DEEPFAKE IMAGES: THREAT TO DIGITAL DIGNITY, LEGAL CHALLENGES AND REGULATORY GAPS IN INDIA

Saumitra Sharma, United University

Soumya Srivastava, United University

Introduction

The word "Deepfake" refers to digital manipulation of content in which the appearance, voice and/or action of a person is changed by artificial means. Hence deepfakes are created images, videos or sounds which are fake but look real, created by using highly developed and advanced techniques by which human features are copied/imitated with great precision.

By deepfake user may replace one person's face with another in a video, or add speech in video as per requirement. Such alterations and changes are so real that it makes it difficult to detect. Therefore, deepfakes questions on reliability of digital evidence and also its authenticity.

So deepfake is nothing but synthetic media where content is artificially produced rather than recorded in the real world. Editing seems to be tiresome work previously demanding tons of effort and skill but deepfake had made it like knife on butter with precision and accuracy which can't be achieved with manual editing.

Deepfake originated in the year 2017 on Reddit. A user with the username "deepfake" began posting doctored videos where celebrities' faces were swapped with other individuals' bodies. These videos attracted viewers because they look real.[1]

The word "deepfake" is made of "deep learning" and "fake." Where "Deep learning" means intelligence that uses neural networks to process and learn from large datasets and "Fake" means something unreal or made artificially. So, they sum up as AI generated fake media.[2] Deepfake is not limited to only face swapping but it goes beyond and includes, lip-sync, cloning of voice and even completely imaginary faces can be crafted. Making it a requirement in the

[1] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021)

[2] Ian Goodfellow and others, 'Generative Adversarial Nets' (2014) Advances in Neural Information Processing Systems.

field of academia, legal and technology in the contemporary world.

Deepfake uses artificial intelligence to replicate and recreate human features. For the same large amount of real data, like images, videos, or audio recordings of a person are studied. Systems learn patterns such as facial expressions, voice tone, gestures, and movements.

The most important technique for deepfake creation is Generative Adversarial Networks (GANs). This involves two AI systems: one generates fake content, while the other tries to detect the reality of content. Symbiosis of these systems give highly realistic content. Achieving the goal of AI to give output as realistic as possible. Though this technology proved to be a boon in the field of education, entertainment, risk of misinformation, identity misuse and digital fraud always remain.[3]

Evolution

Evolution of deepfake technology is a requirement of broader development of digital media manipulation and artificial intelligence. Reliability in reality has been shaken due AI-generated videos. Maximum time AI generated video had roots as they are much simpler techniques that were initially used for creative and entertainment purposes.

Prior to the emergence of AI, image and video editing was tiring which requires skill and human intervention. Software's available on the internet at minimal or no cost provide users with altering video and images at their will. Morphing, rotoscoping, and CGI widely used in films and media which require expertise are now available for any and every individual and results are impressive, often lacking the seamless realism that modern deepfakes achieve.

This transition is the result of advancement in Machine Learning, developing algorithms capable of recognizing patterns in large datasets. Deep Learning uses neural networks modelled after the human brain to process complex data such as images, speech, and video that become gamechanger enabling machines to analyse facial features, expressions, and movements with remarkable accuracy.

[3] Ian Goodfellow and others, 'Generative Adversarial Nets' (2014)

Technological Foundations of Deepfakes

Deepfake technology is built by combining advanced computational techniques, in the field of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning. Knowledge of these concepts is necessary to understand how deepfake works.

Artificial Intelligence (AI) performs tasks which require human intelligence like recognizing faces, understanding speech, making decisions, and learning from data. AI encompasses several subfields, including machine learning and deep learning. In simple terms AI enables systems to analyse visual and audio data and replicate human features.[4]

Machine Learning (ML) is a subset of AI focusing on enabling machines to learn from data without being programmed for every task. They keep studying through a large set of data and deny working as instructed. Deep learning models can learn minute pattern and representation, making them effective for tasks like face recognition, voice synthesis, and image generation. In deepfakes, deep learning allows the system to generate highly realistic outputs that closely resembles. Inspired by the human brain, a computational model has been designed which works as a core in deep learning, were layers of interconnected nodes process information. This layered processing enables the system to build a detailed understanding of the input data. In deepfake applications, neural networks analyse facial movements and map them onto another person's face, creating a realistic illusion.[5]

Working Mechanism of Deepfakes

Deepfake technology operates through a combination of artificial intelligence (AI), deep learning, and large datasets to create highly realistic but doctored audio and visual content.

Can be achieved by collecting a large amount of data. Which includes images, videos, or audio recordings of the target person whose face or voice is to be replicated. Larger the data more accurately will be the final outcome.

Then data is fed into a deep learning model. These models analyse facial features, expressions, angles, lighting conditions, and voice patterns., Once the system learns to recognize and replicate these characteristics, new content can be generated where the target person appears to

[4] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021).
[5] OECD, OECD Principles on Artificial Intelligence (2019).

say or do things they never actually did. Generative Adversarial Networks (GANs) help in face swapping where two neural networks are trained where one is for the source face and another for the target face. During the swapping process, the encoded features of the target face are mapped onto the source video. For blending of swapped faces, lighting, movement of head, expression of face etc is adjusted minutely. The result is a seamless and realistic video.

Not only visuals, it also includes voice cloning where audio samples are used for cloning of voice. The system analyses speech patterns like tone, pitch, accent, rhythm, and pronunciation. After sufficient information and training, the model generates speech that closely mimics the original speaker.

Types of Deepfakes

All deepfakes use AI and deep learning, they differ in how they manipulate or generate content. The major types of deepfakes are face-swapping, voice deepfakes, lip-sync deepfakes, fully synthetic humans, and text-based deepfakes.

i. Face-Swapping Deepfakes

Face-swapping is a commonly recognized form of deepfake technology. Where the face of one person is digitally swapped with another in video or image. Algorithms study the facial features, expressions, and movements of the target face and swap onto the original video, making the result realistic. This technique is commonly used in entertainment, like movies and memes. Misleading and harmful content are also made through this, where controversies are planted by morphing video. Availability of astronomical dataset and advancement in AI models had enabled face-swapping felt authentic.[6]

ii. Voice Cloning Deepfakes

It refers to cloning or mimicking a voice of an individual by the use of an AI-based speech modulator. With the help of samples available and analysis of same makes system familiar and save the time, pitch, accent, pitch and style of person while speaking. Once enough data is accumulated any statement can be made in voice of targeted person. Voice assistants, dubbing,

[6] Danielle Keats Citron, The Fight for Privacy (WW Norton 2022)

and accessibility tools, use this hack and it prove to be very helpful.[7]

iii. Lip-Sync

Deepfakes indulge in lip-sync by altering movements of the mouth in a video to match a different audio track. In the entertainment industry movies shooted in one language can be dubbed in many languages by using lip-sync deepfake. It saves time, money and hard work of the production team. Lip movement of the speaker is altered according to the accent and situation demanded and the task is achieved.

iv. Fully Synthetic (AI-Generated Humans)

Deepfakes create synthetic humans which are completely new and had never existed ever before. By the use of GANs synthesis humans look real as photo, video or as a virtual influencer, where they speak of knowledge tirelessly. No real face is used so they are free to claim whatever the user is desired of and can be used in gaming, advertisement etc.

Uses of Artificial Intelligence

Due to misuse in misinformation, fraud, and privacy violations deepfake is always considered in wrong terms. However, when used responsibly, deepfakes can uplift creativity, improve accessibility, innovate education, and change industries like gaming, entertainment and marketing. In view of this optimism developing a balanced perspective is essential on the technology.

i. Entertainment Industry (Movies and Visual Effects)

One of the prominent and widely accepted uses of deepfake technology is in the entertainment industry. Deepfakes enable creation of highly realistic visual effects (VFX). Enabling de-aging and showing different phases of actor life. Lip-sync help in dubbing movies in various languages without compromising the naturality of acting and emotions behind the dialogue delivery and regional essence.[8]

Overall, deepfakes reduce production costs, save time, and expand creative possibilities,

[7] NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (2023

[8] David Marr, Artificial Intelligence in Practice (Wiley 2021).

making storytelling more immersive and realistic.

ii. Education and Training

Deepfake technology creates engaging and interactive learning experiences by simulating real-world scenarios that would otherwise be difficult, expensive, or dangerous to replicate. Deepfake creates historical figures and scenarios which make long lasting images in mind easy to understand and grasp for students, making learning more dynamic and indulging. Medical students can see more detailed visualization of subjects and have a better understanding of their subjects.[9]

iii. Accessibility (Voice Recreation and Dubbing)

One of the most impactful positive uses of deepfakes lies in improving accessibility, especially for individuals with disabilities. Voice cloning and facial animation technologies enhance ability to communicate. Recorded data being helpful in resynthesizing the natural voice of individuals who had lost their voice due to accident or incident or medical emergency.[10]

Misuse and Risks of Deepfakes

Images, audio and video are manipulated as per requirement with advancements in AI with precision and accuracy, this has undermined the reliability in digitalised media for authenticity and credibility. Raising privacy concerns in the system and also leading to erosion of trust in democratic system. Earlier images, videos and audio are considered as sole evidence but with the intervention of deepfakes it is not considered reliable anymore and their authenticity also questioned. This has by and large put a big question mark on the credibility of the authenticity of evidence and had made people think what could be alternate.

This is not only harming data but also is making a psychological impact on people. For example, if someone has morphed an individual's photo in an adult video, his character gets maligned and this makes him psychologically destroyed, though the individual's innocence is proved later but his character assassination and mental trauma remain forever. In another case if a politician speech is doctored and some ill statement is posted and aired during election campaign, public opinion will be impacted largely and people will follow video content and

[9] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021)
[10] World Health Organization, Global Report on Assistive Technology (2022).

made their mind as per the content, though it may happen to be proved innocence. Misinformation and propaganda are being spread by making fake videos making the credibility of media and social media non reliable. Privacy of people are at risk as individual can be targeted their face morphed and fall the victim of same. Social media has made the access to personal photo and video easier and this made the people to create deepfake video where victim got traumatized and break them emotionally, mentally, psychologically, distress, anxiety, and a sense of helplessness. Deepfake video made a person fall ill especially when video is explicit especially when target is female and presence of internet make it widely spread.[11]

Statement of Problems

The rise of deepfake technology and AI-generated manipulated media has highlighted major shortcomings in India's current legal system, which lacks the necessary tools to effectively control their misuse. These fabricated images pose serious risks to personal privacy, digital dignity, identity, reputation, and the right to control one's own information by allowing unauthorized alterations and distribution of private content. Moreover, the lack of precise legal definitions, regulatory oversight, and reliable enforcement mechanisms results in substantial gaps in how such violations are addressed. As a result, India urgently requires a dedicated and comprehensive legal framework to govern synthetic media created by artificial intelligence, establish clear accountability, uphold constitutional rights, and protect digital dignity in an increasingly digital world.

Legal Framework Governing Deepfakes in India

Manipulation in audio, video, and images is possible with advancements and innovation in deep learning and artificial intelligence, it has transformed the creation of digital content at the same time questioning the credibility of authenticity. Misuse of this technology has brought a serious issue of legality and ethics at the same time it seems to be a great help in the field of education and entertainment and its easy-to-use feature has made it user friendly.

In our country no such law is there to regulate the use of deepfakes which speak about the extent to which its use is limited. Even existing laws, whether cyber law, criminals' law or evidence law fail to speak clearly about misuse. Though laws to keep in check the misuse of

[11] Danielle Keats Citron, The Fight for Privacy (WW Norton 2022).

deepfakes technology and cybercrime required to be drafted and crafted before advancements of such technology, still our legal framework fails to look into such matters and require amendment at great extent to regulate offence and cybercrime resulting from misuses of deepfake. Such crimes are complex and its authenticity factor makes it difficult to sort out and solve the situation as a single victim might be suffering from the violation of privacy which comes from theft of individuals data leading to blackmail and fraud at same time. And the access of people to the internet makes the spread of content at mass level, even beyond borders making the legality more complex.

i. Criminal Law and Deepfakes

The Bhartiya Nyaya Sanhita 2023(BNS), is the basis of a legal framework which deals with offences related to deepfake. Though it doesn't clearly speak about digital or AI-related offences and fails to mention deception, harm and intent of act.

Defamation

As per Section 356 BNS defamation constitutes any false statement that lowers a person's moral/intellectual character, caste, calling, or credit in the estimation of others. By showing false statements or by enacting wrong acts, the reputation of an individual can be maligned beyond revival.[12]

Many actors like Amitabh Bachchan[13] and Anil Kapoor[14] have reached to the court to restrict use of their voice and face to restrict spreading misinformation and false statements leading to rumours and chaos.

One such case is Kamya Buch vs JIX5A[15]

Where an individual's non-consensual morphed image has been disseminated. Delhi High Court intervened into the matter and relieved the victim.

[12] Bharatiya Nyaya Sanhita 2023, s 356.
[13] Amitabh Bachchan v Rajat Nagi CS (Comm) 819/2022 (Delhi High Court).
[14] Anil Kapoor v Simply Life India CS (Comm) 652/2023 (Delhi High Court)
[15] Kamya Buch v JIX5A (Delhi High Court)

Cheating by Personation

Section 319 BNS deals with Cheating by Personation, is a fraud where deception is executed by pretending to be someone else. Though Digital Personation is the subject term of cyber law still Indian Court has given every individual Right to control one's identity.[16]

R. Rajagopal v. State of Tamil Nadu[17], where the Supreme Court agrees with an individual's right for safeguarding his/her privacy and control personal information publication. Here by using deepfake an individual's identity has been misused without consent.

Cheating

Section 318 BNS state cheating as deceiving a person and inducing them to deliver property [18]or consent to an act they would not otherwise agree to and Section 318(4) BNS provides punishment for cheating and dishonestly inducing delivery of property. Voice makeover and impersonation through video became easy with help of deepfake. It increases the risk of deceitfulness by making a trusted character making statement. Through impersonation one can deceit organisations by being senior authority and make fund transfers. All these acts are clear acts of wrongful gain, deception and inducement as of Section 318 BNS.

Sale/distribution/publication of obscene material

Section 294 BNS of the Bhartiya Nyaya Sanhita (BNS) prohibits sale, distribution, public exhibition, or circulation of obscene materials. Importing, exporting, or producing such materials is also prohibited under this law. Aiming to protect public morality, mainly explicit material which is tempting. Deepfake has made it easy to create deepfake and consensual explicit content.[18]

Aveek Sarkar v. State of West Bengal is one such case where Supreme Court goes with community standards test to find out obscenity. Deepfake pornography created without consent is considered obscene and unlawful under this standard raising serious concerns for dignity and violence based on gender.

[16] Bharatiya Nyaya Sanhita 2023, ss 318–319.
[17] R Rajagopal v State of Tamil Nadu (1994) 6 SCC 632.
[18] Bharatiya Nyaya Sanhita 2023, s 294.

Criminal Intimidation

Deepfakes may also be used for blackmail, threats, or harassment. Section 351 BNS define as threatening with injury in terms of person, property or reputation to cause alarm or force an action. Deepfake is used for blackmail, threat, and harassment.[19]

If a person is threatened by saying his/her morphed video will be released and told to not do so only when demand is met, this act is considered a crime.[20]

ii. The Information Technology Act, 2000

IT Act is the primary legislation governing cyber activities in India. This law provides provision which is solely applicable to deepfake-related crime.

Section 66C – Identity Theft

It penalises the forgery by use of electronic signature, password, or other such authentication features. Biometric data is also misused by using deepfake; to name a few features of face or pattern of voice, all these are considered as identity theft in this Section.

Section 66D – Cheating by Personation

It solely refers to cheating by personation using computer resources. Scams involving deepfake, whether impersonation through AI-generated audio or video, are all within this provision.

Sections 67 and 67A – Obscene Content

Criminalize the posting or sharing as well as publishing of obscene and sexually explicit content in electronic form. Strict penalties are levied on non-consensual deepfake pornography.

Section 72 – Breach of Confidentiality

Penalizes access to and disclosure of personal information without authorisation. Creating a deepfake using personal data without consent, constitutes a breach of confidentiality.[21]

[19] Bharatiya Nyaya Sanhita 2023, s 351
[20] Bharatiya Nyaya Sanhita 2023
[21] Information Technology Act 2000, ss 66C, 66D, 67, 67A, 72.

Judicial approach and Case Law

1.Justice K. S. Puttaswamy v. Union of India

In this landmark constitutional judgment, the Supreme Court of India unequivocally recognized the right to privacy as an intrinsic part of the fundamental rights guaranteed under Article 21 of the Constitution. The Court emphasized the concepts of informational self-determination, individual autonomy, and human dignity in the digital era. Although the case did not directly concern artificial intelligence or deepfakes, its broader constitutional reasoning provides a significant jurisprudential foundation for regulating AI-driven identity manipulation, unauthorized synthetic media, and misuse of personal attributes such as facial images, voice, and biometric characteristics. The judgment therefore serves as a crucial constitutional basis for protecting personality rights against technologically mediated exploitation.[22]

2.Shreya Singhal v. Union of India

The Supreme Court, while striking down Section 66A of the Information Technology Act, 2000, reaffirmed the constitutional importance of freedom of speech and expression in the online sphere. At the same time, the judgment highlighted the inherent difficulty in balancing civil liberties with the necessity of preventing digital harm and misuse. In the contemporary context of AI-generated misinformation and deepfake dissemination, the principles emerging from this decision illustrate the judicial challenge of designing regulatory mechanisms that effectively curb harmful synthetic content without disproportionately infringing upon constitutionally protected speech.[23]

Approaches by Different Countries

1.United States[24]

Certain states criminalize non-consensual deepfake pornography and election-related manipulated media.

[22] Justice KS Puttaswamy (Retd) v Union of India (2017) 10 SCC 1.
[23] Shreya Singhal v Union of India (2015) 5 SCC 1.
[24] Deepfake Accountability Act (proposed, US Congress) or relevant state statutes on deepfake election laws.

2.European Union[25]

The EU AI Act imposes transparency obligations and risk-based AI regulation.

3.China[26]

China's Deep Synthesis Regulations require clear labelling of AI-generated synthetic content.

Regulatory Gaps in India

Despite the increasing misuse of deepfake technology, India still lacks a comprehensive legal framework specifically regulating AI-generated synthetic media. Existing laws such as the Information Technology Act, 2000 and the Bharatiya Nyaya Sanhita address deepfake-related harms only indirectly and were enacted before the emergence of advanced generative AI technologies. Consequently, the present legal structure remains technologically inadequate and fragmented.[27]

1. Absence of Dedicated Deepfake Legislation

India currently has no specific legislation defining or regulating deepfakes, synthetic media, or AI-based identity manipulation. This creates uncertainty in enforcement and prosecution.[28]

2. Lack of Definitional Clarity

Existing statutes fail to clearly define concepts such as: synthetic media, AI manipulation, algorithmic impersonation, and digital identity misuse.

The absence of precise legal definitions weakens regulatory effectiveness and judicial interpretation.

3. Delayed Takedown Mechanisms

Victims frequently face delays in: reporting harmful content, obtaining platform response, and

[25] Artificial Intelligence Act (EU) 2024.
[26] Provisions on the Administration of Deep Synthesis of Internet Information Services (China, 2023)
[27] MeitY, Advisory on Deepfakes (26 December 2023)
[28] Digital Personal Data Protection Act 2023; Information Technology Act 2000.

securing removal of manipulated media.

Given the viral nature of digital platforms, delayed intervention often causes irreversible reputational harm.

4. Jurisdictional Challenges

Deepfake content may be created, circulated, and hosted across multiple jurisdictions, making investigation and enforcement increasingly difficult for Indian authorities.

5. Ambiguity in Accountability

The present legal framework does not clearly determine liability among: AI developers, users, intermediaries, and digital platforms.

This creates regulatory ambiguity regarding responsibility and compliance obligations.

6. Inadequate Victim Remedies

Victims often lack access to: speedy legal remedies, compensation mechanisms, psychological assistance, and effective identity restoration measures.

7. Difficulty in Tracing Offenders

Anonymous accounts, encrypted applications, and advanced AI tools make identification of deepfake creators technologically challenging.

8. Critical Analysis

India's current approach toward deepfake regulation remains reactive rather than preventive. Existing laws primarily address obscenity, defamation, or cyber offences but fail to adequately regulate AI-driven synthetic media. Therefore, there is an urgent need for a comprehensive legal framework balancing technological innovation with privacy, dignity, and digital rights protection.

Comparative Analysis of Deepfake Regulation Across Jurisdictions

A comparative study of different jurisdictions demonstrates that countries across the world

have adopted varying approaches toward the regulation of deepfakes and AI-generated content.

1. In India, the current legal framework primarily relies upon the Information Technology Act, 2000 and the Bharatiya Nyaya Sanhita. These laws provide general protection against cybercrimes, identity theft, defamation, and obscene digital content. However, the Indian framework does not specifically define or regulate deepfakes, resulting in significant legal ambiguity and enforcement challenges. The absence of dedicated legislation further weakens victim protection and platform accountability.

2. The European Union has adopted a comparatively comprehensive approach through the EU AI Act. The Act introduces transparency obligations, risk-based classification of AI systems, and disclosure requirements for AI-generated content. Its major strength lies in promoting accountability and ethical AI governance. Nevertheless, the framework has been criticized for its complex compliance requirements, which may impose a substantial burden upon smaller developers and technology companies.

3. In the United States, regulation largely exists at the state level rather than through a unified federal law. Certain states have enacted legislation targeting election-related deepfakes, digital impersonation, and non-consensual synthetic imagery. This approach allows states to address local concerns effectively; however, the lack of uniformity creates inconsistency in enforcement and legal protection across jurisdictions.

4. China has implemented one of the strictest regulatory models through its Deep Synthesis Regulations. These regulations mandate clear labelling of AI-generated or synthetically altered content and impose obligations upon online platforms to monitor and prevent misuse. While the framework strengthens traceability and content control, critics argue that it may also enable excessive governmental oversight and restrictions on freedom of expression.

Suggestions and Framework

Considering the rapid growth of deepfake technology and the limitations of existing laws, India requires a comprehensive and technology-specific regulatory framework. The legal response should focus not only on cybercrime prevention but also on the protection of privacy, dignity, and digital rights.

Dedicated Deepfake Legislation

India should enact a specific law regulating deepfakes, synthetic media, and AI-based identity manipulation with clear definitions and liability provisions.

Mandatory AI Disclosure

AI-generated or manipulated content should contain mandatory labels or watermarking to ensure transparency and prevent misinformation.

Stronger Platform Accountability

Digital platforms should be required to implement: rapid takedown mechanisms, grievance redressal systems, and proactive monitoring of harmful synthetic content.

Victim-Centric Remedies

The framework should provide: speedy legal remedies, compensation mechanisms, psychological support, and digital identity protection for victims.

AI Governance and Ethical Regulation

An independent regulatory authority should be established to monitor AI technologies and ensure ethical and responsible use of generative AI systems.

Digital Literacy and Public Awareness

Awareness programs should educate citizens regarding: identification of deepfakes, online misinformation, and protection of digital privacy.

International Cooperation

Since deepfake crimes often involve cross-border elements, stronger international cooperation is necessary for effective investigation and enforcement.

Concluding Suggestion

India must adopt a proactive and rights-based regulatory approach that balances technological

innovation with constitutional values of privacy, dignity, and individual autonomy.

Conclusion

The rapid growth of deepfake technology has emerged as a serious challenge to privacy, dignity, reputation, and digital security. Although artificial intelligence offers significant technological benefits, its misuse through AI-generated manipulated images has created new forms of cyber exploitation, misinformation, and online harassment.

This study highlights that the existing legal framework in India remains inadequate to effectively regulate deepfake-related harms. The absence of specific legislation, lack of clear definitions, weak enforcement mechanisms, and uncertainty regarding platform accountability continue to create major regulatory gaps.[29]

The research further demonstrates that deepfake misuse directly affects constitutional values such as privacy, dignity, and personal autonomy protected under Article 21 of the Constitution of India. Therefore, India requires a comprehensive and technology-specific regulatory framework combining legal reform, platform responsibility, victim protection, and digital awareness measures.[30]

In conclusion, deepfakes are not merely a technological issue but a broader challenge to digital dignity and democratic integrity in the contemporary digital era.

[29] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2021).
[30] OECD AI Principles (2019).