Jus Corpus Law Journal

Open Access Law Journal – Copyright © 2025 – ISSN 2582-7820
Editor-in-Chief – Prof. (Dr.) Rhishikesh Dave; Publisher – Ayush Pandey

Deepfake: Digital Deception and Intermediaries' Accountability

Laksh Gera[a] Ananya Som Sharma[b]

[a]Institute of Law, Nirma University, Ahmedabad, India [b]Institute of Law, Nirma University, Ahmedabad, India

Received 15 November 2025; *Accepted* 15 December 2025; *Published* 20 December 2025

'Deepfake technology,' fueled by Generative Adversarial Networks, is becoming so adept at blurring the lines between reality and fiction that cases of moral, legal, and security emergencies are emerging as a result. This report traces how synthetic media went from being a niche technological experiment to a means for monetary fraud and, critically, a source for image-driven sex abuse. A disturbing increase in sextortion cases and their psychological toll on victims, particularly women and minorities, is documented, citing FBI statistics and a comment by a U.S. court on how such crimes take place 'clandestinely and across jurisdictions.' A significant part of the discourse situates these risks within the Indian legal system, focusing on the Information Technology Act of 2000. The author quizzes whether Sections 67 and 67A are effective enough for the regulation of obscene content and whether the "Safe Harbour" provision for intermediaries under Section 79 is sufficient. It is submitted that the conventional interpretation of the requirement for "actual knowledge" within Section 79(3) is inadequate, particularly for private encrypted groups. Through the evidence presented by the case of Neetu Singh v Telegram FZ LLC, the point is that the fundamental right of privacy cannot and should not become the reason for violating intellectual property rights, and that platforms need to monitor content proactively. In another study published by the Indian Journal of Law and Technology, the author assessed the constitutionality of Section 79(2) and whether subsequent Section 79(2) judgments comply with the Indian Constitution. The study brought into. Ultimately, the article urges a move from passive protection to active protection, a 'threat turned into a shield.' It suggests that Cyber Threat Intelligence (CTI), derived from Threat Intelligence Platforms (TIP), should be employed for active AI protection. One thing is sure: to restore trust, there has to be an evolved approach, lack of stricter legislation being one of them, that repels AI-powered threats.

Keywords: *deepfake technology, synthetic media, intermediary liability, safe harbour protection.*

INTRODUCTION

Deepfake technology evolves rapidly to merge reality and digital fabrication, thus creating extensive moral, legislative, and security risks. Digital content made with machine learning and artificial intelligence produces deepfakes that present realistic but deceptive media like videos and photos, together with audio. GANs and additional neural networks form the foundation of this technology that delivers synthetic media perfection, surpassing human capabilities in creating realistic simulations.[1]

The world now expresses concern about deepfakes after their original technological novelty became subject to malicious use. Several financial scams and deceptive political messages have been enabled through the improper misuse of this technology.[2] Deepfake technology presents an alarming risk through its implementation of cyber-facilitated criminality, which involves using manipulated images and videos for harassing and extorting or tarnishing the reputation of adult victims, particularly women.[3] On similar footings, a documentary showcasing the dark side of machine culture, named 'Another Body,' showcases a 23-year-old graduate student who logged onto Facebook in 2020, found a porn hub account that featured several deepfakes adult videos of her. The documentary reveals the dark underbelly of internet cultures that enable users to manipulate machine learning for the sake of hurting women and minorities — all while putting a deepfakes human face on screen.[4] These incidents showcase the urgent need for action.

The difference between a deepfake and a simple image is that with the former, the victim is truly there. When a video displays the face of the victim and the peculiarities of their voice are heard in the video, the human brain interprets it as if the victim is actually there. Hence,

[1] Ian Goodfellow et al., 'Generative Adversarial Networks' (2014) 3(11) Advances in Neural Information Processing Systems <https://doi.org/10.1145/3422622> accessed 06 November 2025
[2] Henry Ajder et al., *THE STATE OF DEEPFAKES LANDSCAPE, THREATS, AND IMPACT* (Deeptrace, 2019)
[3] Danielle Keats Citron and Robert Chesney, 'Deepfakes and the New Disinformation War' (*Foreign Affairs*, 11 December 2018) <https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war> accessed 06 November 2025
[4] Trisha Dasgupta, 'Another Body: Documentary Tells Story of Deepfake Pornography Survivor's Search for Justice' *The Daily Texan* (20 March 2023) <https://thedailytexan.com/2023/03/20/another-body-documentary-tells-story-of-deepfake-pornography-survivors-search-for-justice/> accessed 06 November 2025

the trauma quotient in deepfake abuse, including the creation of unwanted intimate images, is so severe. It is not just an image; it is an actual living imitation that appropriates the identity of the victim. The trauma is doubled when it comes to the victim. They face the invasion of privacy as well as the nightmare of their digital duplicate doing something that they never consented to in their lifetime.

The continuous increase of deepfakes requires platform ownership responsibility and regulatory intervention from responsible authorities. The transmission of harmful deepfakes depends on social media platforms and AI developers, alongside content hosting platforms, since they both enable and restrict the propagation of dangerous deepfakes.[5] Better detection technologies and protective legislation must be developed to stop digital deception spread because they represent the path forward.

The blog discusses deepfake deception evolution alongside its middleman problems and possible remedies to counter synthetic media threats. Understanding this digital threat leads to creating implementable measures that transform this internet threat into a defensive tool against cybercriminals and manipulators.

Beyond the pixels and code, deepfakes represent a profound 'epistemic threat,' a danger to our ability to know what is true. For decades, video and audio served as the 'gold standard' of evidence in courts, journalism, and personal relationships. Deepfakes shatter this standard, creating a world where any event can be fabricated with such precision that the average person cannot distinguish it from reality. This leads to a state of 'reality apathy,' where the public, overwhelmed by the difficulty of verifying facts, simply stops believing in anything they see online. This collective cynicism is perhaps more dangerous than any single fake video, as it allows real wrongdoers to dismiss genuine evidence as 'just another deepfake.'

DEEPFAKE: RISE OF SYNTHETIC DECEPTION

Deepfakes have rendered reality and perception as indistinguishable from one another, thereby allowing more space for deception in the virtual world. The threat posed by

[5] 'Deepfake Manipulation: AI and ML-Based Digital Deception' (International Conference on Computers and Applications in Security, New York City, USA, 2020)

deepfakes has long-term implications, eroding confidence in virtual society. The issue of deepfakes has reached both the collective society and its members at the individual level. As Justice Kohli has rightly pointed out, deepfakes cannot be differentiated from real events, posing a serious challenge to the validity of information and the integrity of individual identity.[6] This technological development goes beyond the mere production of AI-generated indecent videos and has led to a chain of other digital offences, such as sextortion, manipulation of geopolitical tension, and a chain of ethical and psychological ramifications.

Sextortion, a grave offence of deepfakes, is where an individual threatens a victim with a doctored objectionable video of the victim for extorting benefits that can vary from financial gains, sexual favours, or acts of revenge, which tend to cause grave psychological trauma. Sextortion is a grave invasion of privacy and a social menace. The offence has implications beyond celebrities and extends to common people as well. Justice Mahajan rightly characterised these abuses and argued that sextortion not only destroys personal dignity but also poses serious challenges to law enforcement agencies because it is "clandestine and cross-jurisdictional in nature.

In 2023, the FBI indicated that there was a surge in the number of victims of sextortion who asserted that fake images or videos, created from publicly available information on their social media or web blogs, were forwarded to the perpetrator upon demand, or recorded during video calls. According to the FBI, law enforcement agencies in the past year (2022) had recorded more than 7,000 offences of online sextortion against children, where at least 3,000 individuals were victimised. Moreover, the warning indicated that more than a dozen victims of sextortion have died by taking their own lives.[7] The surge in the cases of suicides among the victims of sextortion, particularly innocent targets such as women and children, stems from their lack of knowledge about how to navigate through and respond to such situations effectively. As Justice Sandeep Moudgil had rightly stated that acts like sextortion are currently the highest reported form of image-based sexual abuse, a form of online blackmail which has been growing in prevalence since 2021.

[6] 'Emergence of Deepfake Technology Cause of Deep Concern: Supreme Court Judge' *NDTV* (20 February 2023) <https://www.ndtv.com/india-news/emergence-of-deepfake-technology-cause-of-deep-concern-supreme-court-judge-4649510> accessed 06 November 2025

[7] Simon Hendery, 'Deepfakes of victims used in sextortion attacks spike, FBI warns' *SC World* (07 June 2023) <https://www.scworld.com/news/deepfakes-sextortion-spike-fbi> accessed 06 November 2025

Moreover, the exploitation of geopolitical tensions constitutes another important aspect regarding deepfake technology. With the evolution of this technology, it has increasingly been used as a weapon of war against states to create instability among them. This tactic has increasingly been regarded as a conventional technique in today's military warfare. A deepfake video of President Putin, for instance, appeared during the Russia-Ukraine war and was largely perceived as a piece of satire. However, note that there is a fine line between satire and disinformation.

However, the emergence of deepfakes has a strange and two-sided danger in the 'Liar's Dividend.' This refers to the fact that the technology can make a liar credible on one hand; on the other hand, it provides an escape clause from responsibility if one is caught in a genuinely embarrassing situation, if they can prove the video evidence is a deepfake.

However, when deepfakes become common, the threshold of proof becomes impossibly high. If all videos of a bribe or a violation of human rights can be effectively labelled as 'AI-generated,' then the system of judges of truth embodied in our institutions is eliminated, and we are left vulnerable to either the loudest shout or the best story.

The widespread propagation of incorrect information can result in riots and civil disturbances. This goes to demonstrate the gravity of this artificial intelligence technology, which can be used to facilitate far more heinous crimes, ultimately affecting the lives of millions. Since such crimes are grave, policymakers are required to act by implementing tougher laws against such crimes. Moreover, this problem has to be tackled in global forums to foster cooperation among countries, thus making efforts to curb such criminal activity more effective. Moreover, intermediaries have to be held responsible and asked to delete any objectionable material as soon as it is uploaded. There is a pressing need for strong measures to counter the massive threat posed by deepfake technology.

CRITICAL ROLE OF INTERMEDIARIES IN SAFEGUARDING DIGITAL SPACES

As the deepfake technology becomes easily accessible and popular, its misuse continues to increase cybercrimes such as sextortion, misinformation, and geopolitical manipulation. While the individuals suffer from these violations, the platforms that host and distribute content play a crucial role in curbing the spread of such harmful activities.

In India, we have the Information Technology Act 2000,[8] which lays down rules and regulations to prevent the spread of obscene and sexually explicit materials on the digital web. Section 67[9] of the said act deals with punishment for publishing or transmitting obscene material in electronic form, whereas Section 67A[10] deals with punishment for publishing or transmitting material containing sexually explicit acts, etc., in electronic form. However, till now, this act fails to prohibit the spread of obscene, sexually explicit materials or deepfakes in private group chats on different intermediaries, for example, Telegram, Instagram, WhatsApp groups, etc. Deepfakes of different actresses can be found on different private groups on these intermediaries despite their takedown from the web. Intermediaries claim Section 79[11] as their defence in response to failure to prohibit sexually explicit or deepfake content on their platforms. Section 79 gives intermediaries the defence of 'Safe Harbour.' As per this section, intermediaries shall not be liable for any third-party information, data, or communication link made available or hosted by them if the function of the intermediary is limited to providing access to a communication system over which information made available by third parties is transmitted, or the intermediary does not initiate the transmission and they observed due diligence while discharging their duties. Moreover, as per Section 79(3),[12] intermediaries shall not be liable if they don't have the 'knowledge' about the content that is being floated on their platforms.

Therefore, in the majority of cases, like the Boise locker room,[13] etc., intermediaries claim that they didn't have actual knowledge about the content that is being circulated in private group chats, as messages in private group chats are end-to-end encrypted, and it would be a violation of group members' privacy if intermediaries decrypted private group chats. Hence, they did not have 'actual knowledge' of the same; therefore, they must get the defence of Section 79 of the IT Act. But in the present circumstances, it is necessary to consider a wider interpretation of 'Actual Knowledge' under section 79(3)(b[14]. The deepfakes have been prevalent for a long time, and it is way more obvious in the present time that sexually explicit

[8] Information Technology Act 2000
[9] Information Technology Act 2000, s 67
[10] Information Technology Act 2000, s 67A
[11] Information Technology Act 2000, s 79
[12] Information Technology Act 2000, s 79 (3)
[13] 'Bois Locker Room: 10 things you need to know about scandal that has rocked Indian social media' *India Today* (06 May 2020) <https://www.indiatoday.in/india/story/bois-locker-room-10-things-you-need-to-know-about-scandal-that-has-rocked-indian-social-media-1674687-2020-05-05> accessed 07 November 2025
[14] Information Technology Act 2000, s 79 (3)(b)

material, deepfakes, non-consensual intimate images (NCII), etc., can be easily shared on intermediaries where private groups exist. Hence, they must not claim in every situation that they did not have actual knowledge about the content that is being floated on their open platform or in group chats.

Intermediaries have to take stricter measures like proactive monitoring, content filtration, etc, to control illegal material like deepfakes, NCII, etc., and proactive monitoring of suspicious groups will not result in violation of the right to privacy, as privacy exists till a point where one is not infringing someone else's privacy. Hon'ble Apex Court has stated in ***Telegram v Neetu Singh***[15], that disclosure of personal data of individuals who were circulating infringing material, finding that no right to privacy exists, as individuals were undertaking illegal activities. The Court thought that these rights cannot prevent an infringer from facing the consequences of illegal actions. Therefore, the role of intermediaries becomes very much prevalent in curbing deepfakes and other obscene material.

TURNING THE THREAT INTO A SHIELD

Now, as we have discussed above about the extent of harm deepfakes can cause, per se, the readers by now might have been familiarised with the implications of this new technology and to what extent it can cause damage to the social fabric and can create trust erosion. Due to these issues, it is high time that we find some solutions to tackle this surging problem of deepfakes. As the subtitle suggests, turning the threat into a shield means that, as this AI technology can be dealt with by other AI technology, which can be programmed for in-depth analysis of the doctored video, either by analysing the voice modulation or by the movement of the video. Also, open-source AI and other AI websites can implement a mandatory watermark that can help to decipher whether the video is doctored or not.

One of the key solutions to the issue of deepfakes is the application of Cyber Threat Intelligence (CTI). Cyber Threat Intelligence is the process of information gathering, analysis, and use for the purpose of understanding and countering cyber threats. This includes gathering information about the tactics, techniques, and procedures (TTPs) used by the threat

[15] *Neetu Singh v Telegram FZ LLC* CS (COMM) 282/2020

actors. Moreover, threat intelligence based on artificial intelligence enables organisations to determine probable adversaries, predict their behaviour, and take proactive steps.

There should be a Threat Intelligence Platform (TIP) to accumulate and analyse cyber threat information about an organisation. Threat Intelligence Platforms (TIPs) significantly enhance threat detection, response actions, and overall security measures by offering contextually relevant and timely intelligence. TIPs enable organisations to rapidly identify potential threats and take adequate countermeasures against the existing potential harms. By consolidating information from diverse sources, such as open-source intelligence (OSINT) and dark web monitoring, TIPs offer a consolidated view of the threat landscape. Furthermore, machine learning algorithms scan these data streams to conduct cross-correlations and identify threats, with risk and severity analysis. In this way, then, artificial intelligence technology can help reduce the effectiveness of deepfakes.

The other measures that can be taken are to bring all the stakeholders on a common platform to find an accelerated resolution to these matters. In addition to this, there is an urgent need to recognise this crime in the laws, which shall aid in getting recognised, and the general mass shall be made aware of the dangers of deepfakes.

These are short-term measures that can reduce the pace at which the deepfake is proliferating, as a consequence of which all the parties, as well as the intermediaries, must seek a long-term solution for the restoration of trust in the online space.

CONCLUSION

The quick expansion of deepfake technologies transformed the digital space by creating challenges for people to identify artificial human-made alterations from genuine facts. The original positive perception of deepfakes as progressive technology has turned into major ethical problems and legal barriers, combined with security risks. The emergency demand for regulatory action, along with proactive measures from intermediaries, emerges from the existing risks of exploitative sextortion, misinformation and geopolitical exploitation.

Technology companies, along with social media services, constitute essential actors who need to actively combat dangerous deepfake content circulation. The current version of the Information Technology Act 2000 offers small levels of accountability but falls short in

protecting against modern digital deceits. The risks from deepfakes need to be minimised through proactive databases combined with Artificial Intelligence detection algorithms, which should become mandatory security measures.

Security platforms that use artificial intelligence technology like Cyber Threat Intelligence, together with Threat Intelligence Platforms, function as strong defensive barriers against deepfake-based deception. Digital trust restoration requires stakeholders to work together while raising public understanding to develop the necessary measures. Deepfake technology requires multiple approaches from law enforcement, together with technical solutions and regulatory standards, to shift it from a deceptive instrument to a protection against cyber risks.