals, corporations, or entire nations—have struggled to respond in real-time to the damage. Public outrage often triggers short-term action, such as platform takedowns or investigations, but long-term legal and systemic remedies remain limited. Victims of deepfake pornography face lasting trauma and reputational harm, while political and corporate victims contend with trust erosion and economic consequences. These case studies collectively emphasize the urgent need for proactive frameworks that include digital evidence verification, legal reform, AI detection tools, and public awareness campaigns.

Legal and Ethical Challenges

As deepfake technology rapidly advances, legal and ethical frameworks struggle to keep pace, resulting in a grey zone that challenges law enforcement, courts, and civil society. The dual-use nature of deepfakes—capable of both artistic innovation and criminal manipulation—adds complexity to any effort at regulation. One of the primary legal hurdles is the absence of clear, comprehensive legislation specifically designed to address deepfakes. Most cyber laws around the world were formulated before the emergence of AI-generated synthetic media, rendering them insufficient for tackling the distinctive threats posed by deepfakes.

In India, while laws such as the Information Technology Act, 2000 and the Indian Penal Code, 1860 can be invoked for defamation, obscenity, or cyberstalking, they do not explicitly recognize or regulate deepfakes. The challenge is further complicated by jurisdictional issues, where deepfake content created in one state or country can be consumed in another, making enforcement difficult. In the United States, although states like California, Texas, and Virginia have enacted laws addressing deepfakes—such as California's ban on using deepfakes in political campaigns within 60 days of an election—there is no uniform federal law on the matter. The European Union, on the other hand, has adopted broader regulatory frameworks such as the Digital Services Act and the GDPR, which touch on misinformation and data manipulation. However, even these measures fall short of comprehensively addressing AI-generated content and its implications, particularly in cross-border criminal scenarios.

Ethically, deepfakes present serious concerns related to privacy and consent. The unauthorized use of an individual's likeness—whether image, voice, or mannerisms—can result in irreversible reputational damage, emotional trauma, and violations of personal dignity. Victims of deepfake pornography or impersonation often find themselves with limited or no legal remedy, especially in jurisdictions with poorly defined privacy protections. The absence of legal distinctions between synthetic and authentic content further enables perpetrators to act with impunity.

The regulation of deepfakes also raises a critical tension between the need to prevent harm and the imperative to protect freedom of expression. Satirical, parodic, or artistic deepfakes may fall under protected speech in democratic societies. Over-regulation risks unintended consequences such as

[8] Reuters, "Ukraine's President targeted in deepfake surrender video" (16 March 2022) https://www.reuters.com/technology/fact-check-video-zelenskiy-surrendering-ukraine-is-deepfake-2022-03-16/

censorship or misuse by authoritarian regimes to suppress dissent. Thus, any legal approach must balance the right to free speech with the necessity of curbing misinformation and abuse.

Courts are also grappling with forensic and evidentiary issues. As deepfakes grow more sophisticated, the authenticity of audio-visual evidence—a cornerstone of modern litigation—becomes increasingly questionable. If forged media can be passed off as genuine, the evidentiary reliability of digital content in both civil and criminal trials faces a serious threat.

Finally, enforcement remains a major challenge. Many law enforcement agencies lack the technical capacity and forensic tools needed to detect and trace deepfakes. Offenders often exploit VPNs, decentralized platforms, and anonymizing technologies to mask their identities. The rise of open-source deepfake tools has also democratized access, enabling even those with minimal technical skills to produce highly realistic fakes. This expansion of potential offenders puts an immense burden on already stretched investigative bodies.

In conclusion, the legal and ethical challenges surrounding deepfakes demand urgent, multi-layered responses. Laws must evolve to be both technologically informed and ethically robust. Moreover, global cooperation is essential, as the borderless nature of deepfake threats exceeds the capacity of any single legal system to manage effectively.

Policy Recommendations

As deepfake technology becomes increasingly sophisticated and accessible, policymakers must craft a comprehensive and forward-thinking framework to address its potential for misuse. The first step is the formulation of model laws that mandate transparency and accountability in the creation and dissemination of synthetic media. Such laws should require mandatory disclosures for AI-generated content, including clear labeling and digital watermarking to differentiate altered media from authentic sources. Additionally, strict criminal penalties must be introduced for malicious deepfakes—those created with the intent to defame, harass, commit fraud, incite violence, or manipulate legal evidence. These laws must strike a balance between protecting individual rights and preserving freedom of expression in satire or artistic works.

Deepfake threats transcend national borders, making international cooperation essential. A collaborative global framework is needed to facilitate the sharing of forensic data, streamline extradition for cybercriminals, and establish consistent evidentiary standards across jurisdictions. Organizations like INTERPOL, the United Nations, and regional cybercrime task forces can play pivotal roles in drafting treaties and operational protocols tailored to synthetic media crimes.

Simultaneously, digital platforms such as YouTube, Instagram, and TikTok must be held accountable for the content they host. Regulatory policies should compel these platforms to implement deepfake detection systems, provide users with content origin verification tools, and establish swift takedown procedures for harmful synthetic media. Regular transparency reports should also be mandated, documenting the volume of flagged content, the nature of violations, and enforcement outcomes.

Another key strategy involves fostering digital literacy among users. Public awareness campaigns should educate citizens on the nature of deepfakes, how to spot them, and the importance of verifying online information—especially during elections or socio-political events. These programs can be integrated into school curriculums, government advisories, and social media outreach efforts.

Lastly, there must be robust support systems for victims of deepfake abuse. This includes providing free legal assistance, fast-track complaint redressal mechanisms, psychological counseling, and financial

compensation in cases of monetary or reputational loss. Establishing centralized helplines and digital complaint portals can empower victims to report incidents quickly and seek protection.

By combining legal reform, technological intervention, public awareness, and victim support, governments and institutions can build a resilient defense against the criminal misuse of deep fakes.

Conclusion

The proliferation of deepfake technology has introduced a new dimension of threat in the digital age—one where seeing and hearing can no longer be equated with believing. From identity theft and financial fraud to political disinformation and revenge pornography, the malicious use of deepfakes has demonstrated the severe social, legal, and psychological consequences it can inflict. The ease with which this technology can now be accessed—via open-source tools and mobile apps—further amplifies the risks, making it a potent weapon in the hands of bad actors.

This rapidly evolving landscape necessitates urgent regulatory intervention and robust technological preparedness. Governments must act swiftly to formulate comprehensive legal frameworks that penalize misuse while still protecting artistic and creative expressions. At the same time, investment in deepfake detection tools and digital forensics is crucial to equip law enforcement and judicial systems with the ability to discern real from fake.

However, the pursuit of regulation must not stifle innovation. A careful balance must be maintained—ensuring that technological advancements in AI and machine learning continue to flourish, but within a framework that safeguards public trust, democratic processes, and individual dignity.

In the broader context of the AI era, the deepfake challenge serves as a critical test of our commitment to digital ethics. As we navigate this complex terrain, the goal should not only be to curb misuse but also to foster a culture of responsibility, transparency, and respect in the digital world.

References

1. **Information Technology Act, 2000**, No. 21, Acts of Parliament, 2000 (India).
2. **Indian Penal Code, 1860**, Act No. 45 of 1860.
3. California Legislative Information, **AB-730 Elections: deceptive audio or visual media**, 2019. Available at: https://leginfo.legislature.ca.gov
4. Deepfake Report, **"Deepfakes and Democracy: How Synthetic Media is Distorting Truth"**, Brookings Institution (2020). Available at: https://www.brookings.edu/research/the-deepfake-challenge
5. European Parliament and Council, **Regulation (EU) 2016/679** of 27 April 2016 (General Data Protection Regulation), Official Journal of the European Union, L 119/1, 4.5.2016.
6. European Parliament and Council, **Digital Services Act**, Regulation (EU) 2022/2065, Official Journal of the European Union, L 277/1, 27.10.2022.
7. Sharma, Neha, **"Legal Regulation of Deepfakes: Issues and Challenges in Indian Context"** (2023) 65(2) Journal of the Indian Law Institute 212.
8. Citron, Danielle K. and Chesney, Robert, **"Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security"** (2019) 107(7) California Law Review 1753.
9. Jain, R.K., **"Balancing Free Speech and Regulation of AI Content in India"** (2022) 59 Indian Bar Review 101.

10. Kumar, Anurag, **"Deepfakes, Evidence Law, and the Crisis of Authenticity in Indian Courts"** (2021) 63(3) Journal of Law and Technology 87.