DEEPFAKES AND DEFAMATION: A LEGAL PERSPECTIVE ON SYNTHETIC HARM

Akanksha Dubey[*] Ishita Tripathy[*]

ABSTRACT

The emerging technology of AI-generated synthetic media, also known as deep fake, is extensively used to manipulate audio-video content on social media and has introduced a new legal issue of dealing with misinformation campaigns and reputational assault on individuals online. With the emphasis on the Indian legal system, this research examines have definition of laws with the emergence of AI-generated content while also examining comparative judgment with countries like the US, UK and EU. This research also provides a study of AI development before offering reform suggestions for law and policy, gender-based harm and public responsibility and accountability while sharing the data on the internet. Artificial intelligence-generated synthetic media, or deepfakes, are quickly emerging as one of the most disruptive technologies of the digital era. Although they have innovative and instructive applications, their use in disinformation campaigns and reputational assaults presents significant legal issues. With an emphasis on the Indian legal system, this paper examines how defamation law deals with the emergence of deepfakes while also examining comparative jurisdictions like the US, UK and India. It makes the case for the acceptance of "synthetic harm" as a new legal category and draws attention to the shortcomings in the existing legal doctrines. The study also discusses platform accountability, gender-based harms, and ethical issues in AI development before offering reform suggestions for law and policy.

Keywords: Insolvency Resolution, Creditor Rights, Corporate Restructuring, Judicial Delays, IBC Reforms.

[*]BA LLB, FOURTH YEAR, AMITY UNIVERSITY, LUCKNOW.
[*]BA LLB, FOURTH YEAR, AMITY UNIVERSITY, LUCKNOW.

INTRODUCTION

The deepfake is a hazardous tool that is a fusion of Artificial Intelligence and media influence. Deepfakes use generative adversarial networks (GANs) to create videos and audios that can have an audio of a real person. From reputational defamation to political misinformation, the ramifications are wide. The reputation of an individual, which was shaped mostly by social norms and limited by geographical boundaries, is now extended far beyond borders in today's world, where easy access to the network is available. In the digital age the social media can easily manipulate a larger population in a few seconds of a clip reel.

Deepfakes represent a powerful and hazardous shift in how repetition can be harmed by using online sources, like traditional defamation, which usually involves speaking or writing words against an individual. Defect imitates real-life visuals and speech, making the lie appear more realistic and believable to society. This makes them far more damaging, as people are more likely to believe what they see. As a result, existing definition loss fails to protect against the repetition loss by AI and deceptive technology, which need urgent focus.

The Indian laws must keep pace with technology, and a change in law is needed regarding the protection of individual dignity in the digital world.

TECHNOLOGICAL ANATOMY OF DEEPFAKES

GANs and Deep Learning: Two neural networks, the discriminator and the generator, compete against one another in GANs, which are machine learning models. While the discriminator works to separate the generated content from authentic content, the generator produces synthetic content. The output is, over time, enhanced by the system to almost perfect authenticity.

Detection Challenges: It is a highly complex system to detect deepfake content. Standard digital forensic techniques like facial recognition and metadata analysis are no longer able to protect. Since deepfakes themselves use adversarial machine learning techniques to evolve, even AI-powered detection tools have limitations.

REVISITING DEFAMATION LAW IN THE CONTEXT OF DEEPFAKES

The Development of Defamation in Background: Beginning in the common law system, the tort of defamation was established to shield people from untrue statements that damage their

reputation in the eyes of the public. While civil defamation claims in India are based on uncodified tort law, criminal defamation is firmly established by Section 356 of the Bhartiya Nyaya Sanhita.[1] But because they were created before the advent of digital technology, both of these frameworks are mainly unprepared to handle the multifaceted harm caused by synthetic media.

Causation and Intent Issues: For a defamation claim to be successful, plaintiffs must prove the following:

- The statement was inaccurate.
- It was printed.
- It mentioned the plaintiff.
- It damaged the publisher's reputation.
- The publisher was at fault (either careless or malicious, depending on the jurisdiction).

Deepfakes are challenging all of these:

- When the manipulation is seamless, it is hard to prove the falsity.
- Because of anonymised or foreign servers, the attribution is frequently hidden.
- Many people simply share viral content without realising it is fake, so the uploader's intent may be lacking.

Because of this, determining liability under current doctrines is challenging.

GENDER AND DEEPFAKES: DISPROPORTIONATE IMPACT

Deepfake Pornography: More than 90 per cent of videos made through deepfake consist of pornographic content, especially of women, without their consent.As a result of this, individuals, celebrities, and journalists face problems. In India, there have been instances of blackmail, revenge schemes, and WhatsApp groups utilising women's faces superimposed on explicit bodies.

Need for Gender-Sensitive Laws: Laws addressing gender based pornographic content are an essential demand in recent India. The Indian legal system must address the violation of

[1] Bharatiya Nyaya Sanhita 2023, s 356

individual dignity and the right to live with personal liberty freely, under Article 21 of the Constitution.[2]

COMPARATIVE JURISPRUDENCE AND GLOBAL TRENDS

United States: The First Amendment's guarantee of free speech is still firmly upheld by U.S. jurisprudence. Although this restricts defamation lawsuits, states have passed particular legislation:

- "Deepfake election interference" is illegal under California Penal Code Section 653.2.
- Deepfake pornography is prohibited by Virginia's non-consensual pornography laws.

These are state laws, though, and enforcement is still dispersed. A challenging obstacle in deepfake cases is the New York Times v. Sullivan (1964)[3] standard, which calls for evidence of "actual malice."

United Kingdom: By concentrating on whether the statement resulted in "serious harm" to reputation, the Defamation Act of 2013[4] offers a more claimant-friendly approach. In deepfake cases, the UK's stringent privacy and harassment laws may also be used.

The European Union: Platforms are required to comply with the EU's Digital Services Act (DSA)[5] and AI Act by:

- Perform risk assessments;
- Identify and label content produced by AI;
- Quickly remove unlawful content.

This proactive strategy is in contrast to India's reactive one.

[2] Constitution of India, art 21
[3] New York Times Co v Sullivan (1964)
[4] Defamation Act 2013 (UK)
[5] Regulation (EU) 2022/2065 of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) [2022] OJ L277/1

INDIAN JUDICIAL RESPONSE

India hasn't seen a blockbuster legal case about deepfakes yet. Still, our courts have dealt with cases that come pretty close, especially when it comes to tampered content and someone's reputation taking a hit.

Take Subramanian Swamy v. Union of India (2016),[6] for example. The Supreme Court in that case said criminal defamation laws are okay — they don't violate free speech as long as they're used reasonably. It was held that protecting people's dignity is equally important as free speech by an individual.

In another case of Khushwant Singh v. Maneka Gandhi (2002),[7] the Delhi High Court said satire cannot be anything and should not be used to damage someone's image. The ruling hinted that the courts are open to dealing with reputation risks, and that attitude could easily extend to deepfakes once they show up in a big legal fight.

ETHICAL CONCERNS AND THE FREE SPEECH DEBATE

Deepfake technology's widespread use has raised serious concerns and put freedom of speech and expression, which is a fundamental right, at risk. The conflict between technological advancement and abuse of fundamental rights is the centre of the problem, especially when it comes to abuse of free speech, privacy violations and spreading misinformation.

In 2024 a viral video from social media was created a lot of violence in Maharashtra, in the video it was supported to show a minority community discarting a religious side in Maharashtra leading to public outrage in certain cities and crisis which led to violence and damage of public vehicle and property in Maharashtra, later after the investigation it turned out to be a fake video and the owner was not found. This misinformation can lead to significant damage in society and can be used as a weapon to create a nuisance.

Individual personal autonomy and the right to live with dignity are violated when free speech is used for defamatory and derogatory purposes.

[6] Subramanian Swamy v Union of India (2016) 7 SCC 221 (SC)
[7] Khushwant Singh v Maneka Gandhi AIR 2002 Del 58

In India, freedom of speech and expression is given under Article 19 (1)(a)[8] of the Indian Constitution, but it is subject to certain restrictions and is not absolute under Article 19(2). Indian codes have raised concerns about the misinformation being circulated on social media and damaging persons' right to live with dignity quote A firm that, if there is a competition between Article 19 and Article 22[9] of the Indian constitution, Article 22 prevails over the other.

Deepfakes must be acceptable in the form of expression and speech, especially when they are about political criticism, satire, etc, and do not cause provocation in society. The blanket ban on deepfakes may result in censorship, which will infringe the right to freedom of speech and Expression.

POLICY RECOMMENDATIONS

A strong legal and policy framework is necessary to lessen the negative effects while upholding constitutional values like free speech and due process, given the growing threat posed by deepfakes and their intersection with defamation and digital harm. To address synthetic harm in a way that is both morally and legally sound, the following suggestions are put forth:

Statutory Definition of Deepfakes: The official acknowledgement of "deepfakes" in India's legal system is a crucial first step in controlling synthetic media. At the moment, there is no precise statutory definition provided by the Bhartiya Nyaya Sanhita or the Information Technology Act, 2000 (IT Act).[10] A clear, technologically neutral definition of deepfakes that distinguishes between benign uses (like satire or education) and malicious or non-consensual manipulations meant to deceive or harm should be included in the proposed Digital India Act or an appropriate amendment to the IT Act.[11] Such a definition should define thresholds for what qualifies as harmful or misleading deepfakes and cover text, audio, and visual-based synthetic content.

Criminalisation of Malicious Deepfakes: Deepfakes can cause serious harm to one's reputation and psychological well-being, especially when they involve impersonation, defamation, cyberstalking, or sexual abuse. For this reason, making and sharing malicious deepfakes should be considered a separate criminal offence. When there is a clear intent to

[8] Constitution of India, art 19(1)(a)
[9] Constitution of India, art 22
[10] Information Technology Act 2000 (India)
[11] nternet Freedom Foundation, 'Digital India Act: Consultation Draft Analysis' (2024) https://internetfreedom.in accessed 12 july 2025

harass, extort, defame, or mislead, such actions should be punished by law. In order to address the misuse of deepfake technology, specific offences that are similar to those under Section 66E (violation of privacy)[12] and Section 67A (sexually explicit content)[13] of the IT Act may be added to the Bhartiya Nyaya Sanhita or the Digital India Act. Crucially, in order to stop these provisions from being abused against lawful content creators, mens rea, or criminal intent, must be a necessary component.

Accessible Civil Remedies for Victims: Victims must have access to prompt and efficient civil remedies in addition to criminal penalties. This ought to consist of:

- Orders to stop the spread of manipulated or defamatory content.
- Takedown orders are required to eliminate dangerous deepfakes from internet platforms.
- Compensatory damages to compensate for financial loss, emotional distress, and harm to one's reputation.

Establishment of Deepfake-Specific Fast-Track Mechanisms: Time is of the essence when it comes to providing redress because synthetic content is viral. Dedicated fast-track cyber justice procedures should handle complaints pertaining to deepfakes, either under the recently created Digital Harms Tribunals or as a component of the current cybercrime cells. These devices ought to have:

Skilled professionals in digital forensics, accelerated timelines for procedures, and systems for adjudication aided by technology that can differentiate between real and fake media.

This would guarantee that victims can quickly restore their reputations and are not left exposed for extended periods. In situations where the harm is ongoing or the identity of the perpetrator is unknown, courts ought to have the authority to issue ex parte orders. In order to ensure prompt justice, provisions should also be made that permit victims to seek temporary relief while investigations are underway.

Gender-Sensitive Provisions to Combat Deepfake-Based Sexual Violence. Deepfake pornography and image-based abuse disproportionately affect women and marginalised

[12] Information Technology Act 2000, s 66E
[13] Information Technology Act 2000, s 67A

genders. Therefore, gender-sensitive clauses that address the following issues must be included in the legal framework:

- non-consensual sexual deepfakes, frequently utilised in cyberstalking or revenge porn;
- face morphing onto explicit material;
- targeted abuse of public figures, journalists, and activists.

Such behaviour should be explicitly defined as digital sexual violence in amendments to the IT Act and the BNS, which should also include aggravated penalties. Affected individuals must also have access to legal aid, confidentiality protections, and psychosocial support services.

Platform Accountability and Conditional Safe Harbour: The dissemination of deepfakes is greatly aided by digital platforms. Although Section 79 of the IT Act's intermediary protections are crucial for encouraging innovation and free speech, they ought to be subject to responsible conduct. Platforms ought to be mandated by law to:

- Install automated deepfake content detection systems.
- Give AI-generated media a clear label or watermark.
- Put in place quick takedown procedures in response to confirmed complaints.
- Continue to provide transparent reports on your efforts to moderate content.

Legal action against platforms that wilfully permit the spread of harmful content should be made possible by the loss of safe harbour protections for noncompliance with such duties. Additionally, social media companies and government organisations need to work together to create common guidelines for transparency and accountability.

CONCLUSION

Deepfakes are visual lies that pass for reality, not just another type of deception. They put our legal, moral, and technological philosophies to the test, especially when it comes to defamation law. A strong framework is needed to maintain balance between integrity and civil liberties, and decisive actions are required to be taken by Indian courts to acknowledge "synthetic harm"

In recent times, I strong legal Framework is much in demand to maintain a balance between the right to live with dignity of an individual and freedom of Civil liberties. Active measures in recent times are required to be taken by the Indian courts to acknowledge synthetic harm.

The need to address these concerns is important as high-level technology is involved. India needs stringent laws to safeguard against the misuse of the internet for spreading misinformation and hate, and derogatory content.