Lex Scientia Law Review
ISSN 2598-9677 (Print) 2598-9685 (Online)
Vol. 8 Issue 2 (2024) 859-896
DOI: https://doi.org/10.15294/lslr.v8i2.13749
Online since: December 15, 2024

LEX SCIENTIA
LAW REVIEW
https://journal.unnes.ac.id/journals/lslr/index

Digital Deception: The Impact of Deepfakes on Privacy Rights

Karisha Verma [a] ✉

[a] Faculty of Law, University of Lucknow, Lucknow, India

✉ Corresponding email: rs2022law_karishma@lkouniv.ac.in

Abstract

Deepfake technology, which uses advanced artificial intelligence to create synthetic media, poses significant threats to privacy rights. Since its emergence, deepfakes have been used in various malicious ways, raising urgent concerns about their impact on privacy rights. This study investigates the implications of deepfake technology on privacy, with a focus on how it affects individuals and legal frameworks. The research is driven by the need to understand the extent of privacy violations and the adequacy of current laws in addressing these challenges. The article aims to provide a comprehensive analysis of the intersection between deepfake technology and privacy rights. It explores the theoretical implications of deepfakes on privacy, assesses public awareness and concern through empirical research, and evaluates existing legal frameworks in the United States, European Union, and India. A mixed-method approach is used in this article. Doctrinal research involves examining profound impact of deepfake technology on privacy rights and analysing legal frameworks and case law to understand the legal responses to deepfakes. Empirical research includes a survey conducted with diverse respondents in India to gauge public awareness, experiences, and opinions regarding deepfakes and its incursion into privacy rights. The study finds that while some jurisdictions have enacted laws to combat deepfakes, significant gaps remain in protecting privacy rights. Empirical findings reveal varying levels of public awareness and concern, highlighting the need for more robust legal measures and public education. The research underscores the necessity for updated and comprehensive legal frameworks to address the evolving challenges posed by deepfake technology. Recommendations include enhancing legal protections while coming up with technological solutions and increasing public awareness to safeguard privacy in the digital age.

KEYWORDS *Deepfake, Deepfake Pornography, Fundamental Right, Privacy, Regulatory Framework*

Introduction

The rapid rise of Artificial Intelligence (AI) and, more specifically, deepfake technology has ignited critical discussions on its impact on society, privacy, and security. Deepfakes, which involve the manipulation of video, image, and audio content using advanced machine learning techniques such as Generative Adversarial Networks (GANs), first surfaced on Reddit in 2017, where the faces of celebrities were superimposed onto the bodies of pornographic actors.[1] Since then, the technology has evolved at an alarming rate, becoming so sophisticated that it is increasingly difficult to distinguish between genuine and fabricated content. The threat posed by these hyper-realistic, AI-generated images and videos has grown substantially, not only for public figures but also for ordinary individuals. This rise in deepfake technology poses significant risks, particularly in a world where 5.45 billion people which amounts to 67.1% of the global population are connected to the internet, with 5.17 billion or 63.7 % being active on social media.[2] The ease of sharing

[1] Mika Westerlund, "The emergence of Deepfake Technology: A Review," *Technology Innovation Management Review* 9, no. 11 (November 2019): 40, https://doi.org/10.22215/timreview/1282.

[2] Ani Petrosyan, "Number of internet and social media users worldwide as of July 2024," Statista, August 19, 2024, https://www.statista.com/statistics/617136/digital-population-worldwide/. Statista is a global data and business intelligence platform that offers a vast collection of statistics, reports, and insights on over 80,000 topics, sourced from 22,500

content online, combined with the increasing accessibility of deepfake creation tools, has made it easier than ever for malicious actors to use this technology for harmful purposes.

The growth of deepfake content has been staggering. The "2023 State of Deepfakes" report by Security Hero[3] revealed that the number of deepfake videos online had skyrocketed to 95,820, marking a 550% increase since 2019.[4] Similarly, DeepMedia[5], a firm specializing in deepfake detection, predicted that 500,000 video and voice deepfakes would be shared globally on social media by the end of 2023.[6] The proliferation of these synthetic media tools has led to an increase in both personal and public harms. The Deeptrace report titled "The State of Deepfakes: An Overview" identified 14,678 deepfake videos in circulation in 2019, with the number doubling every six months and this number continues to rise exponentially.[7]

The challenge is not only the creation of realistic deepfakes but also the ecosystem that supports their amplification. Deepfakes are shared through social media platforms, websites, and fake accounts that comment, share, and generate misleading information. These media forms create a dangerous mix of misinformation that can distort reality for the viewer. For example, in 2023, UK figures like Sadiq Khan and Keir Starmer were targeted by deepfakes, with falsified audio and video clips circulating widely. A deepfake video of Khan appeared to show him downplaying Armistice Day,[8] while a fabricated video of

different channels across 170 industries. This platform provides valuable data and analysis for various sectors, making it a vital resource for businesses, researchers, and professionals.

[3] Security Hero is a resource website dedicated to helping individuals protect themselves from identity theft and enhance their online privacy. Founded with a mission to prevent 100,000 cases of identity theft by 2025, Security Hero offers guidance, research data, product reviews, and cybersecurity tips to empower users in the digital world. The platform focuses on educating people about identity protection tools and practices.

[4] "2023 State of Deepfakes: Realities, Threats, and Impact," Security Hero, last accessed September 10, 2024, https://www.securityhero.io/state-of-deepfakes/.

[5] Deep Media is a Deepfake Detection and Media Intelligence company.

[6] Justin Mcgill, "24 Deepfake Statistics – Current Trends, Growth, and Popularity (December 2023)," *Contentdetector.AI,* May 23, 2024, https://contentdetector.ai/articles/deepfake-statistics/.

[7] Henry Ajder et al., "The State of Deepfakes: Landscape, Threats, and Impact," Deeptrace, September, 2019, https://finaletheorie.org/download/the-state-of-deepfakes-an-overview/?wpdmdl=30224&refresh=66dfda870b9f11725946503.

[8] Marianna Spring, "Sadiq Khan says fake AI audio of him nearly led to serious disorder," *BBC News*, February 14, 2024, https://www.bbc.com/news/uk-68146053.

Starmer depicted him mistreating staff during the Labour Party conference.[9] In the United States, explicit deepfakes of Taylor Swift were shared on social media in early 2024, gaining over 47 million views on platform X before they were removed.[10] These incidents underscore the significant reputational damage and emotional harm that deepfakes can inflict on individuals. Similarly, deepfake technology has been used to influence international events, such as the video of Ukrainian President Volodymyr Zelensky, which falsely portrayed him urging Ukrainian soldiers to surrender during the war with Russia.[11]

However, deepfakes are not inherently malicious and have legitimate applications in fields such as cinema, education, and healthcare. In the entertainment industry, deepfakes are used to create realistic special effects in movies, bringing historical figures back to life for educational purposes or personalizing digital avatars. They are also employed in healthcare to enhance imaging techniques for the detection of cancerous tumors. Despite these positive uses, the potential for misuse far outweighs these benefits. There is a plethora of possibility of commission of crimes using the technology of deepfake. The technology itself does not pose a threat, however, it can be used as a tool to commit crimes against individuals and society. The potential threats include identity theft, virtual forgery, misinformation against governments, hate speech, online defamation, manipulation of elections, violation of privacy, revenge porn, infringement of personality rights, and the manipulation of digital evidence.

Amid various threats of deepfake, the threat to right to privacy of individuals is the most concerning with vulnerable groups, particularly women and children, being at the forefront. The ability to fabricate realistic, explicit content has led to the rise of non-consensual pornography and sextortion, where deepfake videos are used to exploit and harass victims. A recent case involving Taylor Swift highlights the widespread reach of deepfake technology, with explicit content being shared across various platforms, leading to significant

[9] "Deepfake audio of Sir Keir Starmer released on first day of Labour conference," *Sky News*, October 9, 2023, https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir-starmer-12980181.

[10] Kate Conger and John Yoon, "Explicit Deepfake Images of Taylor Swift Elude Safeguards and Swamp Social Media," *The NewYork Times,* January 26, 2024, https://www.nytimes.com/2024/01/26/arts/music/taylor-swift-ai-fake-images.html#:~:text=One%20image%20shared%20by%20a%20user%20on%20X,spread%20despite%20those%20companies%E2%80%99%20efforts%20to%20remove%20them.

[11] Jane Wakefield, "Deepfake presidents used in Russia-Ukraine war," *BBC News,* March 18, 2022, https://www.bbc.com/news/technology-60780142.

psychological harm. Women, in particular, are often disproportionately targeted, as seen in the 2018 deepfake scandal involving Indian journalist Rana Ayuub, whose manipulated sex-video was circulated widely on social media.[12] As the Warren and Brandeis argument from the late 19th century still holds,[13] these modern technological advancements infringe on an individual's right to privacy, making the lack of robust legal frameworks even more alarming.

Such incidents highlight the urgent need for regulatory intervention to address the growing misuse of this technology. Although certain regions have enacted laws, such as California's regulations against deepfake pornography[14], these measures fail to address the broader impact on privacy rights. In India, where the legal framework is still developing, deepfakes pose a growing threat, yet there is minimal legal infrastructure to combat these crimes effectively.

Although much research has focused on detection techniques and case studies, there remains a significant gap in comprehensive studies examining the global impact of deepfakes on privacy rights. This paper aims to fill that gap by offering a comparative legal analysis of how deepfake-related privacy issues are addressed in the US, EU, and India. Furthermore, it will assess public awareness and the social implications of deepfake technology. The absence of a cohesive, global legal response to deepfakes necessitates a closer examination of how legal systems can evolve to protect individuals from the harmful effects of this growing technological phenomenon.

This study employs a mixed-method approach, utilizing both doctrinal legal analysis and empirical research to investigate the impact of deepfake technology on privacy rights. The chosen methodology ensures a thorough exploration of the topic by addressing both theoretical and practical implications of deepfakes, especially concerning privacy violations in different jurisdictions.

[12] "Incident 769: Investigative Journalist Rana Ayyub Targeted by AI-Generated Deepfake Pornography," AI Incident Database, last accessed August 10, 2024, https://incidentdatabase.ai/cite/769/.

[13] Samuel D. Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4, no. 5 (December 1890): 193-220, http://links.jstor.org/sici?sici=0017-811X%2818901215%294%3A5%3C193%3ATRTP%3E2.0.CO%3B2-C. The authors argued in this paper that the technological developments pose a serious threat to right to privacy.

[14] California Civil Code § 1708.86 (2020). California Assembly Bill 602 (AB 602) - Depiction of individual using digital or electronic technology: sexually explicit material: cause of action. Last accessed August 13, 2024, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602.

The doctrinal research component focuses on the implication of deepfake technology on privacy rights, drawing upon existing legal frameworks and academic literature. The study further examines the United States (US), the European Union (EU), and India, highlighting relevant legal frameworks in each region. In the US, specific state laws such as those in California, Texas, and Virginia are assessed, focusing on their approach to deepfake-related privacy violations, including pornography and election interference. The EU's General Data Protection Regulation (GDPR) is explored in relation to deepfakes, particularly concerning privacy and data protection. In India, the research evaluates the Information Technology Act, 2000, alongside recent developments in Bhartiya Nyay Sanhita, identifying gaps in the regulatory framework surrounding deepfake misuse. Additionally, a comparative legal analysis is undertaken to contrast the legal responses of the US, EU, and India, seeking to identify common approaches and potential areas for harmonizing legal regulations on deepfakes across borders.

For the empirical aspect, this study gauges public awareness and perceptions of deepfake technology's impact on privacy rights. The empirical research is conducted in India, where data was collected through a structured questionnaire distributed via Google Forms to 225 respondents from diverse backgrounds, including students, academicians, and the general public. The survey gathered information on demographics, awareness of deepfakes, experiences with deepfake content, perceptions of privacy violations, and opinions on deepfake regulations. The data collected was analyzed using Statistical Package for the Social Sciences (SPSS) software, focusing on key objectives such as awareness levels, encounters with deepfakes, victimization, privacy concerns, and opinions on legal frameworks.

The study integrates both doctrinal and empirical components to provide a comprehensive analysis of deepfake technology's impact on privacy rights. The empirical findings are used to support and challenge the doctrinal legal theories, thereby allowing a deeper understanding of the effectiveness of existing legal protections. This integration highlights the need for reforms by considering both legal theory and public opinion, offering a balanced and holistic view of how privacy rights are affected by the misuse of deepfakes.

In this study, the scope is limited to the jurisdictions of the US, EU, and India, focusing on the legal frameworks and public awareness of deepfake technology.

Deepfake: A Throttling Attack on Privacy

Privacy is a fundamental human right essential for personal freedom and autonomy and it is safeguarded by multiple international frameworks and domestic legal systems around the world. Article 12[15] of the Universal Declaration of Human Rights (UDHR) and Article 17[16] of the International Covenant on Civil and Political Rights (ICCPR) underscore the importance of privacy, prohibiting arbitrary interference with one's personal life. Moreover, Article 8[17] of the European Convention on Human Rights (ECHR) guarantees the right to respect for private and family life, a provision that has expanded to address digital privacy in today's interconnected world. In India, the Supreme Court's landmark judgment in K. S. Puttaswamy v. Union of India[18] recognized privacy as an intrinsic part of the right to life and personal liberty under Article 21 of the Indian Constitution and categorically held that right to privacy is a fundamental right. Similarly, in the United States, privacy has been upheld as a constitutional right through pivotal cases like Griswold v. Connecticut[19] and Roe v. Wade[20], which protect an individual's right to privacy within the broader framework of personal liberty.[21]

In the digital age, privacy extends beyond physical spaces, encompassing the virtual realm. The digital transformation has led to vast amounts of personal data being shared online, making individuals more susceptible to privacy violations. This expanded concept of privacy which is termed as digital privacy

[15] Article 12 of UDHR provides "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."

[16] Article 17 of ICCPR reiterates the same thing as article 12 of UDHR.

[17] "(1) Everyone has the right to respect for his private and family life, his home and his correspondence. (2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others."

[18] 2019 (1) SCC 1.

[19] 381 U.S. 479 (1965).

[20] 410 U.S. 113 (1973).

[21] The 1965 landmark case Griswold v. Connecticut was a pivotal moment in constitutional law and the broader discourse on privacy rights. The ruling highlighted the implied right to privacy in the U.S. Constitution, specifically limiting government intervention in private matters like contraceptive use. This decision became a key precedent for the U.S. Supreme Court in future cases, including Roe v. Wade, which expanded the right to privacy by affirming individuals' bodily autonomy and legalizing abortion.

now includes protection against unauthorized data collection, online surveillance, and misuse of personal information.

In this modern context, the right to privacy is constantly challenged by evolving technologies, especially Artificial Intelligence (AI)-based tools like deepfakes. Deepfakes exploit personal data, including images, audio, and videos, to create highly convincing synthetic media that often violate privacy in invasive and harmful ways. By manipulating this digital content, malicious actors can fabricate lifelike scenarios that infringe on both the personal and digital privacy of individuals. The danger posed by deepfakes is particularly grave because they blur the lines between reality and fiction, making it easier to exploit individuals without their consent or knowledge. This evolving threat not only harms individuals' reputations but also violates their digital autonomy, making robust legal protections more crucial than ever.

Deepfake, as the name suggests, refers to fake (synthetic) image or audio-video content which is created using Generative Adversarial Networks (GAN)[22], a deep learning architecture.[23] It is the product of advancement in Artificial Intelligence. In a very crude sense, deepfakes are synthetic media often created by using someone else's image or likeness to swap it with existing content. The likeness may include facial features or voice. Through this technology original content can also be created using individual's images and voice showing something which he/she never did or said leading to distortion of reality.[24] The deceptive credibility of deepfake often makes a person believe in something which is not true. These synthetic media forms can be used to create non-consensual explicit content, manipulate public figures, or forge identities—each constituting a serious violation of privacy. The invasive nature of deepfakes allows perpetrators to infringe upon individuals' personal autonomy and reputation, often without their knowledge or consent. For vulnerable groups, especially women and children, deepfakes have been weaponized in the form of

[22] "Generative Adversarial Network (GAN)," Geeks for Geeks, last updated August 9, 2024, https://www.geeksforgeeks.org/generative-adversarial-network-gan/.

[23] Meredith Somers, "Deepfakes, Explained," MIT Sloan School, July 21, 2020, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained; Matthew B. Kugler; Carly Pace, "Deepfake Privacy: Attitudes and Regulation," *Northwestern University Law Review* 116, no. 3 (2021): 611-680.

[24] Bobby Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *California Law Review* 107, no. 6 (2019): 1753-1820, https://www.jstor.org/stable/26891938.

revenge porn, identity theft, and sextortion, causing emotional distress and long-term harm.[25]

In the era of social media, the boundaries of privacy have become increasingly blurred, particularly with the advent of technologies like deepfakes. The ease with which smartphones can capture images and videos highlights the importance of an individual's right to control how their image is used. Yet, despite this need, current legal protections for unauthorized use of an individual's image or likeness remain insufficient. Traditionally, the law has protected individuals' images for commercial gain under the tort of passing off, but it has been inadequate in addressing the personal right to privacy, especially when the use of one's image does not involve commercial exploitation. An individual's right to control their image is rooted in human dignity, a concept that Samuel Warren and Louis Brandeis emphasized in their influential 1890 essay. They argued that an "intrusion-free sphere" is central to preserving human dignity, which warrants legal protection.[26]

The concept of situational privacy[27], as described by Alan Westin, further complicates the issue. Individuals expect a degree of anonymity in public spaces, where, although they are observed, they do not anticipate being personally identified or monitored. This expectation forms the basis of mass privacy- the right to exist freely in public without constant surveillance.[28] NA Moreham echoes this sentiment, noting that people adjust their behavior based on the assumption of anonymity, which is lost when under surveillance.[29] This intrusion violates autonomy and dignity, causing humiliation, emotional distress, and a loss of peace of mind. The tort of misappropriation of personality aims to protect these privacy interests, offering remedies for the violation of one's autonomy, dignity, and the right to control how their image is perceived publicly.

Deepfakes strike at the very heart of these privacy values, intruding upon an individual's image and reputation. They can have far-reaching consequences,

[25] Drew Harwell, "Fake-porn videos are being weaponized to harass and humiliate women: Everybody is a potential target," The Washington Post, December 30, 2018, https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/.

[26] Warren and Brandeis, "The Right to Privacy," 193-220.

[27] Alan F. Westin, *Privacy and Freedom*, (New York: Atheneum, 1967), 459, https://archive.org/details/privacyfreedom00west/page/n22/mode/1up; Daniel Solove and Marc Rotenberg, *Information Privacy Law* (New York: Aspen Publishers, 2006), 37.

[28] Solove and Rotenberg, Information privacy law, 37.

[29] N A Moreham, "Privacy in Public Spaces," *The Cambridge Law Journal* 65, no. 3 (November 23, 2006): 606-635, https://doi.org/10.1017/S0008197306007240.

affecting one's opportunities, livelihood, relationships, and even mental health. Deepfakes disrupt the private sphere where individuals develop their thoughts, goals, and relationships, violating their autonomy and trust. A 2019 study found that 96% of deepfakes on the internet were deepfake pornography, disproportionately exploiting women's images without their consent.[30] The creation of non-consensual deepfake pornography is not merely an invasion of privacy but a violation of sexual privacy, which sits at the core of human dignity.[31]

Sexual privacy dictates a person's right to control their intimate life, including decisions about whether to share or withhold private images of their body. The concept of sexual autonomy allows individuals to manage the boundaries of their intimacy, trust, and vulnerability. Deepfake pornography obliterates these boundaries by forcibly inserting individuals into sexually explicit media without their consent, fundamentally breaching their right to self-determination.[32] Justice DY Chandrachud, in the K.S. Puttaswamy v. Union of India judgment, recognized privacy as the "constitutional core of human dignity," extending its protections to bodily and sexual autonomy. Bodily autonomy, as a subset of privacy, grants individuals the power to make decisions about their body, free from coercion or exploitation.[33]

Deepfake porn not only violates sexual privacy but also diminishes the individual's identity by presenting a manipulated version of their body and face to the public, which they never consented to share. The case of Rana Ayyub, an Indian journalist, serves as a stark example of this. In 2018, a deepfake video of her was circulated on social media in retaliation for her criticisms of the government, accompanied by threats of gang rape.[34] This malicious use of deepfake technology eclipses personhood, stripping individuals of their self-determination and reducing them to mere objects in the eyes of their aggressors. The psychological toll of such violations is immense, often leading to long-term mental health consequences.

The use of deepfakes, particularly in non-consensual pornography, is deeply rooted in control - not only over the individual's body but also over how they are publicly perceived. The misappropriation of identity through deepfakes infringes on personal autonomy and sexual agency, eroding the victim's control

[30] Ajder et al., "The State of Deepfakes," 2019.

[31] Danielle Citron, "Sexual Privacy," *The Yale Law Journal* 128, no. 7 (May 2019): 1870-1960, https://www.jstor.org/stable/45098020.

[32] Citron, "Sexual Privacy," 1870-1960.

[33] K. S. Puttaswamy v. Union of India, 2019 (1) SCC 1.

[34] "Incident 769: Investigative Journalist Rana Ayyub Targeted", 2024.

over their own intimate identity. As Danielle Citron emphasizes, sexual privacy governs the boundaries around intimate life, protecting individuals from having their intimate details, whether fantasies, sexual desires, or private communications, exposed without consent. While deepfakes may not always depict the naked bodies of their victims, their misuse in the form of sexualized face-swapping is still an egregious violation of sexual autonomy and privacy.[35]

At the core of non-consensual deepfake pornography is the lack of consent, with an additional layer of harm caused by the fact that the individual being depicted never actually engaged in the sexual act in question. The profound psychological damage that such portrayals cause to victims underscores the need for legal systems to evolve and address these new, technology-driven invasions of privacy. As Citron notes, the essence of the deepfake pornography crisis lies in the lack of consent and its broader implications for sexual identity and autonomy.[36]

While international frameworks like the UDHR and ECHR emphasize the need to protect privacy in both physical and digital contexts, the legal landscape is struggling to keep pace with the rapid development of deepfake technology. Although jurisdictions like the United States and the European Union have begun enacting preliminary laws to regulate deepfakes, these legal frameworks often lag behind the fast-moving technological advancements. In India, where privacy has only recently been recognized as a fundamental right, the lack of comprehensive legal protections leaves individuals particularly vulnerable to deepfake abuses.

Impact Assessment of Deep Fake on Privacy Rights

This section aims to assess public awareness and perceptions about deepfakes and the impact of deepfake technology on privacy rights with the help of empirical data. A structured questionnaire was distributed via Google Forms to respondents from diverse backgrounds in India, including students, academicians, and the general public. A total of 225 responses were collected and analyzed using SPSS software, focusing on key aspects such as awareness of deepfakes, experiences with deepfake content, and concerns about privacy violations as shown and illustrated on Table 1.

[35] Citron, "Sexual Privacy," 1870-1960.

[36] Citron, "Sexual Privacy," 1870-1960.

Demographics

The Table 1 reveals a diverse distribution of individuals across different age groups, occupations, and genders. The majority of respondents are between 18 and 27 years old, comprising 46.2% of the total, followed by those aged 28 to 37 years at 35.6%. People under 18 years make up a smaller portion at 4.0%, and those 38 years and older account for 14.2%. In terms of occupation, students represent the largest group at 50.2%, while working professionals follow at 40.4%. Homemakers are a minor segment at 3.6%, and those in other occupations make up 5.8%. Gender-wise, there is a higher proportion of males at 55.6% compared to females at 44.4%. The data indicates a strong prevalence of social media usage among the surveyed individuals. A significant majority, 87.1%, are active social media users, while only 12.9% do not engage with social media platforms.

Furthermore, the demographic data shows a majority of young, digitally literate respondents, with students and working professionals making up the largest groups. Gender is relatively balanced, but the data may reflect a slightly male-dominant perspective. Notably, 87.1% of respondents are active social media users, indicating high exposure to deepfakes and related privacy risks. Moreover, active social media users are not only consumers but also content creators, which increases their vulnerability to having their images, videos, or personal data misused for the creation of deepfakes. The pervasive use of social media thus amplifies the risk of privacy violations, as individuals often have a large amount of personal content available online, which can be exploited for malicious purposes. This reinforces the need for stringent legal frameworks to protect individuals from the unauthorized use of their likeness in deepfakes.

Awareness about Deepfake

Figure 1 provides insights into awareness and experiences related to deepfakes. A majority of respondents, 55.6%, have read about deepfakes and possess a basic understanding, while 21.8% have good knowledge and awareness. However, 3.6% have never heard of deepfakes, and 19.1% are aware of them but lack detailed knowledge.

The high level of awareness (77.4% combining basic and good knowledge) reflects growing public familiarity with deepfakes. However, the lack of detailed knowledge in nearly a quarter of respondents (22.7%) indicates a gap that could affect the perceived severity of privacy violations posed by deepfakes. This supports the argument that, while awareness is increasing, a deeper understanding of the technology is necessary to fully grasp its implications for privacy and security. The 3.6% who have never heard of deepfakes highlight that despite the technology's growing presence, public awareness is not yet universal. The lack of detailed awareness may also contribute to underestimating the privacy threats posed by deepfakes, reinforcing the need for educational initiatives alongside legal reforms.

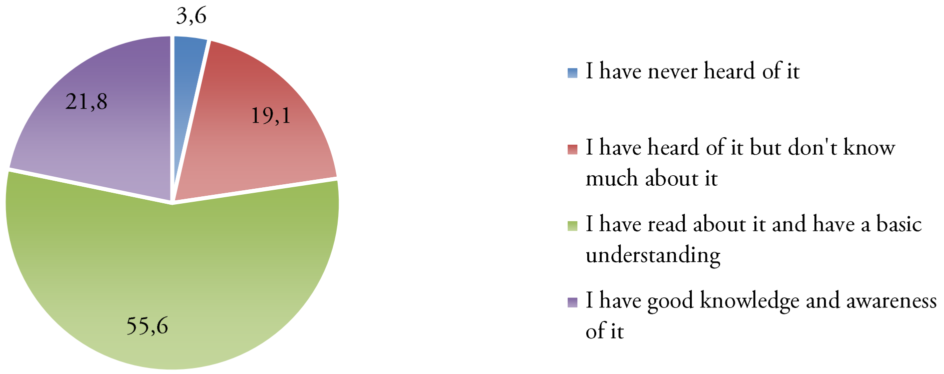

FIGURE 1. Awareness about Deepfake

Encounters With deepfake

A significant portion of the respondents (64.4%) had encountered deepfake content, while 35.6% had not. When it came to identifying deepfakes, 17.8% of respondents reported being able to identify them easily, while 49.3% could recognize deepfakes but were not always certain. A notable 19.6% found it difficult to identify deepfakes, and 13.3% had never attempted to identify such content (see Figure 2).

The Figure 2 indicates that deepfake content is relatively prevalent, with 64.4% of respondents having encountered it. However, the ability to accurately identify deepfakes remains inconsistent. While 17.8% of individuals find it easy to identify deepfakes, a larger segment (49.3%) recognizes them but with uncertainty. This suggests that, although awareness of deepfakes is increasing, distinguishing them from authentic content remains a challenge for many. The 19.6% who find it difficult to identify deepfakes, along with the 13.3% who have never attempted to do so, further highlight a significant gap in recognition skills. This lack of certainty underscores the potential for deepfakes to deceive a large portion of the population, exacerbating the privacy threats and misuse risks posed by this technology. There is a clear need for more public education and tools to help individuals detect deepfakes effectively.

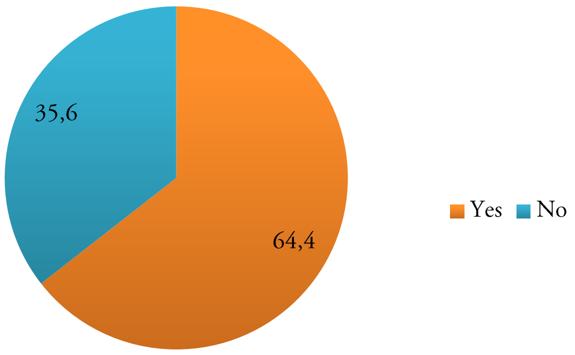

FIGURE 2. People who have come across any incident of deepfake on social media

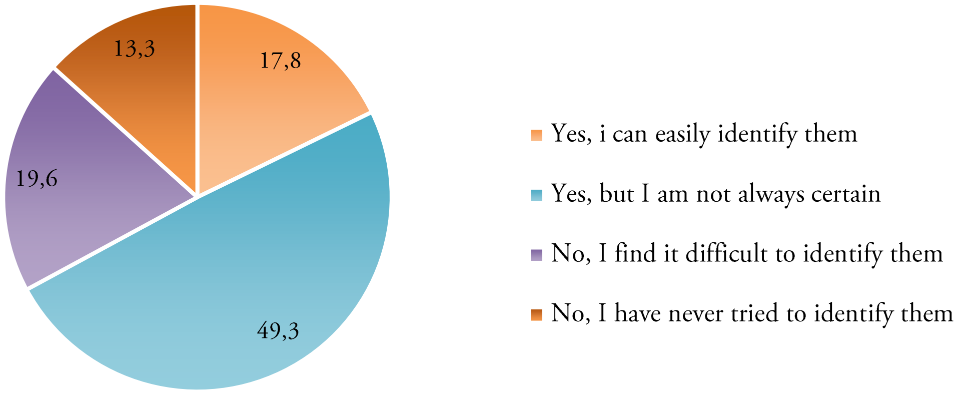

FIGURE 3. Ability to identify deepfake image/audio/video

Victims of misuse of Deepfakes

The Figure 3 indicated that 6.7% of respondents had been victims of deepfake misuse, while 93.3% had not. Among those affected, 9.5% had acquaintances who were victims, 9.0% had close friends affected, and 2.3% mentioned family members. A majority (76.6%) did not report any personal or close connections to deepfake misuse.

Therefore, the Figure 3 reveals that direct victimization by deepfakes is relatively rare, with only 6.7% of respondents reporting personal experiences of deepfake misuse. However, the broader social impact is somewhat more significant, as 9.5% knew acquaintances, 9.0% had close friends, and 2.3% had family members who were victims. Despite this, the majority (76.6%) did not have any personal or close connections affected by deepfake misuse.

This suggests that, while the immediate risk of deepfake misuse might appear limited, it could be underreported or not yet widespread within personal circles. The minority impacted may still represent the potential for harm, especially as awareness grows and misuse becomes more sophisticated. These findings emphasize the importance of pre-emptive legal and regulatory measures, as deepfake misuse may escalate over time, affecting a larger portion of the population.

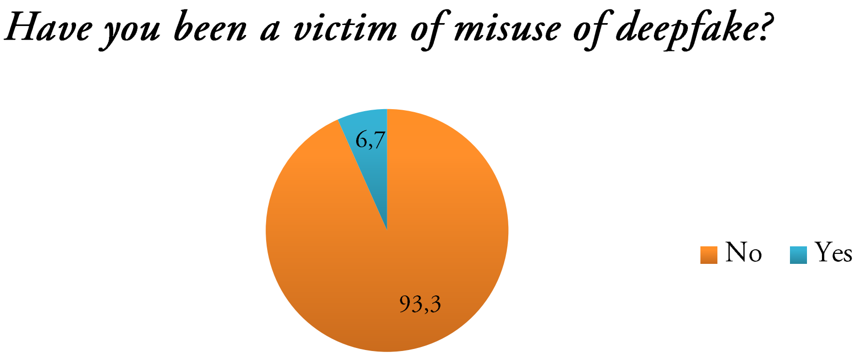

FIGURE 4. Victim identification I

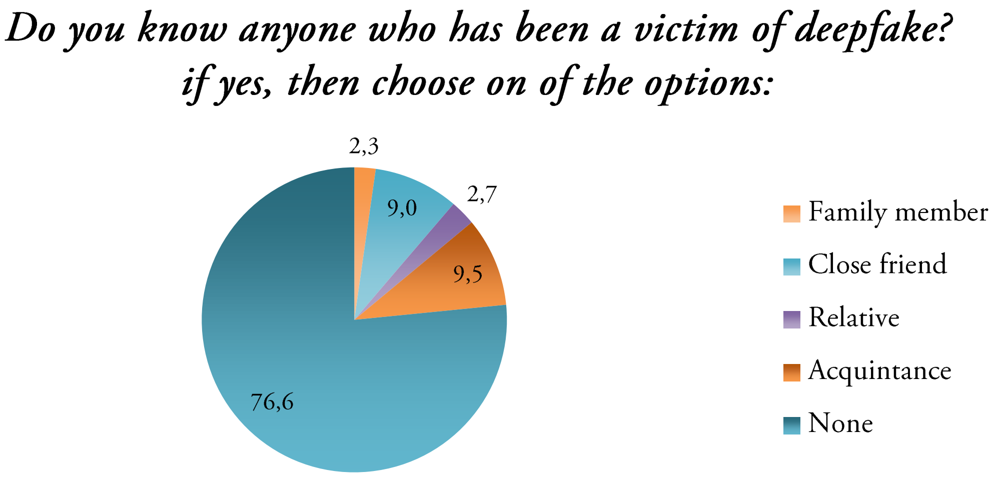

FIGURE 5. Victim identification II

Awareness of misuses of deepfake

The data reveals a widespread awareness of deepfake misuse, with 92.0% of respondents acknowledging the potential harm associated with this technology. The awareness is not limited to a single type of misuse but spans across a range of malicious applications. Commonly reported forms of misuse include the creation of fake news and misinformation (9.1%), revenge porn and explicit content (8.2%), identity theft and fraud (5.8%), and defamation (5.8%). Additionally, political manipulation (3.4%) and violations of personality rights (4.3%) also stand out as recognized concerns. A significant 88.2% of respondents understand that deepfakes can be misused in multiple harmful ways.

This broad understanding indicates a strong recognition of the dangers posed by deepfakes, particularly in exploiting privacy, personal security, and public trust. The diversity of misuse cases emphasizes the urgency for robust legal frameworks to address the wide-ranging impact of this technology.

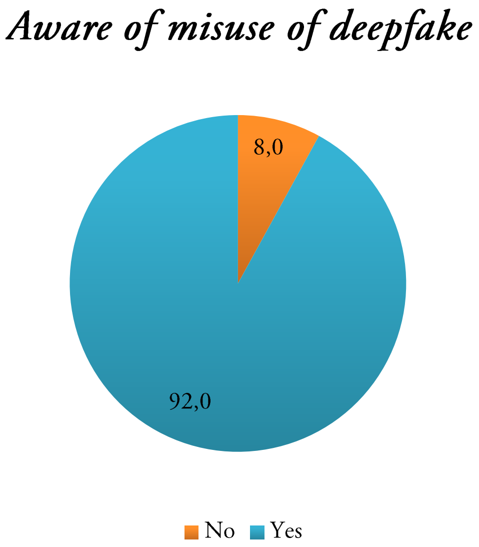

FIGURE 6. Awareness of misuse of deepfake

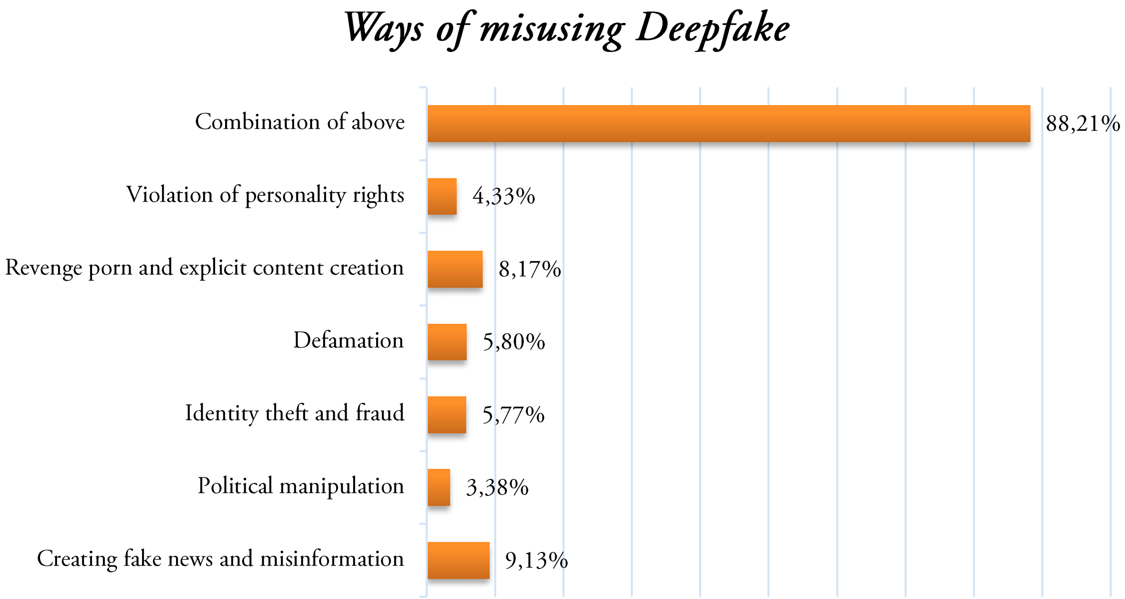

FIGURE 7. Misusing Deepfake Motive

Concern of people regarding impact of deepfakes on privacy rights

The data indicates that the majority of respondents are deeply concerned about the impact of deepfakes on privacy rights. A significant 69.3% are very concerned, while 24.0% are moderately concerned. Only a small fraction feels slightly concerned (4.4%) or not concerned at all (2.2%). Furthermore, an overwhelming 96.9% believe that deepfakes violate individual privacy, underscoring widespread recognition of the threats posed by this technology.

The data highlights a pervasive concern among respondents regarding the impact of deepfakes on privacy rights. An overwhelming 69.3% are highly concerned, with an additional 24.0% expressing moderate concern, indicating that nearly all respondents view deepfakes as a significant threat to personal privacy. Only a small minority (6.6%) exhibit slight or no concern about these implications, which is reflective of either limited exposure or differing views on privacy risks. A striking 96.9% of respondents explicitly recognize that deepfakes violate individual privacy. This broad consensus underscores the serious risks posed by the technology, particularly in terms of compromising personal security, identity, and public trust. These findings strongly indicate that deepfakes represent a substantial and growing threat to privacy rights, validating the argument for a more rigorous and comprehensive legal framework to address these concerns. The near-universal recognition of privacy violations further amplifies the urgency for regulatory measures.

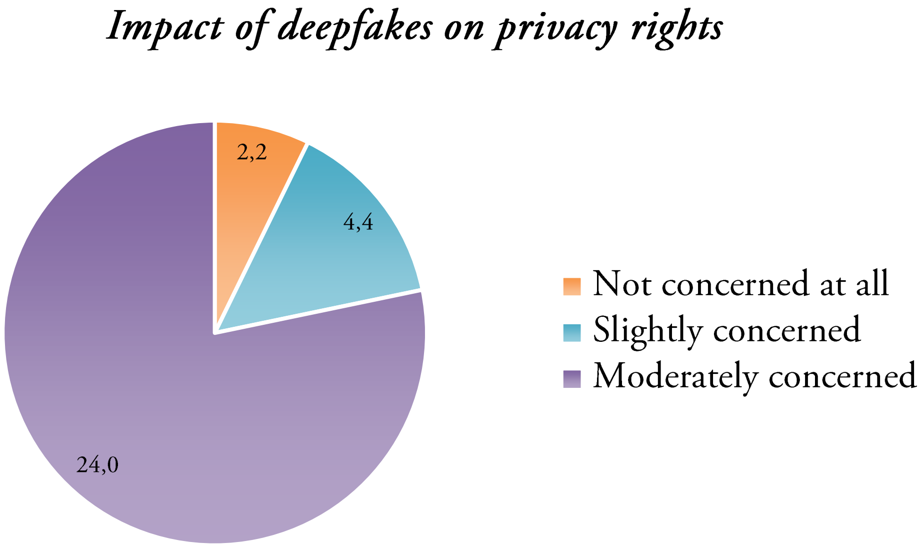

FIGURE 8. Impact of deepfakes I

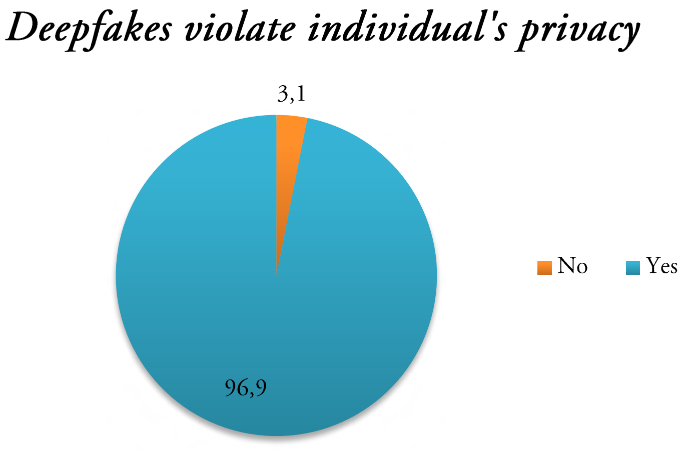

FIGURE 9. Impact of deepfakes II

Vulnerability of specific groups such as women and celebrity to misuse of deepfake

The data highlights strong opinions on the vulnerability of certain groups to deepfakes and the legal status of such content. A substantial majority, 85.4%, agree (47.6%) or strongly agree (37.8%) that specific groups such as women and celebrities are particularly vulnerable to the misuse of deepfakes.

The data reveals a strong consensus regarding the vulnerability of specific groups, such as women and celebrities, to deepfake misuse. A substantial 85.4% of respondents either agree or strongly agree that these groups are particularly susceptible to the harmful effects of deepfake technology. Women and celebrities are often targeted due to the public availability of their images and videos, which makes them prime candidates for non-consensual deepfake content, including explicit videos, revenge porn or defamatory material.

This widespread recognition of the issue supports the argument that deepfakes pose unique privacy risks, especially to individuals with public profiles or those subject to societal and gender-based exploitation. The findings underline the need for legal frameworks to include specific protections for these vulnerable groups, as they face a higher risk of privacy violations, reputational damage, and emotional harm from deepfake misuse. These responses also reflect the growing public awareness of the disproportionate impact deepfakes have on certain demographics, strengthening the case for targeted legal interventions.

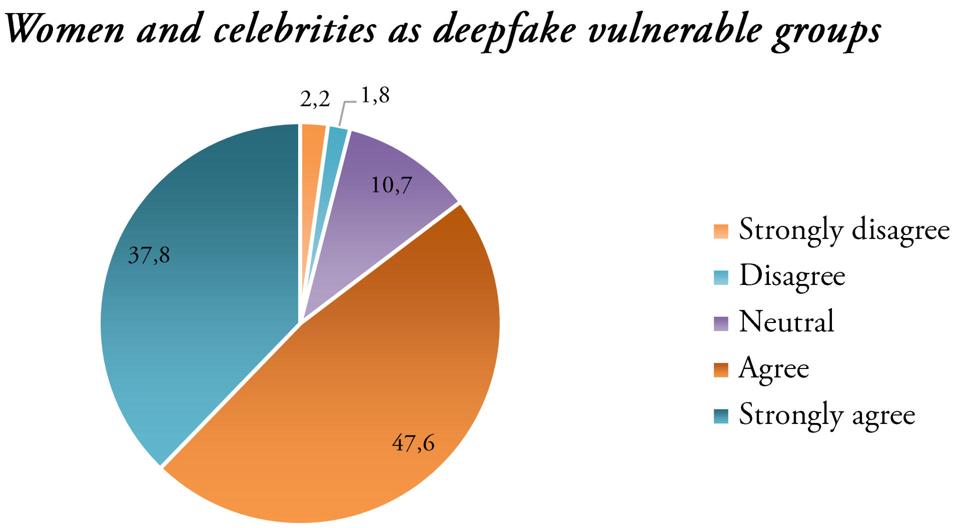

FIGURE 10. Public awareness of deepfakes (women and celebrities)

Opinion regarding blanket ban of deepfakes

The Figure 11 represents the distribution of public opinion on whether making and circulating deepfakes for all purposes should be made illegal. The largest segment, 50.2%, strongly agrees with this proposition, while 36.4% simply agrees. Together, these groups constitute a significant majority. A small percentage, 5.8%, remains neutral on the matter, neither agreeing nor disagreeing. On the other hand, 5.8% of respondents disagree, and a very small fraction, 1.8%, strongly disagrees, indicating that they oppose making deepfakes illegal for all purposes.

The data shows that an overwhelming majority of respondents (86.6%) either strongly agree or agree that deepfakes should be made illegal. This suggests widespread public concern regarding the negative implications of deepfakes, particularly their potential misuse in areas such as defamation, identity theft, and violation of privacy. This support for criminalizing deepfakes likely stems from the rising awareness of their harmful effects, as well as a desire for stricter regulation to protect individuals and society from exploitation and deception.

Meanwhile, the neutral response (5.8%) might reflect uncertainty among a small segment of the population, possibly due to a lack of understanding of the technology or its potential impact. They may be undecided on whether deepfakes should be universally prohibited or believe that regulation should be context-specific.

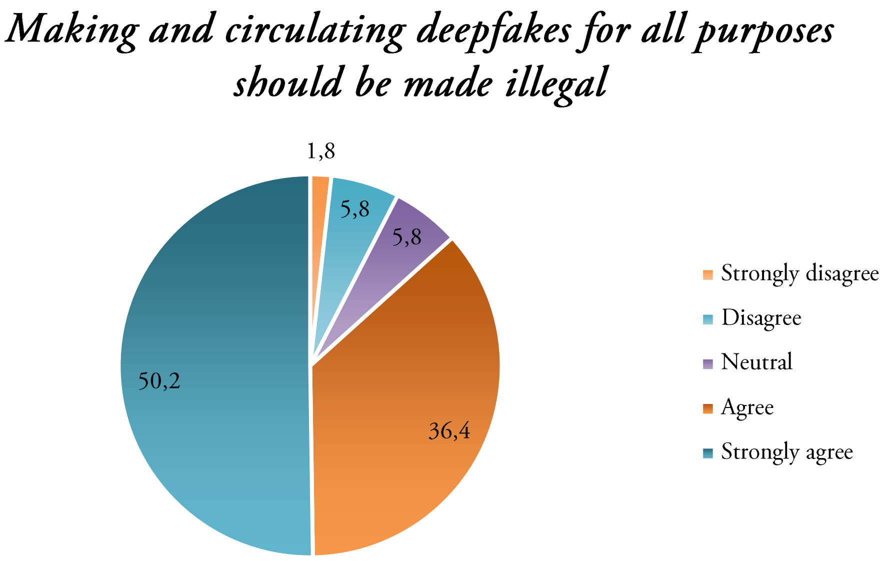

FIGURE 11. Public Opinion about Deepfake I

The 5.8% who disagree and the 1.8% who strongly disagree indicate that some individuals believe in the potential for legitimate or benign uses of deepfake technology, such as in art, entertainment, or education. This group may feel that a blanket prohibition would stifle creativity or technological advancement, or that alternative legal frameworks could address deepfake misuse without imposing a total ban.

Need for an exhaustive legal framework for regulation of deepfakes

The Figure 12 strongly indicates that there is a broad consensus on the necessity of a comprehensive legal framework to regulate deepfakes. With 55.1% of respondents strongly agreeing and 37.3% agreeing, a total of 92.4% of participants support the creation of robust legal measures to address the threats posed by deepfake technology. This overwhelming majority reflects widespread public recognition of the risks deepfakes pose, particularly in terms of privacy violations and misuse.

The 4.0% of respondents who are neutral may indicate a lack of detailed knowledge about deepfakes or uncertainty regarding how such a framework would be implemented. Meanwhile, the small minority who disagree (2.7%) or strongly disagree (0.9%) with the need for regulation could reflect those who believe that deepfake technology has legitimate uses, or that current regulations are sufficient.

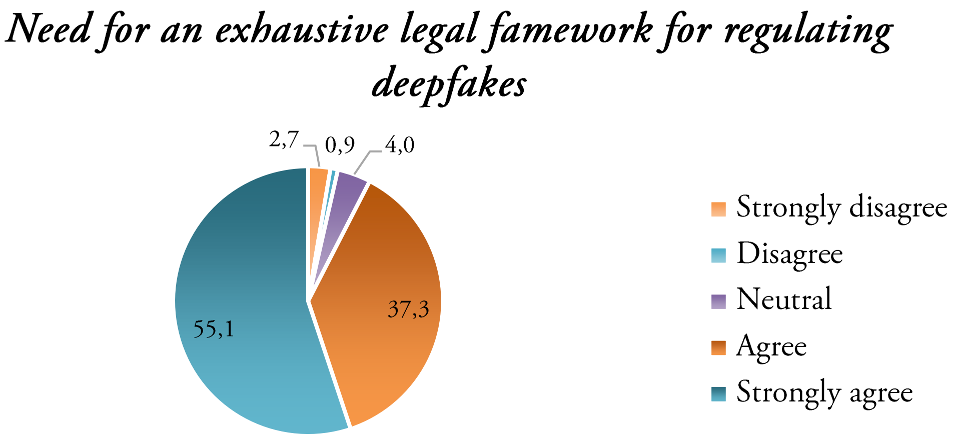

FIGURE 12. Public Opinion about Deepfake II

The empirical data collected in this study strongly validates the argument that deepfakes pose significant threats to privacy rights. The survey findings reveal high levels of awareness of deepfakes, concerns about their misuse, and a consensus that current legal frameworks are inadequate to address these challenges. These insights underscore the need for comprehensive legal reforms that protect individuals' privacy rights in the face of rapidly advancing digital technologies. The findings from this study suggest that future research should aim for more diverse demographic representation, and policymakers must urgently consider comprehensive legal measures to address the complex and evolving threats posed by deepfakes.

Global Legal Landscape

The rapid rise of deepfake technology has created significant challenges for legal frameworks worldwide. Despite efforts to regulate the misuse of deepfakes, existing laws are largely inadequate in addressing the comprehensive privacy threats posed by this technology. This section will explore the legal frameworks in the USA, European Union (EU), and India, analyzing their scope and limitations in combating deepfake misuse, particularly in safeguarding privacy rights.

A. USA Legal Framework

At present, the United States lacks any overarching federal legislation specifically designed to prohibit or regulate the use of deepfakes. Deepfake technology presents complex legal challenges that many existing privacy laws are ill-equipped to address. Traditional privacy laws, such as defamation or false light, often fall short because they only target deception-related harms, not the broader dignitary violations that deepfakes embody. The privacy torts available in the U.S. legal system, such as intrusion upon seclusion or public disclosure of private facts, generally protect true, private information, yet deepfakes blend truth (such as a person's face) with falsity (fabricated scenarios, like pornographic content), creating significant gaps in legal protection.[37]

In response to these limitations, several states have moved to create new regulations specifically addressing deepfakes. However, despite these efforts, the legal framework remains inadequate in fully addressing the wide-ranging implications of deepfakes, especially concerning privacy rights.

California

California has taken a proactive approach to combat deepfake misuse. The state passed two significant laws in 2019, targeting both pornographic deepfakes and election interference.

1. California Assembly Bill 602 allows individuals whose likeness has been used in non-consensual pornographic deepfakes to sue the creators for damages. This law provides a civil remedy to victims of non-consensual pornography, addressing violations of sexual privacy.[38]
2. California Assembly Bill 730 criminalizes the distribution of altered videos or audio meant to mislead voters within 60 days of an election. This law was introduced after a viral deepfake video of Nancy Pelosi, manipulated to appear intoxicated, spread misinformation. The law's primary focus is on combatting misinformation intended to mislead voters.[39]

While California's laws offer some protection, they are limited in scope. The state primarily focuses on pornographic and political deepfakes, leaving broader areas of privacy infringement unregulated.

Texas

Texas criminalized the use of deepfakes in 2019 under SB 751, but its focus is narrow, primarily targeting election interference. The law makes it a criminal offense to create or distribute deepfakes within 30 days of an election with the intent to harm a candidate or influence election outcomes. Violations

[37] Kugler and Pace, "Deepfake Privacy: Attitudes and Regulation", 660.

[38] California Civil Code § 1708.86 (2020), "California Assembly Bill 602 (AB 602) - Depiction of individual using digital or electronic technology: sexually explicit material: cause of action," Last accessed August 13, 2024, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602.

[39] California Election Code § 20010(a) (2020).

can lead to a misdemeanor charge, with penalties including up to one year in jail and fines.[40]

Texas' law only covers election-related deepfakes. It does not address non-political deepfakes, such as those used for personal or financial harm, which leaves many forms of deepfake misuse, particularly those infringing on personal privacy, unregulated.

Virginia

Virginia has also made strides in combating deepfakes by expanding its revenge porn laws. Under Virginia House Bill 2678, the state criminalizes the creation and dissemination of non-consensual deepfake pornography, classifying it as a Class 1 misdemeanor with penalties including up to 12 months in jail and fines.[41]

Similar to California and Texas, Virginia's law addresses specific forms of deepfake misuse (i.e., non-consensual pornography) but does not cover other threats posed by deepfakes, such as their use in identity theft, fraud, or general privacy violations.

The laws in California, Texas, and Virginia provide a starting point in regulating deepfake misuse. However, these laws are narrowly tailored, primarily focusing on pornography and election interference. Deepfakes have far broader applications, and their misuse extends to areas such as Privacy violation, identity theft, harassment, and fraud, none of which are adequately covered by current laws. Current laws fail to adequately protect non-public figures or provide redress for the emotional distress caused by deepfakes, especially when the content is not used for financial gain.

Moreover, these state laws only address criminal or civil penalties and do not offer comprehensive protection for victims of non-consensual deepfakes. The broader privacy implications of deepfakes, particularly their capacity to cause emotional and psychological harm, remain largely unaddressed.

To combat the growing privacy violations associated with deepfakes, there is a need for a federal legal framework that more comprehensively addresses the privacy implications of this technology. Additionally, more stringent platform

[40] Texas Election Code § 255.004 (2019), https://texas.public.law/statutes/tex._election_code_section_255.004.

[41] The Code of Virginia § 18.2-386.2 (2019), https://legiscan.com/VA/text/HB2678/id/1971540#:~:text=Any%20person%20who%2C%20with%20the%20intent%20to%20coerce%2C,image%20is%20guilty%20of%20a%20Class%201%20misdemeanor.

accountability measures are necessary to ensure that digital platforms do not become safe havens for distributing malicious deepfakes.

B. European Union Legal Framework

The General Data Protection Regulation (GDPR) offers several provisions aimed at protecting individuals' privacy and data in the context of deepfakes, but these measures fall short of fully addressing the unique privacy violations that deepfakes can cause.

At the core of GDPR's protection is Article 4(1), which defines personal data as any information related to an identifiable individual.[42] Since deepfakes often involve the manipulation of real personal data, such as images, videos, or voice recordings, they clearly fall under the GDPR's scope when the deepfake portrays a real person. This means that the creation, use, and distribution of deepfakes are subject to the GDPR's stringent requirements for processing personal data. However, a major limitation arises when deepfakes are based on fictional input or the likeness is altered to the point where the individual is no longer easily identifiable. In such cases, the GDPR's protections may not apply, leaving a regulatory gap in protecting fictionalized or generated content that can still be harmful to an individual's privacy or reputation.[43]

Article 6 of the GDPR requires a legal basis for processing personal data, such as consent or legitimate interest.[44] In the context of deepfakes, obtaining explicit consent from the subject is ideal, but in practice, this is difficult to achieve, especially in cases where individuals are unaware that their likeness has been used, as often occurs in non-consensual deepfake pornography or identity theft. Without consent, deepfake creators might attempt to rely on legitimate interest, but this requires a delicate balance between the benefits of using the data and the potential harm to the individual. In cases where deepfakes cause

[42] General Data Protection Regulation, Article 4 (1); "*personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".*

[43] Felipe R. Moreno, "Generative AI and deepfakes: a human rights approach to tackling harmful content," *International Review of Law* (March 29, 2024): 1-30, https://doi.org/10.1080/13600869.2024.2324540; See also European Parliament, "Tackling deepfakes in European policy," July, 2021, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.

[44] General Data Protection Regulation, Article 6 (1) (f).

reputational damage, emotional distress, or privacy violations, legitimate interest would likely not justify the use of personal data, especially given the GDPR's emphasis on protecting individuals' rights and freedoms. However, deepfakes used for artistic or satirical purposes, protected under Articles 11 and 13 of the EU Charter of Fundamental Rights, may still qualify as legitimate, creating grey areas in the law.

The GDPR further strengthens privacy protections through Article 9, which applies stricter conditions for processing sensitive data such as political views, sexual orientation, or health information.[45] If a deepfake involves the misuse of such sensitive data, it would trigger heightened privacy protections, making it difficult for creators to justify their use without explicit consent. This provides an additional layer of protection for individuals whose sensitive data might be exploited in deepfakes. However, the lack of clarity in distinguishing whether deepfakes inherently reveal sensitive data makes the enforcement of Article 9 challenging, potentially allowing harmful deepfakes to slip through legal loopholes.

Article 22 of the GDPR offers protection against automated decision-making[46], which is particularly relevant for individuals impacted by malicious deepfakes used in blackmail, financial scams, or election misinformation. Victims of deepfakes can object to decisions made solely on the basis of AI-generated content. While this provision helps protect against automated discrimination, it is insufficient to combat the broader privacy harms that deepfakes can cause. For example, once a deepfake has been widely disseminated, it is extremely difficult to undo the damage to a person's reputation, even if the content is later proven to be false.

Moreover, the right to be forgotten (Article 17) and right to rectification (Article 16) offer individuals the ability to demand the removal of their personal data from digital platforms.[47] However, in the context of deepfakes, where the content has often gone viral and is replicated across multiple platforms, enforcing these rights becomes complex and often ineffective. Even if the original deepfake is removed, copies can persist, continuing to infringe on the individual's privacy.

While the GDPR provides several important tools for protecting privacy, these provisions are not entirely sufficient to combat the privacy violations caused by deepfakes. The ambiguous nature of personal data in fictionalized deepfakes, the difficulty of obtaining consent, and the challenges in removing

[45] General Data Protection Regulation, Article 9 (1).

[46] General Data Protection Regulation, Article 22 (1).

[47] General Data Protection Regulation, Article 16 & 17.

viral deepfakes leave significant gaps in privacy protection. More targeted regulations, as well as technical innovations, are needed to fully safeguard individuals from the misuse of deepfakes.

C. Artificial Intelligence Act, 2023

The EU Artificial Intelligence Act (2023)[48] introduces several provisions aimed at regulating AI-generated content, including deepfakes, with a focus on transparency and user protection. However, these provisions fall short when it comes to fully addressing the threats to privacy posed by this evolving technology. Article 50(2) requires AI providers to clearly label and tag AI-generated content, such as deepfakes, using methods like watermarks, metadata, or security features.[49] This aims to ensure that the public can easily identify when they are interacting with AI-generated content. However, the provision does not apply to minor edits or instances where AI is used by law enforcement for crime detection and prosecution. This creates gaps in regulation, allowing certain uses of deepfakes to go unregulated.

Article 50(4) mandates that creators of deepfakes must inform the public about the artificial nature of their work, ensuring that the content is properly labeled to indicate its origin.[50] While this enhances transparency, there are key exemptions for law enforcement activities and artistic or creative deepfakes (such as satire or fictional works). These exemptions, particularly for creative expression, raise concerns about potential loopholes, where harmful content could be disguised as art or satire, undermining the law's intent to protect privacy and prevent disinformation.

In addition, Recital 136 of the Act imposes obligations on platforms and search engines, particularly large ones, to detect and disclose manipulated content such as deepfakes.[51] This is designed to address risks like disinformation and election interference. However, these transparency requirements do not replace existing obligations under the Digital Services Act (DSA), leading to potential overlap and fragmentation in enforcement, which may reduce the overall effectiveness of the measures.

The AI Act also promotes the development of voluntary codes of practice and guidelines for deepfake detection and labeling under Article 50(7) and

[48] European Union Artificial Intelligence Act, 2023, https://artificialintelligenceact.eu/ai-act-explorer/.

[49] European Union Artificial Intelligence Act, 2023, Article 50(2) and Recital 133.

[50] European Union Artificial Intelligence Act, 2023, Article 50(4) and Recital 134.

[51] European Union Artificial Intelligence Act, 2023, Recital 136.

Recital 135.[52] These codes encourage collaboration among stakeholders to improve compliance. However, the voluntary nature of these codes means they lack the legal binding force necessary to ensure consistent enforcement across platforms and AI providers, which could weaken the overall regulatory framework.

The failure to classify deepfakes as high-risk AI under Article 5[53] poses a major limitation in protecting privacy rights. Deepfakes, particularly those used for extortion, identity theft, or AI-generated child sexual abuse material, represent severe privacy invasions, yet they are not given the same level of regulatory scrutiny as other high-risk AI systems. This omission creates a patchwork of protections across the EU, leaving many victims of deepfakes without adequate legal recourse or protection against the misuse of their personal data and likeness.

In sum, while the AI Act (2023) introduces meaningful transparency and labeling requirements, its exemptions, voluntary measures, and the failure to classify deepfakes as high-risk make it insufficient to comprehensively combat the privacy threats posed by this powerful technology. More stringent regulations and enforcement mechanisms are needed to adequately protect individuals' personal autonomy, dignity, and privacy in the face of growing deepfake misuse.

D. Indian Legal Framework

In India, there is no dedicated legislation to tackle implications of deepfake. However, by invoking the provisions of few legislations and regulatory framework, the relief can be sought indirectly in cases of misuse of deepfake.

Information Technology Act, 2000 (IT Act)

In the context of deepfakes, the Information Technology (IT) Act, 2000 provides several provisions that can be applied to address the misuse of this technology, particularly with regard to identity theft, impersonation, and the publication of obscene material. While the IT Act does not explicitly mention deepfakes, the following sections can be interpreted to provide some level of protection against their misuse:

1. Section 66A: This section penalized the sending of offensive messages through electronic means, including images, videos, and audio.[54]

[52] European Union Artificial Intelligence Act, 2023, Article 50 (7) and Recital 135.

[53] European Union Artificial Intelligence Act, 2023, Article 5.

[54] Information Technology Act, 2000, Section 66A, https://www.indiacode.nic.in/bitstream/123456789/13116/1/it_act_2000_updated.pdf.

Although the section has been struck down, it is worth noting that deepfakes, which can include grossly offensive content or misleading information, would have fallen under its ambit. Deepfakes used to cause annoyance, inconvenience, or mislead recipients about their origin could have been addressed through this provision. The absence of an equivalent law leaves a gap in protecting individuals from offensive or misleading deepfake content.

2. Section 66B: This section deals with the dishonest receipt of stolen computer resources or communication devices.[55] While not directly related to deepfakes, it could potentially apply if a deepfake was created using stolen digital content or devices. For instance, if someone uses stolen images or videos to create a deepfake, they could be penalized under this section.
3. Section 66C: This section deals with identity theft, criminalizing the fraudulent use of someone's digital identity, which can be a component of deepfake-related crimes.[56]
4. Section 66D: This section criminalizes cheating by impersonation using a computer resource.[57] Deepfakes, which often involve the manipulation of an individual's likeness to impersonate them, could be addressed under this section. For example, deepfakes used to impersonate someone for the purpose of fraud or to cause harm would be punishable with imprisonment of up to three years and a fine. This provision is critical in combating identity theft and impersonation through deepfakes.
5. Section 67: This section penalizes the publication or transmission of obscene material in electronic form. Deepfakes, particularly non-consensual pornographic content, would fall squarely under this section.[58] If a deepfake contains obscene or lascivious content, it could be punishable with imprisonment of up to three years for the first conviction and a fine of up to five lakh rupees. This section serves as an important tool in protecting individuals, especially women and vulnerable groups, from the harassment and emotional distress caused by obscene deepfakes.

[55] Information Technology Act, 2000, Section 66B.

[56] Information Technology Act, 2000, Section 66C "Punishment for identity theft".

[57] Information Technology Act, 2000, Section 66D.

[58] Information Technology Act, 2000, Section 67 "Punishment for publishing or transmitting obscene material in electronic form".

6. Section 67A: This section addresses the publication or transmission of material containing sexually explicit acts in electronic form.[59] Deepfakes that involve sexually explicit content such as revenge porn or non-consensual pornographic deepfakes, are directly punishable under this section. Offenders face imprisonment of up to five years for the first offense, and up to seven years for subsequent convictions, along with heavy fines. This provision is essential in protecting individuals from the severe privacy violations associated with sexually explicit deepfakes.

Despite these provisions, the IT Act does not explicitly mention deepfake technology or AI-generated content, leaving significant gaps in addressing the nuanced challenges posed by deepfakes. The law primarily addresses traditional cybercrimes, but deepfakes introduce more complex issues related to privacy, reputation, and identity, which require targeted legislation to effectively combat their misuse. Thus, while the existing framework provides some protections, it remains insufficient in the face of the growing threat posed by deepfakes.

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (IT Rules, 2021)

The IT Rules, 2021[60] establish a framework for intermediaries (like social media platforms) to manage and mitigate digital content issues, including deepfakes:

1. Rapid Removal: Intermediaries are required to expeditiously remove offensive content (e.g., deepfakes) within stipulated timelines upon receiving notifications from courts or government agencies.
2. Rule 3(1)(d): Mandates intermediaries to take swift action to remove or disable access to harmful content, which includes deepfakes.
3. Cooperation with Law Enforcement: Under Rules 3(1)(j) and 4(2), intermediaries must cooperate with law enforcement in cases affecting national security, public order, or crimes like sexual exploitation where deepfakes often play a role.

In addition, the government has established Grievance Appellate Committees[61] to handle complaints against intermediary decisions, and the

[59] Information Technology Act, 2000, Section 67A "Punishment for publishing or transmitting of material containing sexually explicit act, etc., in electronic form".

[60] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, https://www.meity.gov.in/writereaddata/files/Revised-IT-Rules-2021-proposed-amended.pdf.

[61] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rule 3A.

National Cyber Crime Reporting Portal[62] facilitates the reporting of cybercrimes, including those involving deepfakes.

The Bharatiya Nyaya Sanhita, 2023

The Bharatiya Nyaya Sanhita (BNS), 2023,[63] which replaces the Indian Penal Code (IPC) of 1860, offers more modernized provisions to address challenges brought about by digital technologies, including deepfakes. Under this new legislation, defamation laws have been updated. Section 356(2) of BNS now governs defamation, which was previously covered under Sections 499 and 500 of the IPC. The updated law reflects a modern approach that is likely more attuned to the digital dissemination of defamatory content, such as deepfakes.

The Bharatiya Nyaya Sanhita provides a more comprehensive framework to address digital defamation, but its focus remains limited when it comes to deepfake-specific misuse like non-consensual pornography, identity fraud, or political disinformation. These broader concerns are not explicitly covered under the new law, indicating a gap in comprehensive protections against deepfake violations.

While the IT Act, IT Rules, 2021 and Bharatiya Nyaya Sanhita provide mechanisms to address some cybercrimes, the explicit absence of AI and deepfake-specific legislation leaves significant gaps in protecting against privacy violations. The existing laws mainly focus on misinformation, defamation, and identity theft, but deepfakes introduce more nuanced risks, including non-consensual use of likeness, political disinformation, and exploitation through manipulated content.

To effectively combat the privacy threats posed by deepfakes, India needs more targeted legislation that addresses the complex nature of AI-generated content, ensuring that individuals' privacy rights are adequately protected against the evolving challenges of deepfake technology.

Conclusion

Deepfakes represent a significant and growing threat to privacy, posing challenges that are unprecedented in their scope and potential for harm. As discussed, deepfakes can be weaponized for disinformation, defamation, and identity theft, making them a throttling attack on personal privacy in the digital

[62] National Cyber Crime Reporting Portal, https://cybercrime.gov.in/.

[63] The Bharatiya Nyaya Sanhita, 2023, https://prsindia.org/files/bills_acts/bills_parliament/2023/Bharatiya_Nyaya_Sanhita,_2023.pdf.

realm. This technology undermines trust in digital media and creates a fertile ground for manipulating public perception, causing harm to individuals and society at large.

The findings from this study underscore the serious implications of deepfake technology, particularly in terms of privacy violations and misuse. The empirical data reveals a high level of awareness about the risks posed by deepfakes, with respondents acknowledging potential harm in various forms, including misinformation, political manipulation, identity theft, and revenge porn. This broad awareness is reflected in the overwhelming support—92.4% of respondents—for the establishment of a comprehensive legal framework to regulate deepfakes. While there is clear concern about the misuse of deepfakes, the survey also highlights gaps in understanding and inconsistent ability to identify deepfakes, pointing to a need for public education and better detection tools. The data suggests that certain groups, particularly women and celebrities, are more vulnerable to deepfake misuse, emphasizing the need for targeted legal protections to safeguard these demographics.

The current regulatory frameworks, both globally and domestically, are inadequate in addressing the multifaceted threats posed by deepfakes. India's Information Technology Act, 2000, and the Bharatiya Nyaya Sanhita, 2023, provide some protection against deepfake-related privacy violations but fall short in addressing the unique challenges posed by these technologies. Provisions dealing with cybercrimes, identity theft, and defamation offer some recourse, but they are insufficient in the face of deepfake's growing sophistication. Globally, the European Union's GDPR and the fragmented state laws in the US show varying degrees of readiness to tackle deepfakes, but no comprehensive global or domestic framework yet exists to regulate their misuse effectively especially in the context of privacy violations.

References

"2023 State of Deepfakes: Realities, Threats, and Impact." Security Hero. September 10, 2024, https://www.securityhero.io/state-of-deepfakes/.

"Deepfake audio of Sir Keir Starmer released on first day of Labour conference." Sky News. October 9, 2023. https://news.sky.com/story/labour-faces-political-attack-after-deepfake-audio-is-posted-of-sir-keir-starmer-12980181.

"Generative Adversarial Network (GAN)." Geeks for Geeks. Last updated August 9, 2024. https://www.geeksforgeeks.org/generative-adversarial-network-gan/.

"Incident 769: Investigative Journalist Rana Ayyub Targeted by AI-Generated Deepfake Pornography." AI Incident Database. August 10, 2024. https://incidentdatabase.ai/cite/769/.

Ajder et al. "The State of Deepfakes: Landscape, Threats, and Impact." Deeptrace. September, 2019. https://finaletheorie.org/download/the-state-of-deepfakes-an-overview/?wpdmdl=30224&refresh=66dfda870b9f11725946503.

California Civil Code § 1708.86 (2020), "California Assembly Bill 602 (AB 602) - Depiction of individual using digital or electronic technology: sexually explicit material: cause of action," Last accessed August 13, 2024, https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602.

California Election Code § 20010(a) (2020).

California Election Code § 20010(a) (2020).

Chesney and Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." California Law Review 107, no. 6 (2019): 1753-1820. https://www.jstor.org/stable/26891938.

Citron. "Sexual Privacy." The Yale Law Journal 128, no. 7 (May 2019): 1870-1960. https://www.jstor.org/stable/45098020.

Conger and Yoon. "Explicit Deepfake Images of Taylor Swift Elude Safeguards and Swamp Social Media." The NewYork Times. January 26, 2024. https://www.nytimes.com/2024/01/26/arts/music/taylor-swift-ai-fake-images.html#:~:text=One%20image%20shared%20by%20a%20user%20on%20X,spread%20despite%20those%20companies%E2%80%99%20efforts%20to%20remove%20them.

European Convention on Human Rights, https://www.echr.coe.int/documents/d/echr/Convention_ENG

European Parliament, "Tackling deepfakes in European policy," July, 2021, https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf.

European Union Artificial Intelligence Act, 2023, https://artificialintelligenceact.eu/ai-act-explorer/.

General Data Protection Regulation, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

Griswold v. Connecticut, 381 U.S. 479 (1965).

Harwell. "Fake-porn videos are being weaponized to harass and humiliate women: Everybody is a potential target." The Washington Post. December 30, 2018.

https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/.

Information Technology Act, 2000, https://www.indiacode.nic.in/bitstream/123456789/13116/1/it_act_2000_updated.pdf.

International Covenant on Civil and Political Rights, https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

K. S. Puttaswamy v. Union of India, 2019 (1) SCC 1

Kugler and Pace. "Deepfake Privacy: Attitudes and Regulation." Northwestern University Law Review 116, no. 3 (2021): 611-680.

Mcgill. "24 Deepfake Statistics – Current Trends, Growth, and Popularity (December 2023)." Contentdetector.AI. May 23, 2024. https://contentdetector.ai/articles/deepfake-statistics/.

Meredith Somers. "Deepfakes, Explained." MIT SLOAN Schoo. July 21, 2020. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

Moreham. "Privacy in Public Spaces." The Cambridge Law Journal 65, no. 3 (November 23, 2006): 606-635, https://doi.org/10.1017/S0008197306007240.

Moreno. "Generative AI and deepfakes: a human rights approach to tackling harmful content." International Review of Law (March 29, 2024): 1-30. https://doi.org/10.1080/13600869.2024.2324540.

Petrosyan. "Number of internet and social media users worldwide as of July 2024." Statista. August 19, 2024. https://www.statista.com/statistics/617136/digital-population-worldwide/.

Roe. V Wade, 410 U.S. 113 (1973).

Solove and Rotenberg. Information Privacy Law. New York: Aspen Publishers, 2006).

Spring. "Sadiq Khan says fake AI audio of him nearly led to serious disorder." BBC News. February 14, 2024. https://www.bbc.com/news/uk-68146053.

Texas Election Code § 255.004 (2019), https://texas.public.law/statutes/tex._election_code_section_255.004.

The Bharatiya Nyaya Sanhita, 2023, https://prsindia.org/files/bills_acts/bills_parliament/2023/Bharatiya_Nyaya_Sanhita,_2023.pdf

The Code of Virginia § 18.2-386.2 (2019), https://legiscan.com/VA/text/HB2678/id/1971540#:~:text=Any%20person%20who%2C%20with%20the%20intent%20to%20coerce%2C,image%20is%20guilty%20of%20a%20Class%201%20misdemeanor

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, https://www.meity.gov.in/writereaddata/files/Revised-IT-Rules-2021-proposed-amended.pdf.

Universal Declaration of Human Rights, https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Article%2012%20No%20one%20shall%20be%20subjected%20to,of%20the%20law%20against%20such%20interference%20or%20attacks.

Wakefield. "Deepfake presidents used in Russia-Ukraine war." BBC News. March 18, 2022. https://www.bbc.com/news/technology-60780142.

Westerlund. "The emergence of Deepfake Technology: A Review." Technology Innovation Management Review 9, no. 11 (November 2019): 40, https://doi.org/10.22215/timreview/1282.

Westin, Alan. Privacy and Freedom. New York: Atheneum, 1967. https://archive.org/details/privacyfreedom00west/page/n22/mode/1up .

***

Acknowledgment

The earlier version of this paper has been presented on International Conference on Law, Conservation, and Sustainability (ICCONSIST) held by the Faculty of Law, Universitas Negeri Semarang, Indonesia 2024.

Funding Information

None

Conflicting Interest Statement

There is no conflict of interest in the publication of this article.

Publishing Ethical and Originality Statement

All authors declared that this work is original and has never been published in any form and in any media, nor is it under consideration for publication in any journal, and all sources cited in this work refer to the basic standards of scientific citation.

Notification

Starting from the 2024 issue, our journal has transitioned to a new platform for an enhanced reading experience. All new articles and content will now be available on this updated site. However, we would like to assure you that archived issues from 2017 to 2023 are still accessible via the previous site. Please check the following link: https://journal.unnes.ac.id/sju/lslr/issue/archive

Appendix

TABLE 1. The frequency distribution across different variables of the research

Variables	Frequency	Percentage
Age		
Under 18 years	9	4.0
18-27 years	104	46.2
28-37 years	80	35.6
38 years and above	32	14.2
Occupation		
Student	113	50.2
Working professional	91	40.4
Home maker	8	3.6
Other	13	5.8
Gender		
Male	125	55.6
Female	100	44.4
Active social media user		
No	29	12.9
Yes	196	87.1
Awareness about deepfakes		
I have never heard of it	8	3.6
I have heard of it but don't know much about it	43	19.1
I have read about it and have a basic understanding	125	55.6
I have good knowledge and awareness of it	49	21.8
People who came across any incident of deepfake on social media		
No	80	35.6
Yes	145	64.4
Ability to identify deepfake image/audio/video		
Yes, I can easily identify them	40	17.8
Yes, but I am not always certain	111	49.3
No, I find it difficult to identify them	44	19.6
No, I have never tried to identify them	30	13.3

Variables	Frequency	Percentage
People who made a deepfake image/audio/video		
No	220	97.8
Yes	5	2.2
Victim of misuse of deepfake		
No	210	93.3
Yes	15	6.7
Known victim of deepfake		
Family member	5	2.3
Close friend	20	9.0
Relative	6	2.7
Acquaintance	21	9.5
None	170	76.6
Awareness of misuse of deepfake		
No	18	8.0
Yes	207	92.0
Ways of misusing deepfake:		
1.Creating fake news and misinformation		
No	189	90.9
Yes	19	9.1
2.Political manipulation		
No	200	96.6
Yes	7	3.4
3.Identity theft and fraud		
No	196	94.2
Yes	12	5.8
4.Defamation		
No	195	94.2
Yes	12	5.8
5.Revenge porn and explicit content creation		
No	191	91.8
Yes	17	8.2
6.Violation of personality rights		
No	199	95.7

Variables	Frequency	Percentage
Yes	9	4.3
7.Combination of above		
No	25	11.8
Yes	187	88.2
Impact of deepfakes on privacy rights		
Not concerned at all	5	2.2
Slightly concerned	10	4.4
Moderately concerned	54	24.0
Very concerned	156	69.3
Deepfakes violate individual's privacy		
No	7	3.1
Yes	218	96.9
Certain groups (e.g., women, celebrities) are more vulnerable to deepfake		
Strongly disagree	5	2.2
Disagree	4	1.8
Neutral	24	10.7
Agree	107	47.6
Strongly agree	85	37.8
Making and circulating deepfakes for all purposes should be made illegal		
Strongly disagree	4	1.8
Disagree	13	5.8
Neutral	13	5.8
Agree	82	36.4
Strongly agree	113	50.2
Need for an exhaustive legal framework for regulating deepfakes		
Strongly disagree	6	2.7
Disagree	2	0.9
Neutral	9	4.0
Agree	84	37.3
Strongly agree	124	55.1