# The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2026

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896




## Deepfake and AI Evidence in Law: Assessing Authenticity and Protecting Personality Rights


***Mahak Raikwar***


### Introduction

Imagine waking up to a video of yourself, saying things you never said, doing things you never did. It looks like you. It sounds like you. But you know it's not you. The saying 'seeing is believing' suggests that we trust what we observe. But how long can we depend on this notion? As technology continues to advance rapidly, people have found numerous ways to misuse it, leading to a significant increase in cybercrimes. Deepfakes are one of the serious cybercrimes whose cases have been increasing recently. With generative AI, creating images, videos, and audio of people doing things they never did became much easier.

"Deepfakes" are fake videos, audio, or images created or altered using advanced AI methods, primarily through artificial intelligence. As deepfakes become more sophisticated and believable, the primary question for courts, police, and society as a whole is: How can we trust what we see and hear?

### What is a Deepfake?

If we combine 'deep learning' and 'fake', then it will be known as 'Deepfake'. Deep learning is one of the types of 'machine learning' that works with complex data sets.

Deepfakes use Deep learning, which is an AI technology and is a part of Machine Learning. Deep learning models are made to look at large, complicated datasets and find patterns in them. When it comes to deepfakes, this data is usually videos, pictures, and voice recordings of the person being copied.

In deepfakes, the system has been trained for hours to imitate the video footage, photos, or voice recordings of the person. Generative Adversarial Networks, or GANs, are the framework used in

the most advanced type of deepfake production. The generator and discriminator neural networks are the two opposing artificial networks that make up a GAN.

In an attempt to imitate real-life faces, voices, or behaviors, the generator produces fake content. In contrast, the discriminator evaluates the information and makes an effort to differentiate between what is real and what is not. The generator learns from its errors and creates content that is harder to identify as fake over time as these two networks compete with one another. The generator's ability to produce deepfakes that are almost identical to real content improves with the amount of data you feed it.

Deepfakes made their first appearance on a large scale in 2017, when a Reddit user posted modified pornographic material with the faces of Hollywood celebrities. Public outrage and a growing fear of technology misuse ensued. Soon after, another Reddit user published an app called FakeApp that made it simple for even inexperienced users to produce deepfakes.[1] [2] [3].

**How does it work?**

Using deep learning techniques, deepfakes are being created, primarily through the use of Generative Adversarial Networks (GANs). The process involves several steps:

1. **Collection of Data:** The Primary step is to gather a substantial amount of information about the subject. For example, if our subject is Hrithik Roshan, who portrays Anand Kumar in the film "Super 30," we would collect pictures, videos, and audio recordings of both Anand Kumar and Hrithik Roshan. This data will be used to train the AI model.

2. **Training of the Model:** In this step, the AI model analyzes the collected data for learning the object's facial features, voice patterns, expressions, and mannerisms. It then generates fake content based on this information, while a separate network detector checks the authenticity of the content and makes improvements until the fake content closely resembles the real content.

---

[1] Rakesh Mishra, *The Authenticity Challenge: Addressing the Concern of Producing Deepfake-Generated Media as Evidence in Courts*, The Crim. L. Blog, Nat'l L. Univ., Jodhpur (Apr. 5, 2025), https://criminallawstudiesnluj.wordpress.com/2025/04/05/the-authenticity-challenge-addressing-the-concern-of-producing-deepfake-generated-media-as-evidence-in-courts/.

[2] Gov't of Gujarat Cyber Awareness, *Deepfake Awareness* (n.d.), https://cawach.gujgov.edu.in/dist/documents/sop/cyberAwareness/Deepfake/.

[3] Priya Dabas, *The Impact of Deepfake Technology: Legal Risks and Regulatory Solutions*, MetaLegal (Nov. 26, 2024), https://www.metalegal.in/post/the-impact-of-deepfake-technology-legal-risks-and-regulatory-solutions.

3. **Synthesis and Manipulation:** Once the AI model is trained, it can create new content by covering the subject's face with a different person's face in a video or altering pre-existing footage to change what they say or do.
4. **Post-processing:** The artificial content is improved to make it look more real. This includes changing the lighting, shadows, and the way the lips move to match the speech to make the deepfake look more real.

For example, deepfakes can swap people's faces in a video, or make it seem like someone said or did something they never really did [4] [5].

**Authenticity of the Deepfake evidence**

There is a serious threat from deepfakes to the digital evidence being presented in Indian Courts. Judges and lawyers now have significant difficulties verifying media before allowing it as evidence because deepfakes can be almost identical to real photos and videos. In the case of ***Nirmaan Malhotra v. Tushita Kaul*** [6] The Delhi High Court has refused to accept the photographs that were submitted by the husband, alleging infidelity as a ground to deny maintenance to his wife.

The Court noted that it wasn't clear if the woman in the photos was indeed the wife and emphasized the need for concrete proof, stating:
"We may take judicial notice of the fact that we are living in the era of deepfakes."

Sections 62 and 63 of the Bhartiya Sakshya Adhiniyam, which are the primary legal provisions for electronic evidence in India, were drafted before the emergence of deepfakes. To authenticate digital records, they mostly rely on certificates and device production; however, these standards do not address whether the content is manipulated or fake.

Although some safeguards are added by more recent laws (such as the Bharatiya Sakshya Adhiniyam), such as expert certification and encryption values for files, there are still gaps in the

[4] *Delhi HC Urges Centre to Frame Law to Regulate AI Deepfake*, Hindustan Times (Apr. 2024), https://www.hindustantimes.com/india-news/delhi-hc-urges-centre-to-frame-law-to-regulate-ai-deepfake-101724853407493.html.
[5] Naman Bansal, Regulating Deepfakes in India: A Legal and Policy Analysis (SSRN Working Paper Series, Abstract No. 5153296, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5153296.
[6] Nirmaan Malhotra v. Tushita Kaul, MAT. APP. (F.C.) 180/2024 & CM Nos. 32316–32318/2024 (Del. HC May 28, 2024).

law. For example, it does not specify who is considered an expert or ensure that the content was not created by artificial intelligence from the start.

This indicates that current regulations are unable to keep up with the rapidly changing threat, which could lead to fake evidence, complicated court cases, and a decline in public confidence. To effectively handle deepfakes in evidence, India urgently needs clear, contemporary legislation, improved expert resources, and updated court procedures.[7].

**Legal Framework and rulings**

**The Information Technology Act of 2000's**

- **Section 66E** - Addresses privacy issues related to deepfakes used to collect, publish, or share an individual's photographs in public.
- **Section 66D** - By using a Computer Resource, it punishes identity fraud and online impersonation using AI-generated deepfakes.
- **Section 67** - Posting offensive Content Electronically, criminalizes the transmission of sexually explicit or offensive deepfake videos.
- **Section 72** - Breach of Confidentiality & Privacy: If a deepfake violates a person's privacy by sharing manipulated private images/videos.

**Provisions Under POCSO**

**Section 13** - Using a child in any form of media for sexual gratification commits an offence.

This includes:

1. Representation of a child's sexual organs;
2. The involvement of a child in actual or simulated sexual acts, whether or not penetration occurs, is of concern.
3. Indecent or obscene representations of a child.

**Provisions Under The Indecent Representation of Women (Prohibition) Act, 1986**

[7] *Admissibility of Deepfake Evidence in the Court of Law*, Lawful Legal (n.d.), https://lawfullegal.in/admissibility-of-deepfake-evidence-in-the-court-of-law/.

**Sections 4** - No person shall produce, sell, hire, distribute, circulate, or send by post any book, pamphlet, paper, slide, film, writing, drawing, painting, photograph, or any other representation that contains an indecent depiction of women in any form.

**Digital Personal Data Protection Act, 2023**

The Digital Personal Data Protection Act of 2023 enhances India's data protection framework by prioritizing individuals' rights over their personal data and establishing responsibilities for entities that misuse it.[8]

**Infamous Deepfake Cases**

1. **Rashmika Mandanna Case**

   At the end of 2023, a deepfake video featuring the well-known Indian actress Rashmika Mandanna emerged online. In the widely shared video, a woman resembling Rashmika was shown entering an elevator while dressed in revealing clothing.

   **What Happened**

   - Rashmika's face was expertly attached by the deepfake onto an irrelevant video of another woman.
   - Because of the excellent face-swapping, many viewers initially thought the video was real.
   - Digital investigators and fact-checkers verified it was a deepfake after it went viral.

2. **The False Pentagon Explosion That Disturbed the Stock Market**

   Panic erupted in May 2023 when a photo of an explosion near the Pentagon surfaced on social media. Several verified Twitter accounts shared the image, which appeared genuine. It depicted a plume of black smoke rising from a government building, prompting concerns about a possible missile strike or terrorist attack.

   **What happened?**

   Within minutes, the picture went viral after being featured on financial news providers and even briefly featured in mainstream media.

[8]*Controlling Deepfakes in India*, Chambers & Partners Practice Guides (n.d.), https://chambers.com/legal-trends/controlling-deepfakes-in-india.

The outcome? The U.S. stock market experienced a brief decline, with the S&P 500 decreasing before the rumor was disproved.

3. **The Deepfake Robocall Featuring Joe Biden That Targeted U.S. Voters.**

   Strange robocalls with President Joe Biden's voice were sent to voters in early 2024, right before the New Hampshire presidential primary. "Remain at home and save your vote for the November election," was the message.

   **What happened**

   - The robocalls imitated Biden's voice using AI voice cloning technology.
   - Democratic voters were urged not to cast ballots in the New Hampshire primary.
   - Election officials, cybersecurity specialists, and voters were all immediately alarmed by the incident.
   - The Federal Communications Commission (FCC) has announced an investigation into the use of AI-generated audio in robocalls.

4. **Fraud: The $35 Million Voice Deepfake Scam**

   In 2023, by using the AI-generated voice technology, the scammers impersonated the company's CEO, which led to the theft of $35 million.

   In a Hong Kong-based multinational corporation. During a crucial video call, a deepfake voice that is flawlessly imitated is used. Within minutes, millions were taken away.

   **What Happened**

   - An employee attended a video call with the chief financial officer of the company, along with other senior executives.
   - In between the calls, the CEO of the company instructed the employees to transfer the $35 million to the confidential accounts for an upcoming corporate acquisition.
   - Voices were matched, faces were familiar, and the instructions were obvious, so the call felt normal.
   - The funds were moved—and vanished [9].

**Personality Rights**

[9] Rishi Shetty, *Top Deepfake Incidents*, Arya.ai Blog (May 19, 2025), https://arya.ai/blog/top-deepfake-incidents.

The Deepfake was also violating the Personality Rights of a person. Personality Rights are also called the Right to Publicity, in which a person has the right to control the commercialization of their identity, which includes their name, image, likeness, or voice.

Unlike the right to privacy, which protects people from unwanted intrusions into their personal lives, personality rights focus on economic protection. They acknowledge that a person's identity can have significant commercial value.

These rights are particularly important for celebrities, whose name, image, or voice often carries substantial market value. Personality rights safeguard them from unauthorized commercial exploitation, such as using their identity in advertisements, merchandise, or promotions without their consent.

- **Rajat Sharma Case (2019)**

Zee Media launched a new news channel and published an advertisement stating "India Mein Rajat Ki Adalat Ab Band", indirectly targeting Rajat Sharma, a well-known news anchor.

**What Happened**

- Zee used Sharma's name and show title in a commercial advertisement.
- The statement was allegedly disparaging and misleading.
- Sharma claimed a violation of his right of publicity.

**Court's Decision**

- The decision given by the Delhi High Court is in favor of Rajat Sharma.
- The court held that his persona was clearly identifiable.
- Proof of confusion or deception was not required.

- **Anil Kapoor Case (2023)**

Anil Kapoor filed a suit in the Delhi High Court for the unauthorized use of his image, voice, domains, and the phrase "Jhakaas."

**What Happened**

- His likeness was misused online and through AI tools.

- Domain names resembling his name were registered.

**Court's Decision**

- The court granted interim relief.
- Domain names were transferred to Kapoor.
- Misuse of his persona and "Jhakaas" was restrained[10].

**Conclusion**

In today's world, seeing is no longer believing. Deepfakes can make anyone say or do anything on screen—and that puts truth on trial.

For courts and lawyers, this means more effort is needed to verify every digital video or image presented as evidence. A convincing clip could be fake, and real evidence could be dismissed as fake.

Other countries have recognized the urgent problem that deepfakes pose and have begun implementing specific laws to address it. India can learn from these examples to develop regulations that tackle these new technological challenges while respecting individuals' rights.

India has to carefully address these issues and safeguard its citizens from the risk occurring from deepfakes; it also has to promote innovation and freedom of expression. Only then can we ensure that digital evidence serves to reveal the truth, rather than generating confusion.

[10] Isheta T. Batra, *The Expanding Canvas: A Look at the Evolution of Personality Rights in India*, TBALaw (June 30, 2024), https://www.tbalaw.in/post/a-look-evolution-personality-rights-in-india.