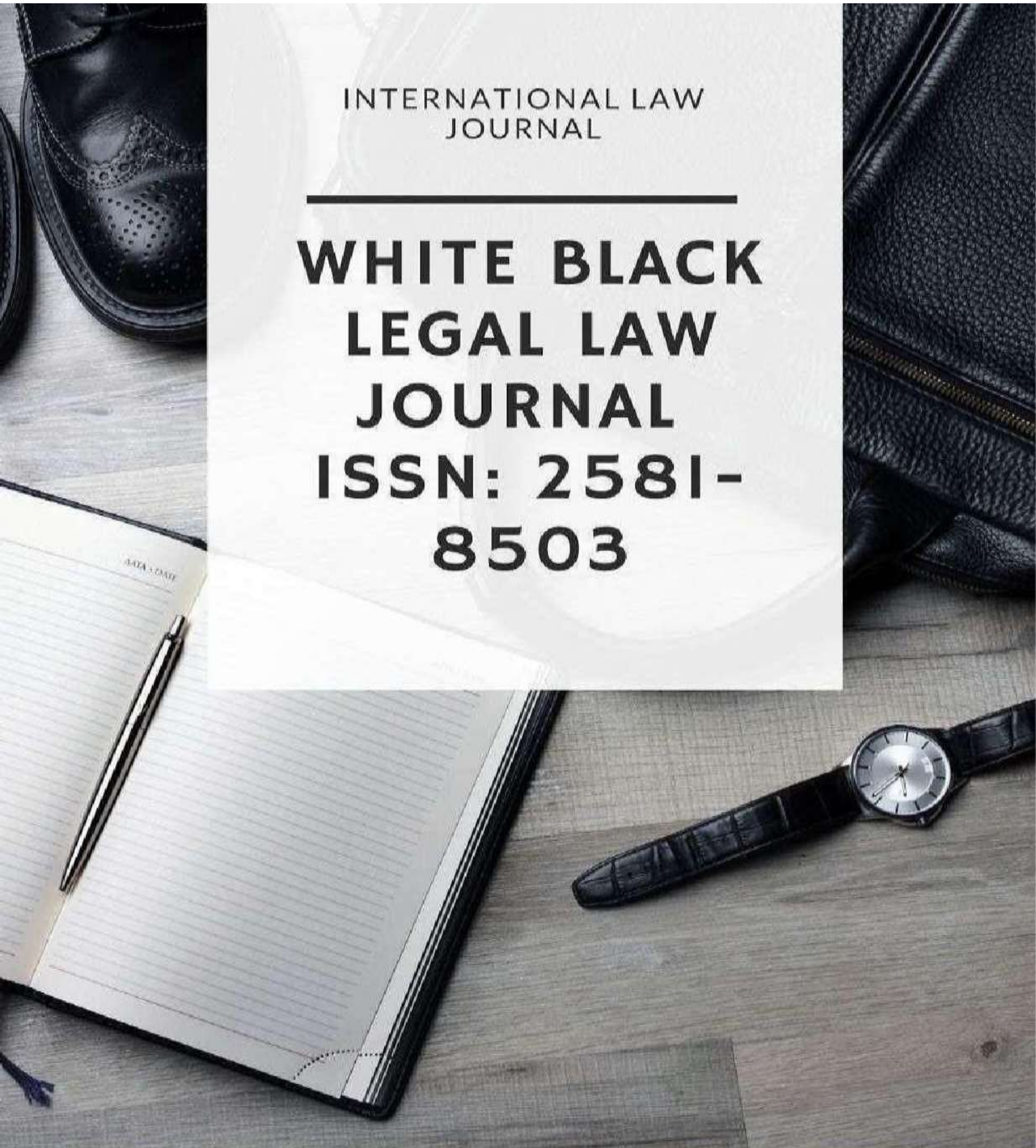

INTERNATIONAL LAW JOURNAL

WHITE BLACK LEGAL LAW JOURNAL ISSN: 2581-8503

Peer - Reviewed & Refereed Journal

The Law Journal strives to provide a platform for discussion of International as well as National Developments in the Field of Law.

WWW.WHITEBLACKLEGAL.CO.IN

DISCLAIMER

No part of this publication may be reproduced or copied in any form by any means without prior written permission of Editor-in-chief of White Black Legal – The Law Journal. The Editorial Team of White Black Legal holds the copyright to all articles contributed to this publication. The views expressed in this publication are purely personal opinions of the authors and do not reflect the views of the Editorial Team of White Black Legal. Though all efforts are made to ensure the accuracy and correctness of the information published, White Black Legal shall not be responsible for any errors caused due to oversight or otherwise.

EDITORIAL TEAM

Raju Narayana Swamy (IAS) Indian Administrative Service officer

Dr. Raju Narayana Swamy popularly known as Kerala's Anti Corruption Crusader is the All India Topper of the 1991 batch of the IAS and is currently posted as Principal Secretary to the Government of Kerala . He has earned many accolades as he hit against the political-bureaucrat corruption nexus in India. Dr Swamy holds a B.Tech in Computer Science and Engineering from the IIT Madras and a Ph. D. in Cyber Law from Gujarat National Law University . He also has an LLM (Pro) (with specialization in IPR) as well as three PG Diplomas from the National Law University, Delhi- one in Urban Environmental Management and Law, another in Environmental Law and Policy and a third one in Tourism and Environmental Law. He also holds a post-graduate diploma in IPR from the National Law School, Bengaluru and a professional diploma in Public Procurement from the World Bank.

Dr. R. K. Upadhyay

Dr. R. K. Upadhyay is Registrar, University of Kota (Raj.), Dr Upadhyay obtained LLB , LLM degrees from Banaras Hindu University & Phd from university of Kota.He has succesfully completed UGC sponsored M.R.P for the work in the ares of the various prisoners reforms in the state of the Rajasthan.

Senior Editor

Dr. Neha Mishra

Dr. Neha Mishra is Associate Professor & Associate Dean (Scholarships) in Jindal Global Law School, OP Jindal Global University. She was awarded both her PhD degree and Associate Professor & Associate Dean M.A.; LL.B. (University of Delhi); LL.M.; Ph.D. (NLSIU, Bangalore) LLM from National Law School of India University, Bengaluru; she did her LL.B. from Faculty of Law, Delhi University as well as M.A. and B.A. from Hindu College and DCAC from DU respectively. Neha has been a Visiting Fellow, School of Social Work, Michigan State University, 2016 and invited speaker Panelist at Global Conference, Whitney R. Harris World Law Institute, Washington University in St.Louis, 2015.

Ms. Sumiti Ahuja

Ms. Sumiti Ahuja, Assistant Professor, Faculty of Law, University of Delhi,

Ms. Sumiti Ahuja completed her LL.M. from the Indian Law Institute with specialization in Criminal Law and Corporate Law, and has over nine years of teaching experience. She has done her LL.B. from the Faculty of Law, University of Delhi. She is currently pursuing Ph.D. in the area of Forensics and Law. Prior to joining the teaching profession, she has worked as Research Assistant for projects funded by different agencies of Govt. of India. She has developed various audio-video teaching modules under UGC e-PG Pathshala programme in the area of Criminology, under the aegis of an MHRD Project. Her areas of interest are Criminal Law, Law of Evidence, Interpretation of Statutes, and Clinical Legal Education.

Dr. Navtika Singh Nautiyal

Dr. Navtika Singh Nautiyal presently working as an Assistant Professor in School of law, Forensic Justice and Policy studies at National Forensic Sciences University, Gandhinagar, Gujarat. She has 9 years of Teaching and Research Experience. She has completed her Philosophy of Doctorate in ‘Intercountry adoption laws from Uttranchal University, Dehradun’ and LLM from Indian Law Institute, New Delhi.

Dr. Rinu Saraswat

Associate Professor at School of Law, Apex University, Jaipur,
M.A, LL.M, Ph.D,

Dr. Rinu have 5 yrs of teaching experience in renowned institutions like Jagannath University and Apex University. Participated in more than 20 national and international seminars and conferences and 5 workshops and training programmes.

Dr. Nitesh Saraswat

E.MBA, LL.M, Ph.D, PGDSAPM
Currently working as Assistant Professor at Law Centre II, Faculty of Law, University of Delhi. Dr. Nitesh have 14 years of Teaching, Administrative and research experience in Renowned Institutions like Amity University, Tata Institute of Social Sciences, Jai Narain Vyas University Jodhpur, Jagannath University and Nirma University.
More than 25 Publications in renowned National and International Journals and has authored a Text book on Cr.P.C and Juvenile Delinquency law.

Subhrajit Chanda

BBA. LL.B. (Hons.) (Amity University, Rajasthan); LL. M. (UPES, Dehradun) (Nottingham Trent University, UK); Ph.D. Candidate (G.D. Goenka University)

Subhrajit did his LL.M. in Sports Law, from Nottingham Trent University of United Kingdoms, with international scholarship provided by university; he has also completed another LL.M. in Energy Law from University of Petroleum and Energy Studies, India. He did his B.B.A.LL.B. (Hons.) focussing on International Trade Law.

ABOUT US

WHITE BLACK LEGAL is an open access, peer-reviewed and refereed journal providededicated to express views on topical legal issues, thereby generating a cross current of ideas on emerging matters. This platform shall also ignite the initiative and desire of young law students to contribute in the field of law. The erudite response of legal luminaries shall be solicited to enable readers to explore challenges that lie before law makers, lawyers and the society at large, in the event of the ever changing social, economic and technological scenario.

With this thought, we hereby present to you

DEEPFAKE REGULATION - BRIDGING THE GAP BETWEEN THE PRESENT REALITIES AND FUTURE IMPERATIVES.

AUTHORED BY - SRI SAI SANTOSH R AND NITHYASRI G

Abstract

This study explores the complex terrain of deepfake technology and its regulatory landscape in the Indian setting. Deepfakes, hyper-realistic synthetic media, have emerged as a pressing concern, challenging the authenticity of visual and auditory information.

This paper examines the present state of deepfake regulation in India and envisions a legal framework that aligns with emerging realities. Beginning with an analysis of the current situation, the paper explores the proliferation of deepfake technology in India's digital ecosystem and its multifaceted implications. It evaluates the existing Indian cyber laws, including the Information Technology Act, of 2000, in light of their applicability to deepfake-related offences. The examination underscores the need for precise legal definitions to effectively address the evolving landscape of synthetic media.

Crucially, the research paper investigates the practical challenges posed by deepfakes, ranging from privacy infringements to potential identity theft and electoral manipulation. Case studies of notable deepfake incidents in India illuminate the real-world legal complexities and provide valuable insights into the scope and limitations of existing legislation.

However, the paper goes beyond the status quo and discusses the imperative "ought to be." It envisions a legal framework for deepfake regulation that considers the rapid evolution of technology. This envisioned framework is designed to bridge the gap between present realities and future imperatives, safeguarding the integrity of digital content and preserving the rights and privacy of Indian citizens.

In conclusion, this research paper conducts a thorough investigation into the intricacies of deepfake technology regulation in India, addressing both the existing challenges and the evolving landscape. By envisioning a legal framework that aligns with emerging technological realities, it aims to contribute to the discourse on tackling this evolving issue, ensuring that the legal landscape remains robust and adaptable in the face of a dynamic digital age.

Key Words: Deepfake, Legal Framework, Deepfake Regulations, Legislations, Proposed Framework.

Introduction

In a recent and sensational development, the issue of a deepfake video featuring popular actor Rashmika Mandanna has captured widespread attention.[1] The video, which circulated widely on social media, prompted a swift response from Union Minister of State for IT, Rajeev Chandrasekhar. The minister, taking to a prominent platform, emphasized the government's unwavering commitment, under the leadership of Prime Minister Narendra Modi, to ensuring the safety and trust of all citizens navigating the digital landscape. Chandrasekhar's tweet not only addressed the specific incident but also underscored the broader dedication to addressing the challenges posed by deceptive technologies in the online realm.[2] This incident serves as a poignant illustration of the urgency and importance of regulating deepfake content, a critical aspect that this research paper aims to explore in-depth.

The term "deepfake" originates from the combination of "deep learning" and "fake." Deep learning is a category of machine learning algorithms utilising artificial neural networks, inspired by the functioning of the human brain. Deepfake technology specifically employs deep learning methods, notably generative adversarial networks (GANs), to produce highly realistic and persuasive counterfeit videos or images.[3]

[1] Jignasa Singha, (2023, November 30). Rashmika Mandanna deepfake video probe hits a wall. The IndianExpress.https://indianexpress.com/article/cities/delhi/rashmika-mandanna-deepfake-video-probe-us-tech-firms-not-sharing-data-police-9047200/

[2] Chandrajith Mithra, (2023, November 6).Rashmika Mandanna's Viral Deepfake Prompts Big Warning From IT Minister. NDTV.https://www.ndtv.com/india-news/rashmika-mandannas-deepfake-goes-viral-rajeev-chandrasekhar-posts-digital-rules-4549472

[3] Ian Sample.(2020, January 13).What are deepfakes – and how can you spot them? The Guardian. https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them

The concept of deepfake technology gained prominence circa 2017, with the term itself coined by a Reddit user who applied it to describe a specific kind of AI-generated, face-swapping pornography. Nonetheless, the foundations of deepfake technology can be traced back to earlier advancements in machine learning and computer graphics.

Generative adversarial networks (GANs), a crucial element of deepfakes, were introduced by Ian Goodfellow and colleagues in 2014. GANs consist of two neural networks – a generator and a discriminator – trained concurrently through adversarial training. The generator produces synthetic data, and the discriminator evaluates it against real data. This iterative process continues until the generator generates data that is virtually indistinguishable from real data. As deep learning techniques progressed, so did the capabilities of deepfake technology. While deepfakes can be entertaining for benign purposes, they also pose ethical concerns and carry the potential for misuse, including spreading misinformation, manipulating public opinion, or creating convincing fake content for malicious purposes. Consequently, there have been efforts to develop tools and techniques to detect and mitigate the impact of deepfakes.

How does it function?

Deepfake technology operates through a sophisticated interplay of machine learning, artificial intelligence, and neural networks, with a particular emphasis on Generative Adversarial Networks (GANs). The process unfolds in several key stages:

1. Data Collection:

- The initiation of the deepfake creation process involves the collection of an extensive dataset related to the targeted individual. This dataset comprises diverse images or videos capturing the person from various angles, under different lighting conditions, and displaying different facial expressions.

2. Training the Model:

- The gathered dataset is utilised to train a deep learning model, often leveraging the capabilities of GANs. GANs consist of two neural networks: a generator and a discriminator.
- The generator's role is to produce synthetic content, such as fabricated images or videos, drawing

inspiration from the training data.

- The discriminator evaluates the generated content against authentic content, aiming to distinguish between the two. The overarching objective is for the generator to create content that is virtually indistinguishable from genuine content, thus deceiving the discriminator.[4]

3. Iterations and Refinement:

- The training process involves numerous iterations of this adversarial interplay between the generator and discriminator. Over time, the generator refines its ability to generate synthetic content that closely mirrors the patterns present in the original data.[5]

4. Fine-Tuning:

- To enhance the realism of the deepfake, the model undergoes further fine-tuning, often incorporating specific target images or videos. This fine-tuning ensures that the generated content aligns with the unique characteristics and nuances of the target individual.

5. Application:

- Once the model reaches a satisfactory level of training, it can be applied to either new or existing images and videos. For instance, the face of the targeted individual in a video can be seamlessly replaced with a synthetic face generated by the model. Similarly, the model may be employed for voice synthesis to mimic the speech patterns of the target individual.

6. Post-Processing (Optional):

- Some deepfake creators may opt for post-processing techniques to further elevate the realism of the generated content. This could involve adjustments to lighting, colour grading, or other visual elements to create a more natural appearance for the deepfake.

[4] Generative Models: Types + Role in Generating Synthetic Data. Questions Pro.https://www.questionpro.com/blog/generative-models/

[5] Unleashing the Power of Generative Adversarial Networks (GANs). The Medium. https://medium.com/@mysterious_obscure/unleashing-the-power-of-generative-adversarial-networks-gans-1a02018dbd5

Types of deepfakes[6]

1. Face-Swapping deepfakes:

- Involving the seamless replacement of faces in videos or images, face-swapping deepfakes utilise deep learning algorithms to analyse and transpose facial features and expressions from one individual onto another.
- This technology has found application in the entertainment industry, allowing filmmakers to convincingly replace actors' faces for certain scenes. However, the potential for malicious use in spreading false information or creating misleading content is a significant concern.

2. Voice Cloning:

- Through deep learning techniques, voice cloning deepfakes replicate a person's voice, mimicking their tone, cadence, and other vocal characteristics to make them appear to say things they did not.
- Voice cloning deepfakes can be employed in various scenarios, from creating fraudulent audio recordings for extortion to impersonating individuals in phone calls. The potential for voice-based social engineering attacks underscores the need for robust authentication measures.

3. Text-to-Speech (TTS) Deepfakes:

- This category involves the conversion of text into spoken words using deep learning to generate synthetic audio with a voice that emulates a specific individual.
- Text-to-speech deepfakes find applications in creating synthetic voices for virtual assistants, audiobooks, or even voiceovers. However, the potential misuse for generating deceptive audio content raises concerns about the trustworthiness of synthetic voices in various contexts.

4. Image-to-Image Translation:

- Deep learning models are used to translate images from one style to another, such as transforming a daytime scene into a nighttime setting, changing the season in a landscape, or altering the appearance of objects.
- Image-to-image translation deepfakes have applications in art, design, and visual effects, allowing

[6] Chiradeep Basumallick. (2022, May 23). What Is Deepfake? Meaning, Types of Frauds, Examples, and Prevention Best Practices for 2022. Spice works. https://www.spiceworks.com/it-security/cyber-risk-management/articles/what-is-deepfake/

for creative transformations of images. However, the risk lies in the potential for creating deceptive visuals that could be used to spread false narratives or manipulate the perception of events.

5. Gesture and Movement Replication:

- Some deepfakes focus on replicating specific gestures, movements, or body language from one person to another, extending beyond facial features to include the nuanced movements and reactions exhibited in a video.
- Gesture and movement replication deepfakes can be challenging to detect, as they involve a comprehensive understanding of body language and non-verbal cues. The potential for creating misleading videos by altering the behaviour of individuals poses risks in various fields, including politics and interpersonal communication.

6. Interactive deepfakes:

- This category explores the creation of AI-driven, interactive characters capable of responding to users in real-time. Although more experimental, it holds potential applications in areas such as gaming and virtual environments.
- Interactive deepfakes raise intriguing possibilities in creating dynamic and responsive virtual characters. However, concerns about ethical use, potential misuse in online interactions, and the psychological impact on users warrant careful consideration as this technology evolves.

Implications and Hazards of Deepfake Technology

Deepfake technology has ushered in a myriad of concerns across various domains, presenting risks that extend from individual privacy to broader societal and political implications.

1. Misinformation and Deception:

- Deepfakes have the potential to propagate misinformation by convincingly portraying individuals engaging in activities they never participated in. This deceptive use of technology can manipulate public perception and distort the truth.
- This misinformation can be particularly harmful in areas such as news reporting, where deepfakes could be used to create fabricated events or statements that mislead the public. The rapid dissemination of false information through social media amplifies the impact, making it challenging

to correct the narrative once the deepfake has circulated widely.

2. Political Manipulation:

- The use of deepfakes poses a substantial threat to political landscapes. Politicians may become targets of fabricated content, harming their reputations and influencing public opinion. This manipulation can have profound consequences for democratic processes and erode trust in political institutions.
- Political deepfakes could be strategically deployed during elections to sway public sentiment or damage the credibility of political figures. The potential for deepfake-driven political polarisation raises concerns about the integrity of democratic systems, demanding vigilant measures to safeguard against such manipulations.

3. Identity Theft and Harassment:

- Deepfake technology can be maliciously employed for identity theft, involving the insertion of an individual's likeness into explicit or compromising scenarios. This form of digital harassment can lead to severe personal and professional repercussions.
- Instances of deepfake-driven identity theft can extend beyond explicit content, infiltrating personal and professional networks. Cybercriminals may exploit these fabricated personas for financial fraud, jeopardizing not only the individual's reputation but also their financial security.

4. Privacy Concerns:

- The manipulation of visual and auditory information through deepfakes raises critical privacy concerns. By generating fabricated content involving private individuals, deepfakes violate personal boundaries and inflict emotional distress.
- Privacy breaches resulting from deepfakes may extend to sensitive personal moments, impacting an individual's relationships, mental well-being, and overall quality of life. The psychological toll of knowing that one's private life can be manipulated and exploited underscores the urgent need for comprehensive privacy safeguards.

5. Security Risks:

- Deepfakes can be weaponized for nefarious purposes, such as creating false audio or video recordings to falsely implicate individuals in criminal activities. This poses both legal consequences and reputational damage.

- Beyond legal ramifications, the security risks associated with deepfakes may extend to corporate espionage or targeted attacks on individuals in positions of influence. The potential for malicious actors to leverage deepfakes as a tool for corporate or political sabotage underscores the importance of robust security measures.

6. Erosion of Trust:

- The prevalence of deepfake technology contributes to the erosion of trust in media and digital content. As awareness of manipulated content grows, scepticism increases, affecting the credibility of information sources.
- The erosion of trust can lead to a broader societal impact, diminishing confidence in various forms of media, including journalism, entertainment, and even interpersonal communication. Rebuilding this trust requires concerted efforts from both technological and educational perspectives.

7. Financial Fraud:

- In corporate settings, deepfakes may be exploited to impersonate executives, potentially leading to financial fraud. For instance, criminals might use deepfake audio to mimic a CEO's voice and orchestrate unauthorized fund transfers.
- The financial implications of deepfake-driven fraud extend to stock market manipulation, where false statements from seemingly reputable sources can influence investment decisions. Safeguarding against such fraudulent activities requires heightened vigilance within corporate communication channels and financial transactions.

8. National Security Threats:

- Deepfakes present risks to national security by enabling the creation of fabricated content capable of influencing geopolitical events or inciting unrest. The deployment of deepfakes during times of crisis can escalate tensions and lead to unforeseen consequences.
- National security threats from deepfakes encompass potential use in disinformation campaigns aimed at sowing discord within populations or destabilising diplomatic relations. The strategic deployment of deepfakes by state actors necessitates international cooperation to address these threats

comprehensively.

9. Challenges for Law Enforcement:

- The dynamic nature of deepfake technology poses challenges for law enforcement in identifying and prosecuting those responsible for deceptive content creation. The anonymity prevalent on online platforms further complicates efforts to trace perpetrators.
- Law enforcement faces not only technical challenges in identifying the origin of deepfakes but also legal complexities in prosecuting individuals across jurisdictions. The collaborative development of international legal frameworks and technological tools is crucial to effectively address these challenges.

10. Social and Psychological Impact:

- Exposure to deepfake content, especially when maliciously used, can have psychological repercussions. The fear of being targeted or uncertainty regarding the authenticity of digital content contributes to anxiety and stress.
- The psychological impact of deepfakes extends to societal trust issues, impacting how individuals engage with digital content and each other. Recognizing and addressing the psychological toll on individuals and communities is essential for fostering a resilient and informed society.

Legal Framework

Information Technology Act, 2008

Section	Provision	Applicability to Deepfake
43A - Compensation for failure to protect data	"Where a body corporate, possessing, dealing, or handling any sensitive personal data or information in a computer resource which it owns, controls or operates, is negligent in implementing and maintaining reasonable security practices and	This section focuses on compensation for failure to protect sensitive personal data. In the context of deepfakes, if an individual's personal data is misused or compromised, there may be grounds to seek compensation.

	procedures and thereby causes wrongful loss or wrongful gain to any person, such body corporate shall be liable to pay damages by way of compensation to the person so affected."	
66C - Identity Theft	"Whoever, fraudulently or dishonestly make use of the electronic signature, password, or any other unique identification feature of any other person, shall be punished with imprisonment of either description for a term which may extend to three years and shall also be liable to fine which may extend to rupees one lakh."	If deepfake technology is used for identity theft, where a person's identity is misrepresented, Section 66C could be invoked.
66D - Cheating by personation by using a computer resource:	"Whoever, by means of any communication device or computer resource cheats by personation, shall be punished with imprisonment of either description for a term which may extend to three years and shall also be liable to fine which may extend to one lakh rupees."	This section could be applicable if deepfakes are used for cheating or impersonation.
66E - Violation of Privacy	"Whoever, intentionally or knowingly captures, publishes or transmits the image of a private area of any person without his or her consent, under circumstances violating the privacy of that person, shall be punished with imprisonment which may extend to	If deepfakes involve the capturing, publishing, or transmitting of images of a private area of any person without their consent, Section 66E may apply.

	three years or with fine not exceeding two lakh rupees, or with both."	
67 - Publishing or transmitting obscene material in electronic form:	"Whoever publishes or transmits or causes to be published or transmitted in the electronic form, any material which is lascivious or appeals to the prurient interest or if its effect is such as to tend to deprave and corrupt persons who are likely, having regard to all relevant circumstances, to read, see or hear the matter contained or embodied in it, shall be punished on first conviction with imprisonment of either description for a term which may extend to three years and with fine which may extend to five lakh rupees and in the event of a second or subsequent conviction with imprisonment of either description for a term which may extend to five years and also with fine which may extend to ten lakh rupees."	If deepfakes involve the publication or transmission of obscene material, Section 67 may be applicable.
69 - Power to issue directions for interception or monitoring or decryption of any information through any computer resource:	(1) Where the Central Government or a State Government or any of its officers specially authorised by the Central Government or the State Government, as the case may be, in this behalf may, if satisfied that it is necessary or expedient to do so in the interest of the sovereignty or integrity of India, defence of India, security of	If deepfake content poses threats to national security, this section may be invoked for interception or monitoring.

	the State, friendly relations with foreign states or public order or for preventing incitement to the commission of any cognizable offence relating to above or for investigation of any offence, it may, subject to the provisions of sub-section (2), for reasons to be recorded in writing, by order, direct any agency of the appropriate Government to intercept, monitor or decrypt or cause to be intercepted or monitored or decrypted any information generated, transmitted, received or stored in any computer resource.	
72 - Breach of confidentiality and privacy	"Save as otherwise provided in this Act or any other law for the time being in force, no person who has secured access to any electronic record, book, register, correspondence, information, document or other material without the consent of the person concerned discloses such electronic record, book, register, correspondence, information, document or other material to any other person shall be punished with imprisonment for a term which may extend to two years, or with fine which may extend to one lakh rupees, or with both."	If deepfake involves the breach of confidentiality or privacy, this section may be applicable.

Other Laws

The Constitution of India, 1950

ARTICLE 21- In the landmark case of *Justice K. S. Puttaswamy v. Union of India,*[7] a nine-judge bench affirmed the fundamental right to privacy under Part III of the Constitution of India. This right is pivotal in protecting individuals from both State and non-state actors who might infringe upon their informational privacy. The judgement emphasises an individual's authority over their personal and digital privacy, recognizing it as a cornerstone of constitutional protection.

This ruling was a response to the argument presented by the Attorney General of India (AGI), who contended that the Constitution of India did not explicitly include a fundamental right to privacy, particularly in connection to the legal challenges surrounding the Aadhaar Card, India's national identity project. The court, however, unequivocally established the existence of the fundamental right to privacy, underscoring an individual's control over their personal and digital privacy.

The implications of this judicial recognition extend to cases involving the creation of non-consensual deepfake content, where private or personal information such as images or video clips is illicitly utilised. Such actions constitute a direct violation of the fundamental right to privacy enshrined in the Indian Constitution.

Moreover, Section 66E of the Information Technology (IT) Act, 2000,[8] addresses the punishment for breaching the fundamental right to privacy. Individuals who knowingly or intentionally click, capture, publish, or transmit an image of a private area of another person without their express or implied consent may face imprisonment up to three years, a fine not exceeding two lakh rupees, or both. This legal provision serves as a deterrent against unauthorised use of private information, reinforcing the constitutional protection of an individual's right to privacy.

In the context of deepfake videos, their widespread use poses a significant threat to individual privacy. The Supreme Court's acknowledgment of the right to privacy in the Justice KS Puttaswamy case highlights the autonomy of individuals and their right to control various aspects of their lives. The

[7] Justice K.S. Puttaswamy (Retd.) v Union of India and Ors (2017) 10 SCC 1.

[8] Information and Technology Act, 2000.Section 66 E

court stressed that privacy is integral to a democratic state, safeguarding individual autonomy, and permitting interference only when just and fair, not arbitrary or oppressive.

The Right to Privacy, as recognized by the Supreme Court, encompasses protection in both physical and mental realms, acknowledging the importance of shielding one's life from public scrutiny. This protection extends to an individual's decision-making authority regarding what information about themselves is released into the public space.

Consequently, it is evident that deepfake videos infringe upon the Right to Privacy in multiple dimensions. While there might not be a specific law expressly prohibiting them, legal claims against such violations would likely succeed in a court of law.

Indian Penal Code, 1860

SECTION 499 AND 500 - A person in India can face liability for defamation under both criminal and civil laws.[9] The legal framework encompasses Section 499 of the Indian Penal Code (IPC), 1860, which defines defamation as a bailable, non-cognizable, and compoundable offence. It includes the publication of information intending to harm the reputation of an individual. Section 500 of the IPC prescribes punishment for defamation, entailing imprisonment for up to two years, a fine, or both.

Earlier, cyber defamation was addressed by Section 66A of the Information Technology (IT) Act, which criminalised the transmission of offensive information with the intent to cause harm, insult, injury, hatred, criminal intimidation, or ill will. However, this provision was invalidated by the Supreme Court in the case of *Shreya Singhal v. Union of India.*[10]

SECTION 468 - Deep Fake content can potentially fall under the purview of Section 468 of the Indian Penal Code (IPC), which defines forgery. Since deepfake videos are typically forged or copied versions of original works and are created with the intent to harm the reputation or image of any party, they may constitute the offence of forgery. Per Section 468, individuals knowingly involved in

[9] T. Pradeep & Aswasthy Rajan, A Critical Study on Cyber Defamation and Liability of ISPS, 119(17) International Journal of Pure and Applied Mathematics, 1717, 1719 (2018)https://acadpubl.eu/hub/2018-119- 17/2/139.pdf

[10] Shreya Singhal v. Union of India (2013) 12 S.C.C. 73

creating such content can face imprisonment for up to three years and may be liable for a fine.

SECTION 124 - Additionally, Section 124 of the IPC applies to punish acts where deepfake content is intended to spread hatred, contempt, or excite disaffection towards the Government of India. Offences falling under this section could lead to charges of sedition.

SECTION 506 - Section 506 of the IPC provides punishment for offences involving the use of videos or images that threaten or intimidate any person, their property, or their reputation. This section can be applicable to cases where deepfake content is used for intimidation or threats.

Furthermore, deepfake content that tends to provoke a breach of public peace and order, promotes communal outrage, or fosters enmity between religious or linguistic groups based on caste, religion, race, language, place of birth, or malicious acts to hurt religious feelings may be subject to legal action. Such actions could be prosecuted under relevant sections of the IPC aimed at maintaining public harmony and preventing the incitement of hatred between different groups.

The Personal Data Protection Bill 2019

The Personal Data Protection Bill of 2019 is designed to protect individuals' privacy and data by preventing unauthorised access and processing of personal information, including images, without explicit consent.

Concerning its impact on deepfake videos, the bill's protective scope extends both territoriality and extraterritoriality. It broadly defines personal data, covering any characteristic of a natural person. Data fiduciaries, whether state entities, individuals, or companies, are obligated to process data for lawful purposes, avoiding misleading practices. Consent is mandatory for processing individuals' data, and unauthorised disclosure leading to a privacy breach is considered a personal data breach.[11]

SECTION 20 - Section 20 introduces a pivotal provision, the 'right to be forgotten,' empowering individuals to cease and erase unauthorised personal data from the public domain. Moreover the right

[11] Parth Tyagi and Achyutam Bhatnagar, Deepfakes and the Indian legal landscape, INFORM.ORG,(July 03, 2020)https://inforrm.org/2020/07/03/deepfakes-and-the-indian-legal-landscape-parth-tyagi-and-achyutambhatnagar/

to be forgotten has already been recognised by some of the High Courts, such as Delhi, Kerala and Karnataka.[12] This provision is especially relevant in cases like revenge porn. The bill prescribes penalties for contraventions, indicating its potential to implicitly prohibit the creation and circulation of unauthorised deepfake videos once enacted into law. In essence, the bill establishes a robust framework for protecting personal data and privacy.

Indian Copyright Act, 1957

SECTION 14 - In the realm of deepfake content, complications arise when altered versions of sound and visual effects from copyrighted works, such as music videos or movies, are incorporated. Section 14 of the Copyright Act, 1957, grants exclusive rights to the owner of a cinematographed work, including any images or sounds within it.[13] Moral rights, as recognized in the case of *Amarnath Sehgal v. Union of India,*[14] empower authors to claim damages for any act of mutilation, distortion, or modification prejudicial to their honour. While copyright owners can seek civil remedies like injunctions and damages for infringements, the remedies may not be advantageous for deepfake victims. Typically, copyright is owned by movie producers rather than actors or photographers, making the actual targets of deepfake content less likely to benefit from the remedies provided by the act.

SECTION 52- In the Indian context, Section 52 of the Copyright Act, 1957, introduces the concept of 'fair dealing,' though not explicitly defined. Aligned with TRIPS Article 13, it aims to confine limitations or exceptions to exclusive rights in special cases that do not conflict with normal exploitation and do not unreasonably prejudice the legitimate interests of the right holder. However, the rigid nature of fair dealing in Indian Copyright law, criticised in comparison to the broad fair use doctrine in the United States, may pose challenges in addressing deep fakes. Even those created with legitimate purposes or for entertainment could be deemed copyright infringements. This calls for a reevaluation to accommodate deepfakes crafted for bona fide purposes within the legal framework.

[12] ibid.

[13] The Copyright Act, 1957.Section 14(d) and Section 14(e)

[14] Amarnath Sehgal v Union of India, 2005 (30) PTC 253 (Del)

International Framework

China:

In 2019, China enacted laws necessitating the disclosure of deepfake technology use and prohibiting its distribution without clear disclaimers. Provisions by the Cyberspace Administration of China (CAC) since January 2023 extend regulations to both deepfake providers and users. These regulations mandate consent, identity verification, government record registration, reporting of illegal content, recourse mechanisms, and the inclusion of watermark disclaimers. China's regulatory framework for deepfakes, introduced in 2019, compels individuals and organizations to disclose the use of deepfake technology. The regulations also mandate clear disclaimers on artificially generated content. Recent provisions by the Cyberspace Administration of China (CAC) require consent, identity verification, government record registration, reporting illegal deepfakes, offering recourse, and providing watermark disclaimers.[15]

Canada:

Canada adopts a comprehensive approach to deepfake regulation, focusing on prevention, detection, and response. Existing laws address the nonconsensual disclosure of intimate images. Initiatives, such as the "plan to safeguard Canada’s 2019 election" and the Critical Election Incident Public Protocol, demonstrate a proactive stance against potential deepfake threats. Canada adopts a three-pronged strategy involving prevention, detection, and response to combat deepfakes. Public awareness campaigns educate citizens about the risks, and the government invests in deepfake detection technology. Legislative efforts are underway to make malicious deepfake creation or distribution illegal.

EU:

The EU actively regulates deepfakes through various frameworks, including AI regulations, GDPR, the Copyright regime, and the Digital Services Act. The Code of Practice on Disinformation imposes significant fines for violators. The proposed EU AI Act aims to subject deepfake providers to transparency requirements. The EU takes a proactive stance on deepfake regulation, emphasising research into detection and prevention. Laws require social media companies to remove deepfakes,

[15] Amanda Lawson. (2023, April 24). A Look at Global Deepfake Regulation Approaches. Responsible ArtificialIntelligenceInstitute.https://www.responsible.ai/post/a-look-at-global-deepfake-regulation-approaches

with the Code of Practice on Disinformation imposing fines for violations. The Digital Services Act increases platform monitoring, while the proposed AI Act mandates transparency for deepfake providers.

South Korea:

South Korea responded to the deepfake challenge by passing a law in 2020, making the distribution of harmful deepfakes illegal. Penalties for offenders include imprisonment for up to five years or fines of significant amounts. Advocates propose additional measures, such as education and civil remedies, to address the multifaceted issue. South Korea, a pioneer in AI research, criminalised the distribution of deepfakes causing harm to public interest. Advocates call for additional measures like education and civil remedies. The government's commitment to technology, legal actions, and societal awareness reflects a comprehensive approach.

United Kingdom:

The UK government is actively addressing deepfake threats through funding research into detection technologies and plans for regulation in the Online Safety Bill. Initiatives like the ENOUGH campaign focus on raising awareness about the harms of deepfake porn. Anticipation surrounds the inclusion of deepfake regulation in future legislation. The UK government funds research on deepfake detection, collaborates with industry and academia, and runs public awareness campaigns. While horizontal legislation is pending in the Online Safety Bill, existing efforts demonstrate a commitment to addressing deepfake threats through research, partnerships, and public education.

United States:

While there are no federal regulations on deepfakes in the United States, certain states like California, Texas, New York, and Virginia have enacted laws, primarily targeting deepfake pornography. The DEEP FAKES Accountability Act, introduced in 2019 at the federal level, proposes disclosure requirements and penalties for those violating deepfake regulations. The focus is notably on addressing the potential misuse of deepfakes in specific contexts.In the US, state laws focus on deepfake pornography, while the DEEP FAKES Accountability Act proposes penalties and establishes a task force within the Department of Homeland Security. This aligns with broader concerns about national security and potential harm from deepfakes.

Proposed Legal Adjustments: Addressing Deepfake Challenges

Unmasking Deepfake Creators:

Within the promising framework of the Data Protection Bill, a notable limitation emerges when grappling with unidentified deepfake creators. A strategic proposal is to establish a specialised governmental entity leveraging blockchain technology. This approach ensures decentralised and tamper-proof verification of video authenticity, particularly in cases where creator identities remain elusive.

Blockchain's Role in Video Authentication:

Inspired by Axon Enterprise Inc.'s successful use of blockchain in authenticating police body camera videos, the recommendation advocates for extending this technological application. By integrating blockchain across various domains, the aim is to establish a robust and comprehensive mechanism for verifying the authenticity of videos.[16]

Government Assurance for Authenticity:

Recognizing the proliferation of videos in the public domain, the recommendation emphasizes the pivotal role of government and regulatory bodies in ensuring authenticity. An illustrative example proposes the mandatory use of Digital Signatures for political campaign videos, thereby introducing a secure mechanism for validating the legitimacy of circulated content.

Enhancing Intermediary Guidelines:

Critical analysis reveals limitations in the current Information Technology (Intermediary Guidelines) Rules 2011 in effectively addressing deepfake challenges. To bridge this gap, the recommendation underscores the necessity of governmental intervention to fortify the legal framework for identifying and regulating deepfakes.

[16] Antonio García Martínez, 'The Blockchain Solution to Our Deepfake Problems' (Wired, 26 March 2018) accessed 30 September 2020

Legislative Refinements: Adapting Laws to Deepfake Realities

Curbing Creation of Synthetic Persons:

Acknowledging the emergence of synthetic personas generated by programs, the recommendation suggests amending the Information Technology (Intermediary Guidelines) Rules 2011. This includes prohibiting the uploading of deepfake videos and incorporating clauses in privacy policies to enhance legal safeguards against the creation and dissemination of synthetic identities.

Protecting Data of Deceased Persons:

Highlighting a legal vacuum concerning the protection of data belonging to deceased individuals, the proposal recommends integrating provisions into the existing Data Protection Bill. This extension ensures comprehensive protection of personal data posthumously, with a specific emphasis on mandatory consent from heirs before any data is publicly circulated.

Empowering Heirs for Legal Action:

Advocating for empowering heirs or interested parties to initiate legal proceedings, the analysis draws inspiration from international practices, citing provisions in the Privacy Act of Hungary and the Spanish Data Protection Act. This inclusion ensures a robust legal framework addressing concerns related to the unauthorized use of data, especially in cases involving deceased individuals.

A comprehensive ban on deepfake AI-generating apps

A comprehensive ban on deepfake AI-generating apps and websites is proposed to address the multifaceted risks associated with this technology. This prohibition aims to preserve informed consent by eliminating the potential for coercion or misinformation when individuals agree to have deepfake content created. To empower individuals and ensure authenticity, the proposal suggests allowing them to actively participate in or oversee the content creation process when providing consent. Encouraging the exploration of non-AI alternatives, particularly for individuals capable of performing acts portrayed in deepfake videos, becomes essential. By mitigating the risk of unethical use, the ban serves as a proactive measure to protect individuals from unauthorized manipulation and potential harm, particularly in light of the ease of access to free deepfake websites. Upholding digital trust and integrity is emphasized, contributing to a safer online environment and preventing the erosion of confidence in digital media. The proposal also considers the economic incentives for malicious use, aiming to minimize the motivation for individuals or groups to exploit deepfake technology for unethical purposes. Aligning with ethical standards, the ban prioritizes the protection of individual rights, privacy, and the prevention of digital deception. Furthermore, it addresses the accessibility issue, aiming to limit casual and unregulated use by imposing a higher barrier to entry, thereby reducing the potential for misuse and harm. In conclusion, this detailed and comprehensive ban is advocated as a proactive and ethical strategy to mitigate the risks associated with deepfake technology and ensure responsible and transparent AI use in content creation.