Deepfake & Pornography: The Coming Crisis of Privacy and Consent

Dr. Rituparna Bhattacharjee[1*], Ms. Muskan Sharma[2]

[1*]*Assistant Professor (Sel.Gr), School of Law, The NorthCap University*
[2]*Research Scholar, School of Law, The NorthCap University*

KEYWORDS
Deepfake; Pornography; Right to Privacy; Artificial Intelligence; Gender Violene

ABSTRACT
The emergence of deepfake technology, powered by artificial intelligence (AI), presents a looming threat to privacy and consent, particularly in the context of pornography. With the ability to seamlessly superimpose one individual's likeness onto another's body, deepfake technology enables malicious actors worldwide to intrude upon anyone's privacy and manipulate an intimate imagery without consent. This paper examines the ethical and legal ramifications of deepfake pornography, emphasizing the urgent need for comprehensive legislation and policy measures to address this growing menace. Drawing upon examples from various jurisdictions, including India, it evaluates existing legal frameworks and highlights gaps in legislation. Additionally, the paper offers suggestions for policy interventions to combat deepfake pornography, safeguard individual privacy rights, and uphold principles of consent in the digital age. As deepfake technology continues to advance, proactive measures are essential to mitigate the potential harm it poses to society.

Introduction

Deepfake refers to the creation of synthetic pornographic content using deep learning technology to superimpose a person's face onto another individual's body in videos or images. The essence of the term "deepfake" lies in its components - 'Deep' represents Deep learning, and 'fake' signifies the ‘act of deception’[1]. It is kind of cybercrime, especially against women, that include manipulation of video and image of a women. Nowadays, the term “cybercrime”[2] has become a well-known and prevalent form of criminal activity, particularly targeting women. In recent time, we have witnessed a rise in cybercrimes against women, including deep fakes and activities on the dark web, potentially fuelled by technological advancements in India. Cybercrime is a dangerous attack on company or individual privacy.

Deepfakes are a video where the face of a person is replaced by another using facial mapping technology and artificial intelligence. It can be said to be an identity manipulation of a person using different tools, such as photoshop. The term "deepfake" was first brought to light in 2017 when a Reddit user used the technology to swap the faces of celebrities with those of adult stars[3]. The ‘deepfake’ videos have become commonplace, ranging from fabricated speeches by Mr. Barack Obama delivering false news to Mr. Mark Zuckerberg[4] asserting that Facebook exploits user information and violates privacy. Detecting any inconsistencies in these videos is nearly impossible, making it easy for viewers to be misled into perceiving them as genuine. Consequently, various nations, including the United States, are actively promoting research in technologies for detecting

[1] Heidari, A., Jafari Navimipour, N., Dag, H., & Unal, M, “*Deepfake detection using deep learning methods: A systematic and comprehensive review*”, 14(2) WIREs Data Mining and Knowledge Discovery (February 9, 2024).

[2] OXFORD UNIVERSITY, available at: https://www.oxfordlearnersdictionaries.com/definition/american_english/cybercrime (last visited March 12, 2024)

[3]The Hard Copy, available at https://thehardcopy.co/personalised-celebrity-messages-using-ai/ (last visited March 12, 2024)

[4] BUZZFEED, available at https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed (last visited March 20, 2024).

deepfakes and offering incentives, such as deepfake prizes[5]. Social media platforms like Facebook also support bug bounty programs aimed at discovering methods to identify deepfake content.
Broadly, deep fakes are machine learning bases software tool that produce realistic synthetic media content[6]. As per the Merriam Webster Dictionary, "deep fakes" defines as "an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said[7]." It is a type of cybercrime against women in India. It is not expressly defining in any Indian statue like Informational Technology Act of 2000. But impliedly section 66E[8] of the IT Act deals with the crime which infringe the privacy of any person. In addition of this, Section 66D of the IT Act, provides that "any individual with malicious intent, to cheat or impersonate someone by using any communicating device or computer resource, can result in imprisonment for up to 3 years or fine upto 1 lakh[9]".
These digitally tampered media are intended to mimic real life but actually threaten privacy and trust and are a very dangerous threat both for the individual and for societal problems. This calls for serious scrutiny of how the dangers brought by deepfakes can be tackled before these things cause untold damage that can never be undone. The emergence of deepfakes has raised alarm bells for the potential harm that they can cause to individuals, especially women in India. There is no provision in the IT Act 2002, which governs the use of technology in India, specifically addressing deepfakes. This leaves a gap in the legal framework that deals with the issue. Deepfakes may be used against women to harass, defame, or blackmail them.

The consequences of deepfakes for women in India are not limited to individual harm but extend to larger societal issues. For example, such manipulated media can create a culture of distrust about the authenticity of information and images, making it increasingly difficult to distinguish between what is real and what is fake[10]. In a world where women already experience so much against them, deepfakes can just increase the complexity in their lives. This would negatively affect women's mental well-being, their confidence levels, and overall well-being in many ways.

The rise in cases of cybercrime against women-the insidious uses of deep fakes and related activities on dark web sites-only indicates the transforming nature of dangers that technological change in the nation is facilitating. Deep fakes are machine-learning-based tools meant to generate media content that imitates real-time synthetic media activity, and a new sophisticated level of exploitation against individuals has thereby emerged[11]. Indian Prime Minister Narendra Modi highlighted concerns over negative use of AI at G20 virtual summit. "The world is worried about the negative effects of AI. India thinks that we have to work together on the global regulations for AI. Understanding how dangerous deepfake is for society and individuals, we need to work forward. We want AI should reach the people, it must be safe for society[12]."

Deepfake Threats to Gender Safety

The term deepfakes was popularized in 2018 by a Reddit user, who made a Reddit forum dedicated to creating and using deep learning software for synthetically face swapping female celebrities into

[5] WORLD ECONOMIC FORUM, available at https://www.weforum.org/agenda/2024/02/ai-deepfakes-legislation-trust/ (last visited March 20, 2024)
[6] Laffier, J., & Rehman, A., "*Deepfakes and Harm to Women*", (20)3 Journal of Digital Life and Learning, (2023).
[7] Merriam-Webster, available at https://www.merriam-webster.com/dictionary/deepfake (last visited March 20, 2024)
[8] Information Technology Act, 2002 (Act 21 of 2000), s.66E.
[9] Information Technology Act, 2000 (Act 21 of 2000), s. 66D
[10] Jack Coffin, *"Asking Questions of AI Advertising: A Maieutic Approach"* 51(5) Journal of Advertising (2022)
[11] Anna Pesetski, "*Deepfakes: A New Content Category for a Digital Age*," 29(2) William & Mary Bill of Rights Journal (2020).
[12] Hindustan Times, available at https://www.hindustantimes.com/india-news/ai-must-be-safe-for-society-pm-modi-on-deepfakes-at-g20-meet-101700656850382.html (last visited March 20, 2024)

pornographic videos. The Amsterdam-based cybersecurity firm, Deeptrace, reported that deepfake pornographic videos are aimed and targeted mainly at women than men, hence increasing gender inequality. Women are the victims in 90% of revenge porn, non-consensual porn, and other harassment where deepfake is one more to that list[13].
Continuing this trend, using artificial intelligence will inevitably include making deepfakes-the use of deepfake technology to create digital fabrications. Deepfakes are synthetic media in which a person in an existing image or video is replaced with someone else's likeness, accomplished using AI techniques. The problem is, deepfakes are very real; it may be hard for anyone to say the difference between real versus manipulated media. And it has the potential and real capacity to promote impersonation, fraud, blackmail, and propaganda as well as misinformation.

An example of this would be the latest release of explicit sexual images created by AI on Taylor Swift that garnered over 45 million views, 24,000 reposts, and hundreds of thousands of likes and bookmarks for almost 17 hours before being taken down. It was shut down eventually, but the images were already out in the wild, reposted in accounts and on social media sites, and this caused huge damage.
At the core of deepfakes is AI, specifically a subset of machine learning called deep learning. Deep learning uses neural networks with many layers to analyze and interpret large data sets. Deepfakes are hyper-realistic digital fabrications, typically video or audio recordings, created using AI and machine learning techniques. They involve superimposing existing images and videos onto the source images or videos using a deep learning technique, particularly generative adversarial networks ("GANs").

GANs consists of two types of neural networks: generator and discriminator. Generator creates images and videos while discriminator evaluates those created images/videos as compared to real footage. The technology creates a convincingly realistic fake content of people saying or doing things that they never really did.
The data that creates deepfakes comes in the form of pictures, soundbites, or even videos of the target person. As more data may produce a more convincing deepfake, the most common targets are celebrities and politicians. The emergence of apps and online services where users can create deepfakes with minimal technical knowledge has further lowered the bar to entry.

AI trains by using collected data to understand the subtleties and nuances of the target person, such as facial features, voice tone and inflections, and movements. The trained model then superimposes the target person's likeness onto a source photo or video. Video manipulations are done on a frame-by-frame basis to create a realistic video.

On November 07, 2023, a day after a lewd video of actor Rashmika Mandanna surfaced on several social media platforms, she came out decrying about its authenticity. The actor's face unwittingly had been superimposed on the body of a British Indian influencer.
Similarly, the resurrection of Paul Walker was created for Fast & Furious 7. In 2020 Indian legislative assembly elections politician Manoj Tiwari's speech delivered in English was manipulated to be disseminated in the 'Haryanvi' dialect.
Initially, in 2017, it was the world's most renowned women who became the first to be affected by this technology. Cases like Scarlett Johansson's manipulations, Gal Gadot's, and Jennifer Lawrence's raised the concerns of the same videos becoming a new means of defamation and humiliation of women.

[13]The State of Deepfakes: Landscape, Threats, and Impact, available at: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf (last visited March 20, 2024)

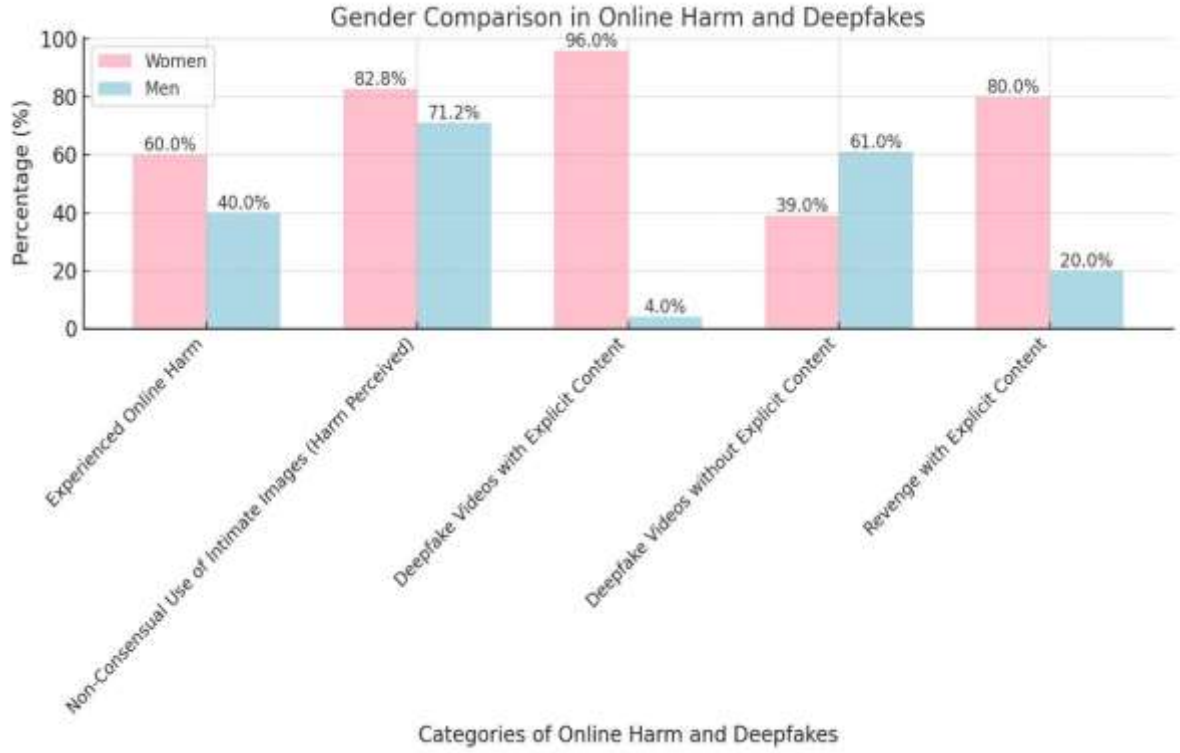
Gender Based Prejudice

The ever-increasing digital era has resulted in increased productivity and efficiency, improved communication, and innovation. But on the other hand, it has given rise to the new and sophisticated form of reproducing other types of violence. Historically, this violence has not affected all people in the same way.

The Centre for International Governance Innovation (CIGI) conducted an international survey on online harm in 2020 in 18 countries, involving 18,149 participants, with a finding of 60% of respondents claiming to have experienced at least one of 13 forms of online harm, with the highest proportion of incidents being reported by LGBTQ+ individuals. In addition, women were a slightly higher proportion than men, though they were more likely to report a more severe impact of these incidents compared to men. Among the online harms surveyed, the non-consensual use of intimate images was the incident most considered harmful, especially for women (82.8% vs. 71.2% for men).[14]

As for deepfakes, research has indicated that women are the main victims. A 2019 report by the cybersecurity firm Deep trace, which conducted research on the status of deepfakes, discovered that 96% of online videos were intimate or sexual in nature. Women, who often are actresses or musicians, are the most victimized; in lesser cases, it is the media professionals. On the other hand, explicit content videos are mainly aimed at men, where 61% of the targeted individuals were politicians and corporate officials[15].

The most terrifying aspect of deepfakes relates to their ability to connect revenge through the release of intimate and sexually explicit content. This takes place when a person's ex-partner or third parties disseminate or threaten to disseminate sexually explicit intimate images of the person in order to control, punish, and/or damage his or her reputation without his or her consent. Some studies have recorded gender patterns in this phenomenon, indicating that it is more prevalent among women than men, and therefore, it is another form of gender-based violence[16].



[14] Centre for International Governance Innovation, available at https://www.cigionline.org/publications/supporting-safer-digital-spaces/. (last visited March 20, 2024)

[15] Supra note 13.

[16] SWGfl, available at: https://swgfl.org.uk/magazine/revenge-porn-research-2019/ (last visited March 20, 2024)

Deepfake victims often deal with guilt, anger, shame, and stigma in along with general anxiety. Their privacy being violated and distorted photographs being circulated can have a severe negative impact on their mental health, which in turn can lower their quality of life and even result in trauma. Although deepfake pornography primarily targets women, men can occasionally fall prey to it in order to extract money. People's lives and reputations can be harmed by pornographic deepfakes. In addition to causing emotional distress, financial loss, and job loss, it treats women like sex objects[17].
On the BBC's The She Word[18], two Zimbabwean women talked about their experiences as revenge porn victims. One of them lost her job, while the other was disowned, which prevented her from finishing her studies. This kind of horrible act raises severe concerns about privacy violations and the obscene portrayal of women. It is typical for victims of revenge porn to experience substance abuse, PTSD, or feelings of hopelessness or worry.

Deepfakes have the potential to cause enormous harm since they are so easily confused for authentic video footage. This is where further research is needed to fully understand the link to image-based sexual assault, also referred to as revenge porn. The UK's 2019 Adult Online Hate, Harassment and Abuse report defines "sexual photoshopping" as one of six forms of image-based sexual assault. The research emphasises that the harm experienced is equivalent to more "traditional" kinds of image-based sexual violence, despite not using the term "deepfake." In light of this, it is important to highlight that the majority of revenge porn victims are women, according to the majority of research (the article references various studies, putting the percentages between 60 percent and 95 percent)[19]. Interestingly, a study emphasized that male victims of image-based sexual abuse report feeling less guilt and less self-blame than female victims in the same circumstance.[20]
Sextortion, also "eWhoring"- Sextortion is a relatively new combination of the words like "sex" and "extortion." In general, extortion happens when "one individual takes advantage of another against his or her will by threatening him or her with violence or injury of any kind[21]". The harm can be physical (to them or their loved ones) or can target their property or reputation[22] typically involving blackmail the threat of revealing damaging secret information or ransom, where something of value is held until the victim fulfils a specified condition[23]. According to the 2023 State of Deepfakes report[24], deepfake videos have increased more than five times, or 550 per cent, since 2019. About 98 per cent of total deepfake videos are porn, and 99 per cent of them target women – a pattern clearly evident in several incidents this year.
The integration of Artificial Intelligence (AI) into the pornography industry marks a significant evolution in how content is created, distributed, and consumed. This technological leap forward is not only changing the landscape of adult entertainment but also raising questions about the future of

[17] Ashish Jaiman, "*The danger of deepfakes*", The Hindu, (January 1, 2023), available at https://www.thehindu.com/sci-tech/technology/the-danger-of-deepfakes/article66327991.ece (last visited March 12, 2024)
[18] Dr. Zubair Ahmed Khan & Ms. Asma Rizvi, "*Deepfakes: A Challenge for Women Security and Privacy*" 5(1) CMR UNIVERSITY JOURNAL FOR CONTEMPORARY LEGAL AFFAIRS (2023)
[19] Adult Online Hate, Harassment and Abuse: A Rapid Evidence Assessment, available at: https://assets.publishing.service.gov.uk/media/5d110e2640f0b6200bb2b2da/Adult_Online_Harms_Report_2019.pdf (last visited March 12, 2024)
[20] Supra Note at 18
[21] Ibid
[22] Ibid
[23] James Lindgren "*The theory, history, and practice of the bribery-extortion distinction*" 141 University of Pennsylvania Law Review (1993) 1695-1740 available at: https://scholarship.law.upenn.edu /cgi/viewcontent.cgi?article=3659&context=penn_law_review
[24] Subham Tiwari, "*Inside Thriving Deepfake Porn Bazaar*" India Today, (December 4, 2023). available at https://www.indiatoday.in/india/story/deepfake-porn-artificial-intelligence-women-fake-photos-2471855-2023-12-04 (last visited June 15, 2024)

human sexuality in a digital age. Deepfake technology, a portmanteau of "deep learning" and "fake," refers to the use of AI algorithms to create or alter video content so that it presents something that didn't actually occur. In the context of pornography, this often involves superimposing the faces of celebrities or other individuals onto the bodies of porn actors. [25]While the technology has legitimate applications in areas such as filmmaking and content creation, its misuse in creating non-consensual pornographic content has raised significant ethical and legal concerns. [26]

Interactive Experiences AI is also being used to create interactive pornographic experiences that respond to the user's actions and preferences in real-time. This can include virtual reality (VR) environments where users can engage with AI-driven characters or scenarios that adapt based on user input.By analyzing user behavior and preferences, AI algorithms can curate and suggest content that is likely to be of interest to the viewer. This level of personalization enhances user engagement and satisfaction, but it also raises questions about privacy and the potential reinforcement of certain preferences or behaviours. [27]

An emerging application of AI in the industry is the generation of erotic literature. Platforms utilizing natural language processing algorithms can create personalized stories based on user preferences and inputs. This innovative use of AI offers a customizable experience, allowing users to explore fantasies in a safe and private manner. The development of AI-driven virtual companions represents a significant leap in interactive adult entertainment. These AI entities can engage users in conversations, learn from interactions, and even exhibit personalities tailored to user preferences.

AI technologies have opened new frontiers in the production of pornographic content, making it possible to create more with less. Using AI-generated imagery and deepfake technology, producers can craft scenes or entire narratives without the need for physical filming locations, traditional actors, or extensive production crews. Moreover, AI-driven content creation democratizes the production process, allowing independent creators and smaller studios to compete with larger entities. By lowering the barriers to entry, AI fosters a more diverse and vibrant ecosystem of content, catering to a wider array of tastes and preferences. Tailoring Experiences to Individual Desires Perhaps one of the most significant impacts of AI on the pornography industry is the unprecedented level of personalization it offers. AI algorithms analyze user data, including viewing patterns and interactions, to curate content that aligns with individual preferences.

The ability of AI to cater to specific fantasies or create hyper-realistic and idealized scenarios can lead to unrealistic expectations about sexual encounters, body image, and performance. This discrepancy between fantasy and reality may contribute to dissatisfaction in personal relationships and an altered perception of healthy sexuality. Impact on Mental Health and Well-Being Moreover, the exposure to non-consensual or extreme content, facilitated by AI's ability to bypass ethical considerations, can have harmful psychological effects on consumers. Influence on Societal Perceptions of Sexuality AI-driven pornography has the power to shape societal perceptions of sexuality, potentially reinforcing harmful stereotypes and norms.

The US has always been ahead of the curve in taking a stance against the production, dissemination, and viewing of even fake child sexual abuse material if the obscene images depict someone who is "virtually indistinguishable" from a real child. The law makes it clear that "*it is not required …that the minor depicted actually exist(s).*" Federal law enforcement has also taken an aggressive approach to AI-generated pornography and "*revenge porn.*" [28]The question then arises as to the relevance of these state laws in the contemporary digital landscape. Legislative Framework

[25] Chen Yu, "*Beyond Flesh and Fantasy: The Rise of AI in Reshaping the Pornography Landscape*" available at https://vixra.org/pdf/2403.0018v1.pdf (last visited June 15, 2024)

[26] Ibid.

[27] Ibid.

[28] Cornell Law School, "*18 U.S. Code § 1466A - Obscene Visual Representations of the Sexual Abuse of Children*, available at https://www.law.cornell.edu/uscode/text/18/1466A (last visited April 14, 2024)

International perspective

It remains in its growing phase as multiple countries and conventions are working in this regard towards the resolution of the issue. Though, it does not hold an express international legal framework for cybercrime as such; it is however controlled by a few international conventions where cybercrime was defined as a form of violence against women through digital or an online platform: Istanbul Convention and Budapest Convention.

Therefore, as per the recommendation of the "Committee on the Elimination of All Forms of Discrimination against Women (CEDAW)" extends the definition of 'violence against women' beyond the physical space to include "technology-mediated environments," thereby addressing online and ICT-facilitated digital violence against women[29]". Additionally, "the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence (Istanbul Convention)" was the first legally binding instrument in Europe that offers a comprehensive framework to end digital violence against women and domestic violence[30].

Moreover, the Istanbul Convention provides a comprehensive definition of the types of violence against women, including online and ICT-facilitated violence. It considers violence against women as a violation of human rights and a form of discrimination against women. As per the Article 3(a) and (b) of this convention, defined "violence against women" as "any acts of gender-based violence that result in, or are likely to result in, physical, sexual, psychological, or economic harm or suffering to women, including threats of such acts, coercion, or arbitrary deprivation of liberty, whether occurring in public or in private life"[31].

Additionally, Article 69 of the Istanbul Convention gives GREVIO the authority to propose general recommendations for the convention's implementation. In light of this, GREVIO agreed to draft its first general recommendation during its 21st plenary meeting, focussing on the Istanbul Convention's applicability to the digital component of violence against women[32]. Because of this, the definition of "violence against women" has expanded to include non-consensual sharing of images or videos, coercion, and threats, such as threats of rape, sexualised bullying, and other forms of intimidation; online sexual harassment; impersonation; online stalking or stalking via the Internet of things; and psychological abuse and economic harm committed against women and girls through digital means[33]. Article 1 Para 2 of the same convention establishes a specific monitoring mechanism, the GREVIO evaluates the compliance of implementing measures with the provisions of the convention[34]. Although, the Istanbul Convention does not contain an explicit reference to the digital dimension of violence against women, but drafter intended to cover violence committed in the digital sphere under Article 2 of the Convention[35].

---

[29] Reyhanne, Nelsy, Marikel Affoum, Isabel Micaela, Santagostino Recavarren, Nayantara Vohra, & Quentin Wodon. "*Protecting Women and Girls from Cyber Harassment: A Global Assessment of Existing Laws*" Global Indicators Briefs No. 18, (June 22, 2023). available at https://documents1.worldbank.org/curated/en/099456506262310384/pdf/IDU0c7c3a5a70b56a04b250a31b0b32b8f5cd856.pdf (last visited April 14, 2024)

[30] Adriane van der Wilk, "*Protecting women and girls from violence in the Digital age*" (December 2021), available at https://rm.coe.int/therelevance-of-the-ic-and-the-budapest-convention-on-cybercrime-in-a/1680a5eba3 (last visited April 14, 2024)

[31] Library of Congress, "*European Union: Istanbul Convention Enters into Force*" available at https://www.loc.gov/item/global-legal-monitor/2023-10-10/european-union-istanbul-convention-enters-into-force/ (last visited April 14, 2024)

[32] Council of Europe, "*Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO)*", available at https://rm.coe.int/grevio-s-baseline-evaluation-report-on-legislative-and-other-measures-/1680ad79b9 (last visited April 14, 2024)

[33] Ibid

[34] Ibid.

[35] Supra Note 30.

In fact, a number of the Istanbul Convention apply in digital space, such as, Article 40[36] is applicable to online and technology-facilitated sexual harassment as per its definition, "*any form of unwanted verbal, non-verbal or physical conduct of a sexual nature with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment*[37]".
Furthermore, the Council of Europe Convention on Cybercrime (the Budapest Convention) is the first and most significant international legally enforceable convention addressing cybercrime and electronic evidence. In November 2001, the Council of Europe's Committee of Ministers accepted the convention and its explanatory report. It was opened for signatures in Budapest and went into effect on July 1, 2004. As of June 2021, 66 states have joined the convention.[38]. India was not a part of this Convention as sharing data with foreign law enforcement agencies violates India's national sovereignty[39].

The convention's main goals are to harmonise the domestic criminal substantive law components of offences and related provisions in the context of cybercrime; to establish a quick and efficient system of international cooperation[40]; and to provide domestic criminal procedure law powers required for the investigation and prosecution of such offences as well as other offences committed using a computer system or evidence related to which is in electronic form. Racist and xenophobic offences perpetrated using computer systems are made illegal under the first Additional Protocol to the Convention on Cybercrime. It became operative on March 1, 2006, after being approved by the Council of Europe Committee of Ministers in November 2002.
As of June 2021, 33 states have signed up to the additional procedure. The convention understands that computer systems facilitate not only communication and freedom of expression, but also the transmission of racist and xenophobic material and speech, and it requires parties to criminalise this dissemination. Subsequently, in September 2017, the Second Additional Protocol to the Budapest Convention was implemented in order to address the issues related to criminal justice in cyberspace and facilitate enhanced collaboration on cybercrime and electronic evidence[41].

The Cybercrime Convention Committee (T-CY)[42] represents the convention's signatory governments and makes sure the Budapest Convention is implemented effectively. Article 46 of the Convention states that the Committee's consultation is intended to facilitate information sharing, the efficient application and implementation of the Convention, and the evaluation of any potential modifications. According to the report "Toxic Twitter" issued by Amnesty International, "online threats of violence against women are often sexualized and include specific references to women's bodies. The aim of violence and abuse is to create a hostile online environment for women with the goal of shaming,

[36] Article 40 of Istanbul Convention, define "Sexual Harassment" as "Parties shall take the necessary legislative or other measures to ensure that any form of unwanted verbal, non-verbal or physical conduct of a sexual nature with the purpose or effect of violating the dignity of a person, in particular when creating an intimidating, hostile, degrading, humiliating or offensive environment, is subject to criminal or other legal sanction."
[37] Meyersfeld, B. C. "*Introductory Note to the Council of Europe Convention on Preventing and Combating Violence against Women and Domestic Violence*" 51(1) International Legal Materials, (2012) 106. available at https://projects.iq.harvard.edu/violenceagainstwomen/publications/introductory-note-council-europe-convention-preventing-and-0 (last visited April 14, 2024)
[38] Supra Note 35
[39] Ibid.
[40] Supra Note 30
[41] Council of Europe, Cybercrime Convention Committee (T-CY). "*Preparation of the 2nd Additional Protocol to the Budapest Convention on Cybercrime, State of play.*" available at: https://rm.coe.int/t-cy-2020-32-protocol-tor-chair-state-ofplay/1680a06a83. (last visited April 14, 2024)
[42] Ibid.

intimidating, degrading, belittling, or silencing women[43]" According to the same research, 25% of respondents—all of whom were active on Twitter—received threats against them and their families, including death threats, bodily harm, sexual violence, and incitement to suicide. Despite the victim's perceived identity, these threats frequently overlap with other hate speech[44].
The first-ever Bletchley Declaration was recently chaired in November 2023 with the goal of addressing the risks and responsibilities associated with AI in a more thorough and cooperative manner by focussing more on promoting scientific cooperation[45]**.** The risks of deliberate abuse and loss of control over AI technologies were discussed in the Bletchley Park Declaration of the summit. This declaration deals with the substantial risks arising from intentional and unintentional misuse over AI. Furthermore, it also acknowledges the much broader risks related to AI, like issues of bias and privacy as well. This declaration has signed by the 28 countries including India and European Union to address the potential risks posed by artificial intelligence.[46]. The declaration outlines several key priorities for the effort. The participating countries will prioritize "identifying AI safety risks of shared concern, building a shared scientific and evidence-based understanding of these risks, and sustaining that understanding." They will also seek to establish "respective risk-based policies across our countries" to address AI-related issues.

United States of America

Cybercrime against women is not specifically covered by any laws in the United States of America. Yet, a number of laws and rules address different facets of cybercrime, including those that can impact women. A federal legislation known as the Computer Fraud and Abuse Act (CFAA) of 1986 makes it illegal to gain unauthorised access to computer networks and systems, steal computer data, or disclose it without authorisation[47]. The CFAA can be used to punish incidents involving unauthorised access to personal information or internet accounts, even if it does not particularly target deepfake as a felony against women. In accordance with growing worries about computer hacking, the regulation was passed as an amendment to the Comprehensive Crime Control Act of 1984. Blackmail, threats, cyber pornography, publishing pornographic material, stalking, bullying, defamation, morphing, and creating false profiles are among the cybercrimes committed against women in the United States[48].
Over the past few years, deepfakes—manipulated media files that show a phoney image or video of a person—have become more and more common. According to earlier estimates by Wired, the top 35 websites hosting deepfake pornography had at least 244,635 deepfake films uploaded in the first nine months of 2023. Although there is currently no federal law in the United States, ten states, including Virginia and Texas, have criminal prohibitions prohibiting deepfakes. The Preventing Deepfakes of Intimate Images Act of 2023 was introduced in May 2023 by New York Democrat Rep.

[43] Amnesty International, "*Toxic Twitter – a toxic place for women*", (March 21, 2018), available at: https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1-1/ (last visited April 14, 2024)
[44] Ibid.
[45] Milin Stanly, "*A peep into the future of AI: The Bletchley Declaration and International Collaboration for AI Safety*" Indiaai, (November 9, 2023), available at https://indiaai.gov.in/article/a-peep-into-the-future-of-ai-the-bletchley-declaration-and-international-collaboration-for-ai-safety (last visited April 14, 2024)
[46] Kevin Poireault, "*28 countries sign Bletchley Declaration on AI safety" Infosecurity Magazine*, (November 1, 2023). available at https://www.infosecurity-magazine.com/news/28-countries-bletchley-declaration/#:~:text=The%20Bletchley%20Declaration%2C%20an%20international,Nigeria%2C%20Israel%20and%20Saudi%20Arabia. (last visited April 14, 2024)
[47] Investopedia, "*Computer Fraud and Abuse Act (CFAA)*" (September 9, 2022) available at https://www.investopedia.com/terms/c/computer-abuse.asp#:~:text=1984%20(CFAA).-,The%20Computer%20Fraud%20and%20Abuse%20Act%20of%201984,both%20civil%20and%20criminal%20matters. (last visited March 12, 2024)
[48] GeeksforGeeks, "*Cyber Crime Against Women*" available at https://www.geeksforgeeks.org/cyber-crime-against-women/ (last visited on March 12, 2024)

Joe Morelle in an effort to make it illegal to share sexual deepfake photos online without consent[49]. This Act was passed in order to give victims of damaging deepfakes legal remedy and to safeguard national security from the dangers posed by deepfake technology[50].
The bipartisan Deepfake Task Force Act was introduced by the United States to help the Department of Homeland Security combat deepfake technology. A new measure known as the Disrupt Explicit Forged Image and Non-Consensual Edits (DEFIANCE) Act of 2024 was introduced by US senators on January 30. It lets victims of deepfakes and porn produced by AI file lawsuits to get compensation[51]. In addition, on January 10, US congressmen Madeleine Dean and María Elvira Salazar introduced the No Artificial Intelligence Fake Replicas and Unauthorised Duplications (No AI Fraud) Act of 2024, which would shield Americans from having their voices and images altered.

United Kingdom

The primary law in the UK pertaining to offences or attacks against computer systems, like hacking or denial of service, is the Computer Misuse Act 1990 (also known as "CMA1990")[52]. The majority of crimes committed against women and girls in the technological age include coercive and controlling behaviour crimes, cyberstalking and harassment, and the disclosure of private sexual images without consent. Online activity is used to plan and orchestrate violent acts as well as to humiliate, control, and threaten them.
According to the Sexual Offences Act of 2003, criminals may use social media or online dating services to facilitate their crime by setting up a meeting with a victim with the goal of committing rape or other sexual offences, much like in online romance fraud. At a roundtable on the enforcement and prosecution of "violence and harassment" online in 2013, the End Violence Against Women Coalition (EVAW) collected testimonies and expressed concerns that criminal justice authorities approached online violence and harassment differently and less successfully than they did offline[53].
The National Centre for Cyberstalking Research (NCCR) was established in the United Kingdom in 2009 with the goal of conducting research and analysis on the incidence, causes, effects, and risk assessment of cyberviolence against women and girls. The centre is now undertaking a survey to look into the incidence and impact of revenge porn. In 2011, the centre published the findings of a study on the prevalence, nature, and impact of cyberstalking[54].
The fundamental law governing the processing of personal data in the UK is the Data Protection Act of 2018 (DPA 2018), which is enforced in conjunction with the UK General Data Protection Regulation[55]. All facets of how businesses, organisations, and governmental bodies handle and handle personal data are governed by this framework for data protection. All UK data controllers—

[49] "Text of H.R.3106 - 118th Congress (2023-2024): Preventing Deepfakes of Intimate Images Act" available at https://www.congress.gov/help/legislative-glossary#glossary_billsummary (last visited on March 12, 2024)

[50] Janvhi Rastogi, "*Deepfake Pornography: A Legal and Ethical Menace*" (October 16, 2023) available at https://tclf.in/2023/10/16/deepfake-pornography-a-legal-and-ethical-menace/#:~:text=India%2D%20There%20is%20no%20distinctive,create%20liability%20against%20the%20perpetrators. (last visited on March 12, 2024)

[51] Legal Guidance, "*Cybercrime Prosecution Guidance*" available at https://www.cps.gov.uk/legal-guidance/cybercrime-prosecution-guidance (last visited on March 12, 2024)

[52] Ibid

[53] End Violence Against Women "*New Technology: Same Old Problems. Report of a roundtable on social media and violence against women and girls*" (2013). available at https://www.endviolenceagainstwomen.org.uk/wp-content/uploads/Report_New_Technology_Same_Old_Problems.pdf (last visited on March 12, 2024)

[54] Maple, C., Shart, E., & Brown, A. "*Cyber stalking in the United Kingdom: An Analysis of the ECHO Pilot Survey*" University of Bedfordshire. (2011) available at: https://www.beds.ac.uk/__data/assets/pdf_file/0003/83109/ECHO_Pilot_Final.pdf (last visited on March 12, 2024)

[55] UpGuard, "*Cybersecurity Laws and Regulations in the UK*" available at https://www.upguard.com/blog/cybersecurity-laws-regulations-uk (last visited on March 12, 2024)

businesses and organisations that oversee the processing of personal data—are required by the DPA 2018 to set up and keep up suitable security measures to safeguard personal data.
The UK parliament passed the Online Safety Act of 2023 to control internet and online safety and to defend against online crime. One of the British laws that includes provisions to punish "downblousing" and deepfakes of pornography is this one[56]. The goal of the Act was to make the UK "the safest place to be online." According to the Bill, offenders will be subject to harsh punishment regardless of whether the originator meant to embarrass or upset the victim.

A variety of new "communication offences" are introduced under Part 10 of the Act, which goes into effect on January 31, 2024, for individuals who send harmful communications via messaging services, dating apps, social media platforms, or "airdrops." The new communication offences will apply to users variously across the home nations. Part 10 of the Act will have retrospective effect[57]. These violations include messages sent through dating apps, social media, and device-to-device sharing. Offenders risk a fine in addition to a maximum two-year jail sentence. The Act's crimes target the rising prevalence of digital and online abuse and seek to make the internet a safer place. Prosecutors have so far tended to resist these activities by citing statutory or common law charges, such harassment, that are unsuitable for certain online conduct.

China
In 2022, China has taken significant steps to address issues related to deepfakes and digital impersonation. The Cyberspace Administration of China (CAC), the Ministry of Industry and Information Technology (MIIT) and the Ministry of Public Security (MPS) jointly issued the Provision on the Administration of Deep Synthesis of Internet (Deep Synthesis Provisions) based service [58]. The Deep Synthesis Provisions will become effective on January 10, 2023 as part of the Chinese government's efforts to strengthen the supervision of deep synthesis technology and services.

The provisions provide a comprehensive definition of 'deep synthesis technology' and include techniques like face generation, face replacement, gesture manipulation. Under the regulations, deepfakes cannot be produced without the users' consent and those which are created should be clearly labelled as 'edited.' The Deep Synthesis Provisions clearly stipulate the main responsibility of "deep synthesis service providers[59]"
The Deep Synthesis Provisions is an elaboration of "Regulations on the Administration of Online Audio and Video Information Services 2019," which broadly banned the use of machine-generated images, audio and video to create or spread "rumours[60]. The new regulations are aimed at deep

[56] Hern, A. "*Online safety bill will criminalise 'downblousing' and 'deepfake' porn*" The Guardian (2022, November 24). Available at: .https://www.theguardian.com/technology/2022/nov/24/online-safety-bill-to-return-to-parliament-nextmonth#:~:text=Online%20safety%20bill%20will%20criminalise,porn%20%7C%20Internet%20safety%20%7C%20The%20Guardian. (last visited on March 12, 2024)
[57] Legislation.Gov.UK. "*Online Safety Act 2023*" available at https://www.legislation.gov.uk/ukpga/2023/50/enacted (last visited on March 12, 2024)
[58] The Diplomat. "*China's New Legislation on Deepfakes: Should the Rest of Asia Follow Suit?"* The Diplomat, (2023, March). available at https://thediplomat.com/2023/03/chinas-new-legislation-on-deepfakes-should-the-rest-of-asia-follow-suit/ (last visited on March 12, 2024)
[59] Deep synthesis service providers" refers to companies that offer deep synthesis services as well as those who provide them with technical support.
[60] China Briefing, "*China to Regulate Deep Synthesis (Deepfake) Technology from 2023*", available at: https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/ (last visited Apr 14, 2024).

synthesis service providers and emphasize cybersecurity, real-name verification of users, data management, marking of synthetic content to alert viewers and "dispelling rumours[61]".
They expand the Chinese government's efforts to prevent social and political disruption by increasing its control of the Internet. These efforts are tied to the actions of tech platforms and companies. According to the Deep Synthesis Provisions, deep synthesis service providers must strengthen data management by taking necessary measures for personal data protection according to the existing Data Security Law, Personal Information Protection Law, etc.[62].
The Deep Synthesis Provisions call for the creation of a mechanism for dispelling fake news so that when deep synthesis services are used to produce, copy, publish and disseminate false information, deep synthesis service providers are required to take measures to dispel such news, keep records, and report them to the relevant authorities (such as the Internet Information Department)[63]. In addition, the new measures make it mandatory to add labels or tags on information generated from using deep synthesis technologies. These include voice simulation, intelligent conversation or writing that simulate the style of a real person, face image synthesis, or face manipulation.

India

In India, offences related to deepfakes, especially those involving computers, are addressed in large part under the Information Technology Act of 2000. Instances where a computer resource or communication device is utilised maliciously for personation and cheating are particularly covered by Section 66D of the Act. This clause covers situations in which people are technologically coerced into saying or doing things that result in fraud[64]. The Information Technology Act of 2000's Section 66E also addresses the use of deepfakes for privacy breach. This section describes the invasion of privacy that happens when deepfakes are used to obtain, share, or publish someone's private images or videos without that person's knowledge or consent[65].
Sections 67A and 66B of the Information Technology Act of 2000 apply to deepfakes that contain explicit or pornographic content. These sections outline the fines and penalties associated with publishing and disseminating explicit and sexual content involving children and adults. Notably, in order to adhere to these legal requirements, a number of companies, including Pornhub, have banned deepfake sexual content. Under Section 79 of the Information Technology Act of 2000, intermediaries—the websites where deepfake content is frequently posted—are subject to liability regulations. This clause mandates that intermediaries remove content either when they become aware of its existence or when they receive a court order. The court emphasised in the *Myspace Inc. v. Super Cassettes Industries Ltd.*[66] decision that, even in the absence of a court order, intermediaries are required to take down infringing content as soon as private parties notify them of copyright infringement.
With the introduction of the Information Technology Rules in 2021, Social Media Intermediaries (SSMIs) were subject to new regulations. In order to be classified as SSMIs, intermediaries with more than a specific number of registered users must assign staff members to monitor and identify the source of information and specific types of content. The rules guarantee a more dependable structure for addressing content-related issues and contain a grievance settlement process for intermediaries to address user complaints and grievances.

Deepfakes of pornography have emerged as a result of the exponential expansion of machine learning and artificial intelligence technologies. Deepfake technology has significant roots in India, especially

[61] China Briefing, "*Cyberspace Administration of China Announcement on Regulating Deep Synthesis Technology*" available at: https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/ (last visited on March 12, 2024).
[62] Supra Note 60
[63] Ibid.
[64] Information Technology Act, 2000 (Act 21 of 2000), s. 66D.
[65] Information Technology Act, 2000 (Act 21 of 2000), s. 66E
[66] 2016: DHC:8178-DB

the extensive use of the technique in pornography, politics, the film business, and cases of revenge-defamation. According to the Puttuswamy Case[67], it violates the basic right to "privacy" guaranteed by article 21 of the Indian Constitution. The non-consensual use of people's photos for deepfakes is another violation of the right to privacy guaranteed by Article 21.

Digital Personal Data Protection Act 2023 and Deepfake[68]

Although, prior to the adoption of the DPDP Act of 2023, there was no comprehensive data protection law in India. As a result, the Information Technology Act (IT Act) of 2000, which grants legal recognition for transactions conducted through electronic data interchange and other electronic communication channels—often referred to as "electronic commerce"—governed data protection[69].

The DPDP Act also defines the term "data" as "means a representation of information, knowledge, facts, concepts or instructions which are being prepared or have been prepared in a formalised manner, and is intended to be processed, is being processed or has been processed in a computer system or computer network, and may be in any form (including computer printouts magnetic or optical storage media, punched cards, punched tapes) or stored internally in the memory of the computer[70]"

With the spurt in the technological development, these amendments in IT Act become ineffective for the protection of the personal data. Therefore, parliament enacted Digital Personal Data Protection Act of 2023 (DPDP Act), a separate or comprehensive Act, "to protect digital personal data in a manner that recognises both the right of individuals to protect their personal data and the need to process such personal data for lawful purposes and for matters connected therewith or incidental thereto[71]".

The DPDP Act, similar to GDPR, applies to digital personal data but is narrower and does not include entities outside India monitoring Data Principals' behaviour[72]. Further, the DPDP Act imposes narrowly defined obligations for processing digital personal data, establishes purpose limitation obligations, and creates rights for individuals whose data are collected and used[73]. It also establishes a supervisory authority, known as, the Data Protection Board, that investigates complaints and issues fines but it lacks the power to issue guidance or regulations[74].

The DPDP law is a transformative law that extends the scope of personal data processing to all entities, regardless of size or private status. It borrowed from the European Union's General Data Protection Regulation (GDPR) approach.[75] Further, it has significant extraterritorial application if the procession of digital personal data is in the connection with any activity related to offering of goods or services to Data Principle within the territory of India[76].

---

[67] Justice K.S. Puttaswamy (Retd.) & Anr. vs. Union of India & Ors (2017) 10 SCC 1

[68] Sarvagya Chitranshu, "*The Deepfake Conundrum: Can the Digital Personal Data Protection Act 2023 Deal with Misuse of generative AI?*" Indian Journal of Law & Technology, (December 23, 2023). available at https://www.ijlt.in/post/the-deepfake-conundrum-can-the-digital-personal-data-protection-act-2023-deal-with-misuse-of-ge (last visited on March 12, 2024).

[69] Digital Personal Data Protection Act of 2023 (Act no. 22 of 2000)

[70] Information Technology Act, 2000 (Act 21 of 2000), s. 2(1)(o)

[71] Digital Personal Data Protection Act of 2023 (Act no. 22 of 2000)

[72] Robyn Annetts, & Matthew R, Cin, "*Unpacking India's Digital Personal Data Protection Act*", Lexology. Ropes & Gray LLP, (December 8, 2023), available at https://www.lexology.com/library/detail.aspx?g=505cf55a-8bd8-4202-9938-bb999a746faa (last visited March 12, 2024)

[73] Ibid.

[74] Digital Personal Data Protection Act of 2023 (Act no. 22 of 2000), s. 18

[75] Raktima Roy & Gabriela Zanfir-Fortuna, "The Digital Personal Data Protection Act of India, Explained" Future of Privacy Forum. (August 15, 2023). available at https://fpf.org/blog/the-digital-personal-data-protection-act-of-india-explained/ (last visited March 12, 2024)

[76] Digital Personal Data Protection Act of 2023 (Act no. 22 of 2000), s. 3(b)

However, it's unclear if this would also entail behaviour monitoring because the final version of the law omitted a clause mentioning this[77]. Additionally, Section 3(c) of the DPDP Act, includes a household exemption when data processing occurs for solely personal or domestic purposes[78]. Similarly, personal data made publicly available by the data subject or by law is not be covered by the Act.

In other words, the government can retain "personal data"[79] for an unlimited period regardless of whether the purpose for which it was collected has been served. A data protection law must safeguard and balance peoples' right to privacy and their right to information, which are fundamental rights flowing from the Constitution[80]. Therefore, DPDP Act 2023 is not effectively sufficient to protect the personal data in the digitalized world. Also, it unable to protect the dignity of the women, as it is not able to protect the personal data and that can increase various cybercrimes against women, like, cyber pornography, deepfake videos and so on.

The Data Fiduciary's responsibilities regarding AI-generated media are not expressly mentioned in the DPDPA. The duty outlined in Section 8(5), however, can be extended to guarantee that any unlawful content produced using information that is accessible to a Data Fiduciary is eliminated as soon as it is discovered. A Data Fiduciary must guarantee the "accuracy" and "completeness" of data under Section 8(3), even if it is "likely" that the data will be used to make a decision that has an impact on the Data Principal.

The word "likely" broadens the scope of Data Fiduciaries' obligations to consider deepfakes, including social media businesses. It should be noted that the business strategy of Data Fiduciaries is based on using users' personal information to target advertisements and show them content that is tailored to their interests. Even a deepfake itself could be considered personal information that is "likely" to be used to detect and decide activities that affect the Data Principal, as was previously stated. There are two ways to approach this.

First, a Data Fiduciary can display targeted material to the individual featured in it by using a deepfake as one of the input pieces of information for its algorithm. Second, a deepfake can be recommended to another user who consumes or is 'likely' to consume the content of the person featured in it.

In simple words The DPDP Act concerns itself with safeguarding the personal data of the individual from misuses by 'Data Fiduciaries'. A Data Fiduciary is defined in the Act to be any person who determines the means and the purpose for which the personal data of an individual would be used. A Data Principal is defined as an individual whose data is taken by such Data Fiduciary. The two more important definitions under the Act are that of 'personal data' and 'personal data breach.' Personal data under the act is defined as any piece of data relevant to a person which can be used to identify them. The use of the phrase 'any piece of data' widens the ambit of this definition to a large extent. This interpretation can be borrowed from the interpretation of the term 'any information' used in the definition of personal data under the EU Data Protection Laws. The obligation of the Data Fiduciary is attached by the definition of this term and, to a significant extent, the Act's whole framework. According to section 4, a Data Fiduciary may only use personal information for those for which the Data Principal has given clear authorisation. In addition, a Data Fiduciary may use personal information for additional permissible purposes that are explained in section 7. They mostly consist of following any ruling or legal obligation. This ensures that a Data Fiduciary cannot legally collect personal information about an individual without that person's explicit consent and utilize it to train any generative AI models. This approach might work well because most social media platforms and search engines have access to an astonishingly large amount of personal data.

---

[77] Hunter Dorwart, H., Gallan, J., & Rezzouk-Hammachi, V. "*Decrypting India's New Data Protection Law: Key Insights and Lessons Learned*". Lexology; Bird & Bird LLP. (October 13, 2023). available at https://www.lexology.com/firms/bird-and-bird/hunter_dorwart (last visited March 12, 2024)

[78] Digital Personal Data Protection Act of 2023 (Act no. 22 of 2000) s. 3(c)

[79] Digital Personal Data Protection Act of 2023 (Act no. 22 of 2000), s. 2(t)

[80] Anjali Bhardwaj & Amrita Johri, "The Problems with the Data Protection Bill" The Hindu. (February 21, 2023) available at: https://www.thehindu.com/opinion/op-ed/the-problems-with-the-data-protection-bill/article66531928.ece (last visited June 16, 2024)

Under this clause, Data Fiduciaries ought to be required to make sure that fraudulent AI-generated material disappears from their platforms. Since identifying and detecting deepfakes has been a significant problem, these regulations should require Data Fiduciaries to incorporate procedures for doing so. The signatures that are already there and specific to each piece of material are used by the current best practices to detect any content that may have been tampered with. But deepfake detection is not entirely covered by current technologies.
Consequently, any material that Data Fiduciaries may reasonably suspect to have been altered must have a warning attached to it. In addition, a Data Principal may request a reprieve under section 8(10), which requires a Data Fiduciary to set up a functional grievance redressal system. Upon receiving a complaint, a Data Principal has the right to request that their personal data be removed from a Data Fiduciary's platform if they discover that their data has been misused.

Right to be Forgotten
The "right to be forgotten" that the Data Principal has under section 12 of the DPDP Act 2023 further supports the Data Fiduciary's removal action. Article 17(2) of the General Data Protection Regulation (GDPR) also mentions this right. Finding a balance between this right and the freedom of speech and expression is crucial. The European case of *Google Sapin SL Google Inc.,v. AEPD,Mario Costeja Gonzalez* [81] addressed this issue by saying that the public's right to know that information must be balanced with the sensitivity of the personal data. It is also necessary to assess the degree to which it impacts a Data Principal's life. A deepfake has a far more serious effect on the person it features because it basically spreads false information about them.
The right to be forgotten has been acknowledged by Indian courts as a crucial component of the right to privacy guaranteed by Article 21 of the Indian Constitution. In the case of *Zulfiqar Ahman Khan v. Quintillion Business Media Pvt. Ltd*[82] "The right to be forgotten and the right to be left alone are inherent aspects" of the right to privacy, according to a Delhi High Court single-judge bench.

Furthermore, in the case of *Mahendra Kumar Jain v. State of W.B.*,[83] it was observed that by preserving an individual's right to privacy, Section 8(1)(j) of the Right to Information Act strengthens the protection granted to a person's reputation and dignity under Article 21. According to the Section, some information is deemed to be personal in character and is not subject to publication as it would not be in the public interest. In light of this, a Data Fiduciary is required to assist an individual in exercising their right to be forgotten in relation to a deepfake.

Nevertheless, appears that the issue of fraudulent generative AI-based media is not fully addressed by the DPDP Act. The Act's Section 3(c), which outlines the circumstances under which the Act would not apply, is pertinent in this situation. According to the first clause, when a person processes data for any domestic or personal reason, the Act is not applicable. The Act doesn't clarify what is meant by "personal or domestic." The issue of whether someone who obtains another person's personal information unlawfully may be regarded as a Data Fiduciary under the Act is also brought up. Artificial intelligence-generated bogus information utilised for domestic distribution might quickly become out of control, even if the answer is yes. This is due to the fact that a deepfake may spread over an ever-expanding social circle in the era of social media. Once shared with anybody in their family or social network, it would be practically hard for the creator to monitor where their deepfake is going.

[81] Global Freedom of Expression. "*Google Spain SL v. Agencia Española de Protección de Datos - Global Freedom of Expression*," (November 4, 2023). available at https://globalfreedomofexpression.columbia.edu /cases/google-spain-sl-v-agencia-espanola-de-proteccion-de-datos-aepd/. (last visited June 25, 2024)
[82] (2019) SCC Online Del 8494
[83] W.P.A. No. 17293 of 2021

Furthermore, it should be mentioned that the DPDPA does not distinguish between sensitive and personal data. Sensitive personal data is not given an extra layer of sufficient security since there is no such distinction. Sensitive personal data includes information that might be used to discriminate against a Data Principal, even while their personal data can be used to identify them. This contains details about their political views, medical history, sexual orientation, and ethnic origins. Under extremely certain conditions, access to such sensitive data is permitted under Article 9 of the GDPR. A deepfake created using private information might have more serious repercussions, such as systematic discrimination or swaying political opinion. Since the DPDPA does not make this difference, the current research examines the security of both of these data types collectively as well as the deepfakes that are produced from them.

A lawsuit was launched in Delhi High Court against the unregulated use of deepfake and artificial intelligence technologies, *Chaitanya Rohilla v. UOI* [84]. Although the Center's legal representative argued that the PIL's issues are within the purview of legislation and that the government is actively addressing them, the court emphasised the need to weigh competing interests and carry out lengthy discussions in order to reach a resolution.

The DPDP Rules, 2025 build upon the DPDP Act, 2023 by enhancing personal data protection mechanisms that indirectly strengthen safeguards against the misuse of data for deepfake creation. Rule 3 emphasizes the need for informed consent by mandating that notices to data principals clearly present details of personal data collection and its intended purposes, ensuring transparency. Rule 6 requires reasonable security safeguards, such as encryption and masking, to protect personal data against unauthorized access, which is often exploited in the creation of deepfakes. Furthermore, Rule 7 introduces strict timelines for notifying personal data breaches. Data fiduciaries must report breaches immediately to the Data Protection Board (DPB) and the affected individuals, with a detailed report required within 72 hours. This ensures timely intervention and reduces the risks of personal data being misused for generating deepfakes. Rule 4 also introduces Consent Managers, intermediaries registered with the DPB, to streamline and monitor the process of obtaining, managing, and withdrawing user consent. These measures collectively aim to minimize the risk of data exploitation and enhance accountability in data handling.

However, the rules leave critical gaps in directly addressing deepfakes. For instance, there are no explicit provisions or definitions for synthetic media or deepfakes, leaving ambiguity in their regulation. The rules do not mandate content authenticity standards, such as technologies to detect or label manipulated media, which could help protect users from deepfake-based misinformation or fraud. Rule 12 mentions the requirement for algorithmic audits for significant data fiduciaries, but it lacks specific directives to regulate AI models used to create deepfakes. Additionally, while penalties are specified for data breaches, accountability for entities actively creating or distributing deepfakes remains unclear, leaving room for misuse. Public education on recognizing and mitigating deepfakes is also absent, despite the increasing impact of such technologies on digital democracy and trust. Addressing these gaps in subsequent amendments or policies would provide a comprehensive framework for combating deepfake-related threats.

Policy Recommendations and Future Directions

Insufficient public understanding and consciousness, along with inadequate laws, worsen the impact of deepfakes. Laws and regulations have not kept up with technological advancements in many nations. Nonetheless, there have been developments in adjacent fields. For example, Mexico's Ley Olimpia, which forbids acts of digital violence in all its manifestations, has extended to other Latin

[84] Dixit, Pranav. "*Delhi High Court Grants Centre Two Weeks to Respond to PIL on AI and Deepfake Regulation*." Business Today (January 9, 2024). available at https://www.businesstoday.in/technology/news/story/delhi-high-court-grants-centre-two-weeks-to-respond-to-pil-on-ai-and-deepfake-regulation-412382-2024-01-09. (last visited June 25, 2024)

American nations.[85] By making it simpler to report this and other types of cyberaggression, such efforts enable the effective punishment of individuals who commit these crimes. Disseminating knowledge about deepfakes also aids in preventing a culture that minimises or ignores these practices. Public security agencies, together with digital platforms like social networks and other websites, need to put deepfake detection technology and other examples of digital violence into practice. Promoting stringent guidelines and measures is essential to halting the dissemination of photos and information that have been altered without the participants' permission. Furthermore, it's critical to use technology to generate pertinent reports. A few instances of technology being used for the benefit of people are El Salvador's LegalTech programme for LGBTQ persons. These resources make it easier for victims of gender-based violence who identify as other genders to get services.[86]

In the modern digital environment, deepfakes and other types of online abuse pose an increasing threat to women's safety and dignity. It is critical to handle this issue in its entirety in order to stop the wave from growing larger. It is imperative that the legal and security sectors, the technology sector, and the educational system work together and coordinate their efforts to safeguard the rights and well-being of women and other impacted populations.

The digital realm ought to be an area dedicated to development and ingenuity, where women can freely and securely express themselves without worrying about falling prey to another deepfake. Image generation and morphing are the two main methods used to make deepfakes. The method of creating a picture involves a neural network analysing facial expressions and using the provided samples to generate an image of its own. There has to be a way to distinguish between "cheapfakes" and "deepfakes."

Deepfakes are gradually beginning to attract the attention of a wider variety of individuals due to their increased accessibility and prevalence. Better quality and proliferation are two linked variables that have likely contributed to this trend. To begin with, compared to last year, the deepfakes that are currently doing the rounds are superior. The astounding rate at which deepfakes are evolving has led to a situation where it is almost difficult to identify well-made deepfakes.

Depending on the prevailing perceptions of women at play, the impact of deepfakes, particularly those that are sexually graphic, on women can vary depending on the nation. Journalist Rana Ayyub faced a deepfake pornographic smear campaign in India when she fought for the rights of an eight-year-old girl who had been sexually assaulted and died many times. These are not unusual experiences; women who have been the victims of image-based sexual abuse have stated that it is difficult for them to keep or obtain jobs. To make matters worse, several of the internet service providers where this kind of online abuse occurs have been hesitant to recognise the issue and have found it difficult to respond to it.[87]
Hence, there are three security implications associated with deepfakes: first, strategically, deepfakes have the potential to upset fragile peace on a global scale; second, deepfakes may be used as a weapon to unduly influence elections or the political process, or to discredit opposition, which is a national security concern—especially if foreign powers are involved in their creation and dissemination; and

[85] INFOBAE, available at https://www.infobae.com/america/mexico/2022/01/21/ley-olimpia-que-es-y-como-denunciar-si-comparten-fotos-intimas-sin-consentimiento/ (last visited June 25, 2024)

[86] Hidalgo, Nidia. "*LegalTech Para Personas LGBTQ+ En El Salvador - ¿Y Si Hablamos de Igualdad?" ¿Y si hablamos de igualdad?,* September 22, 2023. https://blogs.iadb.org/igualdad/es/servicios-legales-digitales-inclusivos-para-las-personas-lgbtq-un-prototipo-de-legaltech-en-el-salvador/. (last visited June 25, 2024)

[87] Ananya Bhattacharya, "*I Was Vomiting: Journalist Rana Ayyub Reveals Horrifying Account of Deepfake Porn Plo*t." India Today, (November 21, 2018) available at:. https://www.indiatoday.in/trending-news/story/journalist-rana-ayyub-deepfake-porn-1393423-2018-11-21. (last visited June 25, 2024)

third, on a personal level, the potential for using deepfakes to create sexually explicit video could disproportionately harm women, especially in public.

Encouraging service providers to get this consent prior to receiving an order to produce a deepfake is one way to implement this policy. Legislators might wish to take into account "know your customer" laws that are now in place in the banking and financial services sectors, given the possible harm that deepfakes might do. Of course, offshore companies can easily get around such laws because internet services can be accessed from anywhere in the world, but customers within the jurisdiction risk legal consequences if they falsely claim to have provided images with the consent of the person featured.

Legislative or legal measures such as those above are unlikely to end the harmful use of deepfakes because of the boundaries that separate countries and the cyberspace. Policymakers should thus think carefully about the best ways to educate digital literacy. The idea behind deepfakes is that "seeing is believing." [88]Media literacy initiatives are thriving like mushrooms to counteract this deeply ingrained prejudice, and Microsoft is among the corporations releasing tools that should assist the general public in determining the veracity of video material. Worldwide school curriculum are being established to incorporate global standards for digital literacy. Additionally, special consideration has to be given to cohorts who can be difficult to reach because of things like restrictions on general technical literacy and the effects of the digital divide.
The advent of deepfake technology presents hitherto unseen obstacles to the global legal and social structure of human civilisation. While certain parts of deepfakes are covered by the framework that now exists, there is an obvious need for targeted regulation, judicial clarification, and public involvement to stop the improper use of this technology. While genuine AI-generated pornography enterprises confront a distinct set of legal challenges to ensuring they operate within the law, deepfakes continue to haunt communities. As AI-generated photos and films become increasingly lifelike through technological improvements, society must confront the moral and legal ramifications of these developments. Individual rights must be upheld, and appropriate action must be taken against those who abuse these tools for malevolent ends. a wide range of legal concerns, such as liability, consent, privacy, intellectual property, and regulatory compliance. It takes careful preparation, strong procedures, and adherence to ethical norms to navigate these legal complications. Thus, proactive legal risk management is essential in this developing industry.

.

[88]Rituparna Bhattacharjee, "*Understanding Interpretation via Jurisprudential Lenses*." *Issue* 5, P 623. *INT'L JL MGMT. & HUMAN* (2022).