# DEEPFAKES, DEFAMATION, AND DIGITAL EXPLOITATION: LEGAL AND ETHICAL CHALLENGES IN THE AGE OF AI

**Pavithra Kakkanatt Rajesh***

## ABSTRACT

*Deepfake technology has evolved from an amusement tool to a serious threat to privacy, reputation, and democratic stability. It is driven by artificial intelligence models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). This technology makes it simple to produce realistic-looking but phoney audio, video, and images. The distinction between fact and fiction is blurred by deepfakes. This poses significant problems for personal safety, misinformation control, and defamation law. In non-consensual pornography, the harm is particularly severe. Deepfakes have turned into instruments of digital exploitation in this case, primarily affecting women and resulting in long-term psychological and reputational damage. Case studies, like the Rashmika Mandanna incident in 2023, demonstrate the speed at which manipulated content proliferates and the shortcomings of the detection and removal techniques in use today. While states—including India—have invoked older statutes like the Information Technology Act and attempted to craft sector-specific codes for artificial intelligence, the statutes themselves lack teeth, and the pace of enforcement lags behind technical evolution. Spurious audio and video creations jeopardise not only private reputations but also the fundamentals of governance, boardrooms, and civic debate, corroding the substratum of trust on which democracies rely. Against this backdrop, the present discussion examines the intersecting ethical, legal, and social layers associated with deepfakes. It concludes that the window for effective mitigation is narrow and recommends the immediate scaling of robust detection technologies, the codification of calibrated and proportionate prohibitions, the creation of streamlined grievance redress for individuals of all backgrounds, and the forging of multilateral agreements that transcend domestic silos to confront the menace in its transnational, virulent form.*

*BBA LLB, SECOND YEAR, CHRIST (DEEMED TO BE UNIVERSITY), BANGALORE.

**Keywords:** Deepfake Technology, Artificial Intelligence, Defamation, Misinformation, Non-Consensual Pornography, Cybercrime and Digital Exploitation.

## INTRODUCTION

Deepfake technology, driven by artificial intelligence, allows for the rapid and effortless production of hyperrealistic videos that replace one person's face with another's body. The technology uses neural networks such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) to create highly realistic, deceptive media.[1] The ability to create hyper-realistic but entirely fabricated videos, images, and audio recordings has raised significant concerns about reputational harm, misinformation, and the adequacy of existing legal protections.[2] The most intriguing implications of deepfakes revolve around the issue of whether specific content is defamatory. While states define defamation in different ways, the main idea is that a publication must be harmful enough to cause people to think less of the plaintiff. Figuring this out can be complicated with videos. Courts recognise that a skillfully created mix of partial truths and opinion-like statements, combined with carefully chosen images and dramatic audio, can be very damaging when shown on television.[3] The combination of deepfakes and harmful bots can be very damaging. Imagine a scenario where a deepfake video is made to defame someone, and then a swarm of bots shares the video on social media. The sheer volume and seeming authenticity of the content can seriously harm a person's reputation. The internet has become a breeding ground for new and harmful ways to attack someone's reputation.[4] Deepfakes and malicious bots represent two particularly concerning trends today. Although there are positive uses for this technology, it has primarily been exploited to create nonconsensual pornographic content. Deepfake porn is a serious sexual crime that has affected hundreds of thousands of women. Deepfake pornography is a serious tool for digital exploitation. It creates convincing but false evidence that can harm victims deeply. Once released, these videos spread quickly on anonymous platforms, making it almost impossible to remove them and recover one's reputation. Let’s explore how cybercriminals are

[1] Keepnet Labs, “Deepfake Pornography: Understanding the Threat and Protecting Your Employees” (*Keepnet Labs*, March 12, 2025) https://keepnetlabs.com/blog/deepfake-pornography-understanding-the-threat-and-protecting-your-employees

[2] Tziolis I, “The Deepfake Dilemma: Navigating Defamation in the Age of Deepfakes” (*BlackBay Lawyers*, May 20, 2025) https://www.blackbaylawyers.com.au/post/the-deepfake-dilemma-navigating-defamation-in-the-age-of-deepfakes

[3] Milkovich v Lorain Journal Co 497 US 1 (1990).

[4] Aviv Ovadya and Sam Woolley, ‘Computational Propaganda: Bots, Deepfakes and the Epistemic Crisis’ (2019) 36(3) Journal of International Affairs 23.

using deepfake pornography for blackmail, corporate espionage, and social engineering attacks.[5]

## CASE STUDY: THE RASHMIKA MANDANNA DEEPFAKE CASE

A viral deep fake, in which actress Rashmika Mandanna is shown entering an elevator in a suggestive black costume, went public in November 2023. The video was doctored through AI, where the face of a British-Indian influencer, Zara Patel, was replaced with the face of Mandanna, which caused an outcry on the internet. Mandanna also called out the video publicly as being extremely scary, noting how it put people in a vulnerable position, and other celebrities such as Amitabh Bachchan, Mrunal Thakur, and Naga Chaitanya voiced the need to regulate more and keep those who go through with the act accountable.[6]

The government of India acted fast. The Ministry issued GSPCC removal advisories to social media companies, requiring deepfake material to be removed within 36 hours in accordance with the IT Rules 2021; failure to remove such content exposes the company to a fine of up to 3 lakhs and one year imprisonment under Rule 7. On November 10, 2023, Delhi Police filed an FIR under Sections 66C (identity theft) and 66D (cheating by personation) of the IT Act,[7] 2000, which carry up to three years of jail term and a fine of up to 1 lakh and IPC 465 (forgery), 469 (forgery with intent to harm reputation and 509 (insults the modesty of a woman).[8] The principal suspect behind this, Eemani Naveen, a 24-year-old from Andhra Pradesh, was arrested in January 2024 because he made the video in an attempt to increase his Instagram followers.[9] The incident highlighted the looming danger of deepfakes not only on celebs but also in politics and daily life. In the 2024 elections in India, politicians were targeted by fake information generated by AI, and 25 per cent of Indians had been exposed to this misinformation.[10] It stimulated the debate around ethical AI and the demand to better protect

[5] Keepnet Labs, "Deepfake Pornography: Understanding the Threat and Protecting Your Employees" (*Keepnet Labs*, March 12, 2025) https://keepnetlabs.com/blog/deepfake-pornography-understanding-the-threat-and-protecting-your-employees

[6] Deepfake video of actor Rashmika Mandanna: How to identify such videos, what to do if you are a victim' The Indian Express (India, 8 November 2023) https://indianexpress.com/article/explained/explained-sci-tech/rashmika-mandanna-deepfake-video-9019394/ accessed 18 August 2025

[7] Information Technology Act 2000, s 66C & 66D (India).

[8] Indian Penal Code 1860, ss 465, 469, 509 (India).

[9] 'Rashmika Mandanna deepfake case: Main accused arrested from Andhra Pradesh' Hindustan Times (India, 20 January 2024) https://www.hindustantimes.com/india-news/rashmika-mandanna-deepfake-case-main-accused-arrested-from-andhra-pradesh-101705766888965.html accessed 18 August 2025.

[10] Matt Kamen and Nilesh Christopher, 'Indian Voters Are Being Bombarded With Millions of Deepfakes. Political Candidates Approve' WIRED (24 April 2024) https://www.wired.com/story/indian-elections-ai-deepfakes/ accessed 18 August 2025.

it, and with specific law, current provisions against a similar offence are no longer covered by the current frameworks, such as the Bharatiya Nyaya Sanhita (successor to IPC), which are limited to these offences and do not provide a framework. The Indian Cybercrime Coordination Centre (I4C) named Mandanna National Ambassador of Cyber Safety in October 2024, making her an ambassador of awareness against cyber risks such as deepfakes, phishing and fraud.11 This case in point draws attention to the importance of technological protection, cross-border coordination, and the modernisation of relevant laws to counter AI misuse.

## ANALYSIS

Deep-fake technologies are backed by technologies such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs),12 what was once used for entertainment has rapidly turned into a pervasive threat of hyper-realistic media that blurs the line between what is real and what is not, even hindering international relations. By 2025, the use of deep-fakes will be alarmingly accessible, with the emergence of OpenAI, where text commands are being interpreted and converted into highly realistic deep-fakes. These artificial intelligence programs enable non-experts to create models which integrate text, voice and video to give a seamless output. However, these creation tools proliferate while the detection tools significantly lag. It is imperative to understand that the societal impacts of deep-fake technologies are profound. Deep-fake technologies erode the trust in media and amplify harms like misinformation, defamation and non-consensual pornography.

Deep-fakes are often used as political weapons to disregard adversaries or promote their own ideas. The release of the "Obama Arrest" deep-fake video by Donald Trump, which was captioned as "No one is above the law", was used as a political weapon to disregard Barack Obama, the former President of the United States. This deep-fake video was used to amplify the allegations against Obama for alleged administrative misconduct. Even though some social media users praised the video as a bold political move, others condemned it for being baseless, reckless and spreading misinformation.

[11] 'Rashmika Mandanna appointed National Ambassador for Promoting Cyber Safety' Hindustan Times (India, 10 October 2024) https://www.hindustantimes.com/india-news/rashmika-mandanna-appointed-national-ambassador-for-promoting-cyber-safety-101728567677774.html accessed 18 August 2025.

[12] Reza Babaei and others, 'Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis' (2025) 14(1) J Sens Actuator Netw 17 https://doi.org/10.3390/jsan14010017 accessed 17 August 2025

Courts recognise the amplified harm of visual media when dramatic videos and curated images create negative perceptions. Misinformation, especially in politics, can be detrimental to the democratic structure of our government. In the 2024 elections, 1 in 4 voters were influenced by these deep-fake videos, influencing public opinion and electoral integrity.

Nonconsensual pornography remains the most insidious form of this abuse, disproportionately affecting women and creating a digital form of sexual abuse. These fabricated videos or photos create false evidence and are rapidly spread through every medium, which makes it nearly impossible to remove them from these platforms. In 2023-2024 alone, the number of cybercrimes exceeded the record of all the prior years, with victims facing extortion, defamation, bullying and the erosion of personal privacy. Children, particularly teenagers, are undergoing extensive bullying because of the democratisation of Deep-fake technologies. Globally, around 60% of people encountered deep-fakes, with human detection accuracy at 62% heightening vulnerability.

Particularly in a country like India, deep-fakes have usually been targeted at celebrities like Rashmika Mandanna.13 In 2023, Rashmika Mandanna's face was swapped with British - Indian influencer Zara Patel. This sparked public outcry and swift action under the IT Act's Section 66D for impersonation. By 2025, India will have fortified its framework with the AI TRA Bill 2024, imposing up to five years of imprisonment for unauthorised deep-fake creation or distribution.

It is imperative to understand that improvements are urgently needed to address the enforcement gaps, technological advancement and cross-border challenges. First and foremost, states should initiate a statute criminalising the creation and not just the distribution, with a clear definition to not overstep on satire. Governments, in order to detect deep-fakes, can use Data Augmentation, Dynamic Face Cutout, Adversarial training, multi-modal analysis and other advanced techniques like GAN-fingerprint detection. Finally, and most importantly, victim support is the most important part of the fight against deep-fakes. Expedited trial, legal

[13] 'Deepfake video of actor Rashmika Mandanna: How to identify what's real and what's not' The Indian Express (New Delhi, 8 November 2023) https://indianexpress.com/article/technology/tech-news-technology/rashmika-mandanna-deepfake-video-how-to-identify-9017948/ accessed 17 August 2025.

aid, psychological aid and public-educational campaign can mitigate the harms of digital exploitation and reputation damage.14

## CONCLUSION

Deepfake technologies driven by new age technologies like GAN and VAN have transformed from the context of entertainment purposes to a deep, pervasive threat, producing hyper-realistic media that blurs the finite line between reality and fiction. Accessibility and that is fueled by new age Open AI platforms has enabled non-experts to create deceptive content, outpacing detection technologies, amplifying harms such as defamation, misinformation and non-consensual pornography. High-profile cases like the aforementioned case of actress Rashmika Mandanna deep-fakes underscore the personal and societal consequences, with the victims facing reputational damage, extortion and psychological trauma.

As deep-fakes continue to proliferate, balancing innovation and regulation is paramount. Positive applications in education and media must be preserved, but unchecked misuse threatens societal trust and individual rights. A collaborative, proactive approach -integrating advanced technologies as aforementioned, robust laws and global coordination offers the best way to curb this digital menace and use these technologies for educational and productive purposes.

[14] Momina Masood and others, 'Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward' (2021) arXiv:2103.00484v2 https://doi.org/10.48550/arXiv.2103.00484 accessed 17 August 2025.