

---

# PLATFORM LIABILITY FOR DEEPFAKES IN INDIA: REASSESSING INTERMEDIARY RESPONSIBILITY UNDER DIGITAL GOVERNANCE FRAMEWORKS

---

Gurnoor Kaur, B.COM. LL.B. (Hons), University Institute of Legal Studies, Panjab  
University, Chandigarh <sup>1</sup>

## ABSTRACT

Section 79 of the Information Technology Act, 2000, grants safe harbour to the intermediaries (social media platforms) in cases of any violations of personal data taking place over these. With no adequate legislation addressing deepfakes, it becomes really hard to address this issue when such an exclusion from liability is provided. The right to free speech and expression under the Constitution of India is at a high risk when it is so likely to be manipulated for ulterior gains. Looking at the global perspective, it comes as an indication that India ought to have a structured legislation addressing it. Further, looking into structural reforms, such as identification and marking of such data, is the need of the time before this technology does more harm than its pros.

**Keywords:** deepfakes, intermediaries, safe harbour, personal data and structural reforms.

---

<sup>1</sup> B.COM.LLB (Hons) at University Institute of Legal Studies, Panjab University, Chandigarh

## I. INTRODUCTION

As technology advances, it becomes vulnerable to be weaponised by “bad actors” who use it for illegal acts. Deepfakes are the latest tool used to mislead victims by using generated media to gain the trust of prospective targets. The generated media is used to either misrepresent the victims in an attempt to defame them or to mislead them for ulterior motives.

There is often a delay before authorities are made aware of any such issues due to a combination of the victims not realising they have been targeted followed by an over-complicated reporting process. The damage is often non-reversible and, because of the pervasive nature of electronic media, cannot be completely removed from the internet.

Through this article, the aim is to address this concern in respect to the statutory framework of India and the rest of the world with some suggested solutions. To start with it, one needs to understand what deepfakes really are and the kind of threats associated with them.

## II. UNDERSTANDING DEEPAKES AND THEIR SOCIETAL RISKS

### Conceptual understanding of Deepfakes

The term ‘deepfakes’ has originated from two words: ‘deep’, signifying deep learning system, and ‘fake’, denoting fabricated content. Deep learning systems are a subset of artificial intelligence and machine learning to analyse the complex data, in form of visual and aural media to manipulate the perceptions of human brain by using sophisticated algorithms, which alter the data to generate new content. Originating in the entertainment industry to make realistic special effects, it was quickly adopted in use for illegal purposes because of their ability to be accepted for being authentic. Regulating this technology poses two main threats: first, mitigating the harm caused by deepfakes while preserving the right to freedom of expression; and second, conditional safe harbour protection available to intermediaries under the law.<sup>2</sup>

### Risks associated with Deepfakes in India

India’s internet user base has significantly expanded to 954.40 million people in 2024 from

---

<sup>2</sup> Adv. Yashkumar S Bedmutha, ‘Regulating Deepfakes in India: Intermediary Liability and Constitutional Limits’ (LAWYERSCLUBINDIA, 9 March 2026) <<https://www.lawyersclubindia.com/articles/regulating-deepfakes-in-india-intermediary-liability-and-constitutional-limits--18208.asp>> accessed 5 May 2026

251.59 million people in 2014, reflecting widespread penetration in both rural and urban areas.<sup>3</sup> In a socio-culturally diverse and rapidly digitalising landscape, deepfakes pose serious threats by undermining democratic integrity, social cohesion and individual dignity. The non-consensual use of a person's identity or data for person gain or to cause harm is legally indefensible.

Recent incidents have reaffirmed these harms. In 2025, a case at IIIT- Nyaya Raipur involved deepfakes of female students for blackmail.<sup>4</sup> Similarly, public figures have been targeted through manipulated media including circulation of Rashmika Mandanna's deepfake video;<sup>5</sup> and Ranveer Singh and Aamir Khan's manipulated video of campaigning for political parties.<sup>6</sup> Each time we were told that the law will catch up, but it never did and this has led to harassment of individuals with no limits.

The rapid and borderless online dissemination ensured content is propagated worldwide and persists even after the manipulated content is deleted. Consequently, the victims of deepfakes, be they females, politicians, or celebrities, lose control of their public image and any subsequent correction, rarely mitigates the harm.

### III. INTERMEDIARY LIABILITY FRAMEWORK

Mass online communication is the most efficient way of communication in the modern world. Social media platforms, such as Facebook, WhatsApp, Instagram, etc., not only facilitate connectivity, but also permit circulation of unverified content publicly with no preventive measures and no liability at their end. Clearly, such channels accept no accountability and yet make hefty gains. In this relevance, the Supreme Court has given a proportionality framework in *Justice K.S. Puttuswamy v. Union of India*, reaffirming that any restriction must pursue a legitimate aim, adopt rational means, be necessary in the absence of less restrictive alternatives,

---

<sup>3</sup> Ministry of Communications, 'Universal connectivity and Digital India initiatives reaching to all areas, including tier-2/3 cities and villages' (Press Information Bureau, 02 August 2024)

< <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2040566&reg=3&lang=2> > accessed on 13 April 2026

<sup>4</sup>, Madhavi Ravikumar 'The new face of digital violence' (Frontline, 28 November 2025) <<https://frontline.thehindu.com/social-issues/gender/ai-deepfake-abuse-misogyny-crisisindia/article70335218.ece>> accessed on 13 April 2026

<sup>5</sup> Meryl Sebastian, 'Rashmika Mandanna calls for action against 'scary' deepfake video (Delhi, 07 November 2023) BBC News <<https://www.bbc.com/news/world-asia-india-67305557> > accessed on 13 April 2026

<sup>6</sup> Meryl Sebastian, 'AI and deepfakes blur reality in India elections' (Kochi, 16 May 2024) BBC News <<https://www.bbc.com/news/world-asia-india-68918330> > accessed on 13 April 2026

and include adequate procedural safeguards.<sup>7</sup>

## In India

IT Act<sup>8</sup> doesn't expressly regulate deepfakes but its provisions particularly Sections 66C, 66D, 67, 67A, and 67B address the harms arising from the synthetically generated media, including identity theft, impersonation, and obscene content. This framework is supplemented by Digital Personal Data Protection Act,2023<sup>9</sup> and Bharatiya Nyaya Sanhita,2023<sup>10</sup>, which criminalise privacy breaches, disinformation, and organised cybercrime, all of which are vectors for advanced fraud. But IT Act under Section 79 clearly provides conditional safe harbour to intermediaries for third-party content, contingent upon their neutrality as facilitators and compliance with due diligence obligations. It needs to be noted that providing an open interface in itself is facilitation and the term 'due diligence' is to be studied with respect to IT Rules, 2026<sup>11</sup>.

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021<sup>12</sup>, in 2026, adopted various provisions in reference to deepfakes. Wherein it has defined 'synthetically generated information' stating that any act, of this kind, done in good faith as an exception to it. It has mandated the intermediary to periodically inform users about compliance with laws and to have accountability for any such information if found to be relevant to them. Rule 3(1) of the Act deals with the description of due diligence to be exercised by intermediaries by ensuring compliance with the terms and conditions to curb synthetic media. Rule 3(2) mandates intermediaries to establish a grievance redressal mechanism and publish the name of the Grievance Officer, his contact details, and the mechanism to lodge a complaint. Further, it provides a deadline of twenty-four hours to acknowledge the complaint and seven days to resolve it from the date of its receipt. It also states that any complaint in respect to removal of any information or communication link shall be expeditiously resolved within 36 hours of such reporting. The Act also states various additional due diligence measures to be observed by significant social media intermediaries, which also includes online gaming intermediaries. Then, Rule 7 clarified that in cases of failure in compliance with the rules of

---

<sup>7</sup> *Justice K.S. Puttaswamy (Retd.) & Anr. v. Union of India & Ors* [2017] 10 SCC 1

<sup>8</sup> The Information Technology Act 2000

<sup>9</sup> The Digital Personal Data Protection Act 2023

<sup>10</sup> Bharatiya Nyaya Sanhita 2023.

<sup>11</sup> *Infra* Note at 11.

<sup>12</sup> The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021

this Act, protection under Section 79 shall not be applicable, and such an intermediary shall be liable for punishments under any law in force, including this Act and BNS, 2023<sup>13</sup>. Additionally, DPDA, 2023<sup>14</sup> focuses on protecting individual rights and in reinforcing the need to process data for legitimate purposes, primarily focusing on reasonable grounds of processing, proper notice, consent, and so on.

The Supreme Court in *Shreya Singhal v. Union of India*<sup>15</sup> under Sec 66A of IT Act stated that limitations on online speech should be subject to limited reasonable conditions as per Article 19(2) of the Constitution<sup>16</sup>. The regulation of deepfakes too has to maintain a constitutional balance of necessity to overcome synthetic fake news and assurance of legitimate free speech and expression.

### In Rest of the World

The adverse dynamics of deepfakes are dealt by various countries in their own manner, be it remedial or regulatory. The European Union Artificial Intelligence Act, 2024 (EU AI Act) under Article 50 focuses on transparency obligations for providers and deployers, ensuring that outputs of AI systems shall be marked in a clear, readable, and unambiguous manner.<sup>17</sup> The United States, under the Take It Down Act, 2025, has prohibited the non-consensual online publication of intimate visual depictions, both authentic and computer-generated, and required subsequent removal upon receiving notice of their existence within 48 hours of the notification.<sup>18</sup> China has strict regulations for online platforms mandating proactive moderation, user verification, labelling synthetic media, and taking measures such as deletion or blocking to control it.<sup>19</sup> This all has demonstrated a growing shift toward explicit statutory regulation of deepfakes by clearly defining the platform obligations and safeguards, whereas India continues to deal in fragmented intermediary liability provisions, highlighting the need

---

<sup>13</sup> *Supra* note at 9.

<sup>14</sup> *Supra* note at 8.

<sup>15</sup> *Shreya Singhal v. Union of India* [2015] 5 SCC 1

<sup>16</sup> The Constitution of India 1950

<sup>17</sup> Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) OJ L 2024/1689, art 50.

<sup>18</sup> TAKE IT DOWN Act 2025

<sup>19</sup> Xuanting Liu, 'Normative construction of platform liability in the governance of deepfake technology' (2025) 16(4) *Advances in Social Behaviour Research*

<[https://www.researchgate.net/publication/392556873\\_Normative\\_construction\\_of\\_platform\\_criminal\\_liability\\_in\\_the\\_governance\\_of\\_deepfake\\_technology](https://www.researchgate.net/publication/392556873_Normative_construction_of_platform_criminal_liability_in_the_governance_of_deepfake_technology)> accessed 25 April 2026

for a comprehensive regulatory framework with no resolution mechanism.

#### **IV. RETHINKING THE PLATFORM LIABILITY IN INDIA**

Addressing the concerns of Indian framework, taking into account international safeguards of deepfakes, some suggestions are listed hereunder:

1. There is a need to mandate the labelling of synthetic media. It may include tagging, watermarking or any other means. It needs to be properly formatted in user's language so that it is visible to a naked eye. Since both literate and illiterate people can have access to such media, marking needs to be done accordingly.
2. Accurate legal provisions must be made for deepfakes. A clear definition of 'deepfakes' along with degrees of punishments for distinct offences, for instance, the intention with which such deepfakes were generated. Also, gender-neutral laws ought to be promoted in this aspect.
3. Remedies available to victims do not correspond to the damage that has been caused. The speedy dispersion of media online is irreversible. The social media takedowns take too long; by then even if deleted, it has reached everyone around the globe. It causes extreme self-esteem and reputational losses. One of the way out in this scenario is to make such platforms accountable and subject to damages.
4. Platforms need to be provided a list of mandates and directories at universal level. It is necessary to strike a balance between the right to freedom and limitations on such freedom.
5. Special Authorities, in form of tribunals, can also be established to initiate any legal action in cases of non-compliance or any disputes for speedy redressal of specifically deepfakes.
6. Proper awareness is a must so that any victim doesn't feel alone in this battle. It should not only be about the facts of deepfakes but also legal authorities, remedies, and punishment to the offenders. It is necessary to disseminate such information in local language and in public areas.

Hence, there is a need to evolve the laws to address this issue of deepfakes and platform liability since the root cause is unregulated dissemination of media, which causes irreversible damages.

## **V. CONCLUSION**

Deepfakes represent one of the most complex regulatory challenges in the modern era. There is no specific legal framework to deal with, making it hard to not only to address but also prevent such future violations. This issue is further complicated by providing safe harbour to platforms under Section 79 of the IT Act. Platforms should not be treated as passive facilitators when their algorithms amplify such content and delay their effective takedowns. The comparative development in jurisdictions of EU, UK, USA, and China has demonstrated clearly that transparency obligation, mandatory labelling and platform accountability are both possible and necessary. Ultimately the question is not whether deepfakes can be eliminated completely but whether the law can evolve fast enough to minimise their harmful consequences while keeping pace with the chilling ability of AI to generate ever-more convincing materials.