The Maastricht Journal of Liberal Arts
2024, Volume Fifteen 15-25

THE CURRENT STATE OF PORNOGRAPHIC DEEPFAKES

A Science and Technology Studies Perspective

Author: Jola Gockel

Abstract This paper analyses the current state of deepfake pornography from a Science and Technology Studies (STS) viewpoint. Looking at the phenomenon from a social constructivist perspective shows that misogynistic power structures are embedded in certain deepfake technologies and that deepfake pornography reflects and reinforces such power structures. Additionally, the risk perspective points to the need for effective (federal and global) legislation and to the need for increased public awareness. Finally, the vulnerability perspective reveals how not everyone is affected equally by the potential of being featured in deepfake pornography, with celebrities having a higher risk of being featured in deepfakes and private individuals experiencing greater difficulty disproving deepfakes of themselves. Implications and questions for future research are discussed.

Keywords: Deepfakes, Deepfake Pornography, Artificial Intelligence, Science and Technology Studies

1. Introduction

Picture this: you are scrolling through your social media feed as usual when you suddenly see a video of yourself. It is not the usual kind of video depicting you with your family or friends but one of you engaging in sexual activities. The body in the video is not yours, but the face is. This was the reality faced by Indian journalist Rana Ayyub in 2018 when a video of her allegedly starring in a pornographic video was shared on Facebook, Twitter and WhatsApp. Although it came to light that the video did not actually depict Ayyub, it undermined her work as a journalist and made her the victim of online harassment, including rape threats (Kerner & Risse, 2021; Reid, 2021). Ayyub fell victim to a technology known as deepfakes. The term "deepfake" stems from the terms "deep learning" and "fake". A deepfake can be any digital product–videos, images, audio, text–that has been artificially created through machine deep learning technology (Giles et al., 2019). In the case of deepfake videos, the deep learning programme analyses dozens of videos or images of a person's face from various angles, to learn what it looks like and then superimposes it onto somebody else's body (Kerner & Risse, 2021; Öhman, 2020).

Meskys et al. (2020) identified four categories of deepfakes: pornographic, political, commercial and creative. While deepfakes can be used with harmful intent in any of these categories, the authors highlight the potentially damaging consequences of the first use case. Aijder et al. (2019) found that 96% of deepfake videos online are pornographic. This is an alarming development for several reasons. Recall the example of Rana Ayyub, who endured physical, mental and emotional harm from the fake video, describing it as cruel and invasive (Reid, 2021). Pornographic deepfakes can undermine a person's chances of finding a job or a romantic partner (Chesney & Citron, 2019). Furthermore, deepfake pornography can be intentionally used to humiliate, harass or blackmail (Franks & Waldman, 2019; Hancock & Bailenson, 2021).

As the prevalence of deepfake videos, including pornographic deepfakes, is expected to rise in coming years and the technology is expected to produce increasingly realistic outputs (Harris, 2021), it is important to examine this phenomenon from different disciplines and angles. To contribute to the growing scholarship on the subject and address societal needs, this paper poses the following research question: *How can concepts and theories from Science and Technology Studies (STS) contribute to the understanding of the current situation regarding deepfake technology in pornography?* This paper first gives additional background information on deepfake technology and pornographic deepfakes. It then looks at the phenomenon from a social constructivism viewpoint, and applies the risk- and vulnerability perspectives. Finally, this paper concludes that the interdisciplinary nature of STS points towards the misogynistic power structures underlying deepfake pornography, as well as a lack of legislative solutions for and public awareness of the impacts of deepfake technology. STS thus demonstrates that, in addition to technological approaches, there is also a need for legal and societal solution approaches to deepfake pornography, including legislation, education, and a broader societal effort against misogyny.

2. The Rise of Deepfake Technology

Deepfake technology first emerged as a public phenomenon in 2017, when a Reddit user began posting deepfake videos depicting female celebrities in pornography (Delfino, 2019). In 2018, another Reddit user created and posted FakeApp, an application that allowed anyone to create their own deepfakes, provided they had several hundred images of their target (Meskys et al., 2020). A month after the release of FakeApp, Reddit banned the deepfake subreddit, but other dedicated forums quickly emerged (Aijder et al., 2019). The first dedicated deepfake pornography website was registered in early 2018 and within a year and a half, the top four websites of this kind had gathered over 134 million views (Karasavva & Noorbhai, 2021). As Aijder et al. (2019) identified, there are currently three ways in which individuals can create (pornographic) deepfakes of their choice. First, there are computer applications and accompanying tutorials which enable users to create their own deepfakes. Second, there are service portals where anyone can upload sets of pictures and purchase deepfakes at little cost. Third, there are individual deepfake creators who use online forums and marketplaces to advertise their paid services. These different options mean that anyone, regardless of their skill level, can create deepfakes. As Delfino (2019) highlights, the implication is that “everyone, everywhere is a potential deepfake victim” (p.887). At present, the programmes developed to detect deepfakes are lagging well behind those made to create them. Any weakness used to detect deepfakes, such as irregular blinking patterns, is quickly picked up by deepfake technology creators and incorporated into the fake videos (Gosse & Burkell, 2020).

3. A Social Constructivist View on Deepfake Pornography

To understand the social context around the development and usage of deepfake technology, this paper now first turns to social constructivism, which is a central paradigm within which Science and Technology Studies operates. Social constructivism points to the political nature of the development and usage of any technology (Bijker, 2001). Political, in this case, aligns with Winner's (1980) definition of "arrangements of power and authority in human associations as well as the activities that take place within those arrangements" (p. 123). Although Winner is not a constructivist himself, his theory of the politics of artefacts aligns with social constructivism's examination of power. According to his theory, politics, i.e. power structures, are built into technologies in the way they are designed (Winner, 1980). This can seem abstract at first, but a case study of DeepNude can clarify Winner's ideas. DeepNude is an application that allows its users to upload a full-body photo of another person. Within less than a minute, the uploaded picture is returned, depicting the subject completely naked. However, DeepNude only works on pictures of women. This is because the data that was used to train the algorithm underlying the application was solely of female bodies (Aijder et al., 2019). The creators of DeepNude designed the application in such a way that it grants users authority over female bodies only, thereby manifesting misogynistic power structures in its design.

Importantly, social constructivism sees the power structures reflected in technology as an active choice that was made (Kleinmann, 2005, Chapter 1). Certain power structures—in this case misogyny—are built into the technology while others are not. But this design is not considered as predestined; it could have been otherwise (Pinch & Bijker, 1984). It is not technologically impossible for an application like DeepNude to work on pictures of men, the creators simply chose not to programme it like that. According to social constructivism, social context plays the crucial role in determining which technological designs are successful (Bijker, 2001). The design of DeepNude thus makes sense, considering the social context in which objectification of females is well-established (Harris, 2021) and female deepfake pornography has found great popularity. For example, before it was shut down, a subreddit dedicated to pornographic deepfakes was followed by more than 100,000 users. Apart from such videos, posts in this subreddit also included people asking how to create pornographic deepfakes of their ex-girlfriend or of a girl they go to school with (Chesney & Citron, 2019). It is open to question whether deep fake pornography featuring men's bodies would hold the same appeal. Long before deepfake pornography existed, the same structural inequalities were already visible in traditional pornography and men were accused of objectifying women for their own pleasure (Harris, 2021).

To further illustrate the political nature of deepfake technology, even when it is less inherently political than DeepNude, it is helpful to consider Woolgar and Cooper's (1999) notion of the ambivalence of artefacts: "technology is good and bad; it is enabling and it is oppressive; it works and it does not; and, as just part of all this, it does and does not have politics" (p.443).

While other deepfake pornography applications may seem less inherently political, everything surrounding their use is nevertheless shaped by power structures. Consider the following: Who is developing deepfake technologies? Who is using deepfake technology to create pornography? Who is depicted in this pornography? The answers are simple: it is overwhelmingly men who, without consent, insert images of women into sexual scenarios in which they never took part (Aijder et al., 2019; Öhman, 2020). As researchers have pointed out, this is highly problematic, as it takes away women's autonomy over their bodies, depicting them as sex objects for men to enjoy (Chesney & Citron, 2019). From a social constructivist perspective, this also exemplifies the co-shaping that takes place between society and technology (Bijker, 2001). Social values are thought to shape how technology is developed—these values are embodied in and reinforced by the technology itself (Pinch & Bijker, 1984; Wyatt, 1998, Chapter 1). In this case, the technology reflects the misogyny that exists within society, e.g., in how the DeepNude application is programmed. Once this discrimination has been built into the technology, its merit and consequences might no longer be questioned but taken for granted (Kerner & Risse, 2021). Furthermore, the technology may now reinforce the societal structures that were built into it (Chesney & Citron, 2019; Pinch & Bijker, 1984; Wyatt, 1998, Chapter 1), potentially encouraging the development of other similar misogynistic applications.

Overall, these aspects, as highlighted by social constructivism, show that simply trying to develop technological solutions to deepfake pornography will not suffice. The underlying structural societal problems are too inherently intertwined with technological development and would not therefore be solved by developing a perfect deepfake detection software.

4. A Risk and Vulnerability Perspective on Deepfake Pornography

The previous section applied concepts from the general STS perspective of social constructivism to the case of deepfake pornography to offer a clearer picture of the social context surrounding it. To deepen the analysis, this paper now turns to a more specific part of STS: the risk and vulnerability frameworks (Bijker et al., 2014).

4.1 The Risk Perspective

One part of the risk perspective consists of theories of risk governance, as described by Bijker et al. (2014). Risk governance refers to “the various ways in which many actors, individuals, and institutions, public and private, deal with risks surrounded by uncertainty, complexity, and/or ambiguity” (Van Asselt & Renn, 2011, p. 432).

A first consideration is how institutions are dealing with the rise of deepfake technology. The law commonly lags behind technological development (Eggestein & Knapp, 2014). Deepfake technology—and specifically the rise of pornographic deepfakes—is no exception. As of 2023, there is no federal law in the United States which provides remedy to the victims of deepfake pornography (Beaumont-Thomas, 2024). Instead, victims usually have to rely on other legislation, such as copyright claims or suing for defamation (Chesney & Citron, 2019). While this is sometimes successful, often it is not, as these laws were not formulated with deepfake pornography in mind (Harris, 2019). Although there is no federal legislation addressing deepfake technology, some individual US states have passed laws on it. For example, California has passed a bill that allows victims to pursue legal action where they did not consent to be featured in a deepfake pornographic video (Kerner & Risse, 2021). However, this reveals an important limitation in risk governance: legislation is bound by state or country borders, whereas media on the internet is not. If a victim lives in California but the perpetrator does not, the victim cannot make use of the Californian legislation. For this reason, scholars are calling for federal US laws concerning deepfake pornography (Delfino, 2019; Harris, 2019). From a global perspective the challenge is even greater, since more legislative borders are concerned. While the UK, for example, introduced legislation making deepfake pornography illegal in 2023 (Yousif, 2024), other countries, such as Canada or various members of the European Union, do not have appropriate legislation (Karasavva & Noorbhai, 2021; Mania, 2024). This points to a dire need for countries to cooperate globally in establishing legislation that offers remedies to deepfake pornography victims, who are often left powerless. On a more positive note, social media platforms have been faster in adapting to this technological development than governments. Several companies, such as Pornhub, Reddit, Discord and Twitter, have expanded their terms of service and banned deepfake content from their platforms (Ratner, 2021).

Beyond the institutional response, it is worth examining how the general public is dealing with the risks of deepfake technology in pornography. In general, little attention is being paid to this issue, since large parts of online communities do not seem to care about the spread of deepfake pornography (Meskys et al., 2020). In mainstream media, the risks of political deepfakes have been discussed significantly more than those of pornographic deepfakes (Delfino, 2019). While such concerns are justified, this limited focus results in more political attention and resources being dedicated to combatting political deepfakes than pornographic ones (Delfino, 2019). A study by Gosse and Burkell (2020) found that the misogynistic nature of pornographic deepfakes is barely ever mentioned in news coverage. Thus, such coverage fails to acknowledge the larger societal context that plays into how deepfake technology is being developed and employed, as discussed in the previous section. This public neglect points to an additional need for non-technical approaches to combat deepfake technology. Delfino (2019) stresses the importance of improving legislation, as well as educating and training law enforcement, the judiciary and the general public. Public awareness may be growing, as an instance of deepfake pornographic images of singer-songwriter Taylor Swift caused thus-far unmatched public outrage in early 2024 (Olson, 2024). The public attention also resulted in US politicians and the White House advocating for improved legislation against deepfake pornography (Yousif, 2024). The long-term impacts of this incident are yet to be seen.

At this point, many questions remain open regarding the governance of deepfake pornography. These questions include: Who is the victim of deepfake pornography—only the person whose face is used or also the other people depicted in the video? Who is the perpetrator—the person creating the deepfake or the person distributing it? How can they be located? Can and should platforms hosting deepfakes be held accountable? How can we ensure that deepfakes are removed from the internet? Currently, there are no clear answers to these questions (Delfino, 2019; Meskys et al., 2020), thus scholars should continue to investigate them and policymakers should follow discussions on this subject closely to implement effective and clear legislation.

4.2 The Vulnerability Perspective

While the risk perspective gives important insights into how society deals with the impacts of deepfake pornography, the vulnerability perspective extends this view and shows how not all members of society are equally impacted by developments in deepfake pornography (Bijker et al., 2014). As of 2019, 99% of victims of deepfake pornography were female actors and musicians (Aijder et al., 2019). This reflects the abovementioned gendered dimension of deepfake pornography, as well as public figures being at higher risk. The vulnerability of celebrities can be explained by deep learning algorithms' need for vast amounts of data on the targeted individuals to generate realistic deepfakes (Ratner, 2021). This leaves celebrities especially vulnerable to being victims of deepfake pornography, as immense volumes of pictures and videos of them are available. The vulnerability perspective also offers a second consideration: not every individual will suffer the same consequences if they are victimised. For a public figure, it is comparatively easy to debunk a fake video, whereas for a person who is not in the public eye it is considerably harder (Giles et al., 2019). Therefore, private individuals may be less vulnerable to being featured in deepfake pornography than famous people but when they are, they are likely to be more vulnerable to the consequences. Importantly, these vulnerabilities are not fixed but context dependent. For example, the risk of being featured in deepfake pornography may increasingly shift toward private individuals. People with an active online presence are already at an increased risk, as many images and videos are available of them. However, as deepfake technology progresses, it will likely need increasingly lower input to produce realistic outcomes (Giles et al., 2019). This would also make people with less of an online presence more vulnerable. In a more optimistic scenario, women could be rendered less vulnerable if the societal context becomes less misogynistic.

Additionally, Bijker (2006) points to the potentially positive and even necessary role of vulnerabilities as a prerequisite for innovation. Although deepfake technology creates vulnerabilities in certain use cases, it is also used in beneficial ways. For example, it enables victims of abuse to share their stories while ‘wearing’ a mask, thereby hiding their identity but accurately conveying their emotions (Kerner & Risse, 2021). Furthermore, it allows for the creation of engaging educational content, such as realistic videos of historical figures talking to students (Chesney & Citron, 2019). Linking this back to the risk governance considerations discussed earlier, these examples show why an outright ban of deepfake technology is neither desirable nor beneficial (Spivak, 2018), rendering a possible response more complicated.

5. Conclusion

Having applied theories and concepts from Science and Technology Studies, this paper offered insights into the current state of deepfake technology in pornography. By looking at the phenomenon through the lens of social constructivism, the political nature of certain deepfake applications and the political structures surrounding all forms of deepfake technology were demonstrated. Specifically, this demonstrated the role that misogynistic power structures play in the development of deepfake technology. The risk perspective revealed a lack of legislation, especially on a federal and global level, as well as a lack of public awareness of the risks of pornographic deepfakes, although recent developments point to potential improvements. The vulnerability perspective emphasised the context-dependent nature of the impact of deepfake pornography, where celebrities are currently more likely to be targeted but private individuals have greater difficulty disproving deepfakes of themselves. Additionally, it drew attention to the potentially beneficial uses of deepfake technology, questioning the utility of an outright ban. Taken together, these insights show that in addition to developing technological responses to deepfake pornography, societal responses are also needed. Such responses should include education on deepfake pornography, (global) legislation controlling it and general effort against the present misogynistic underpinnings.

Finally, limitations of this paper shall be mentioned. While the risk governance framework did touch on legislation surrounding deepfake technology, it could not adequately capture the extensive discussion surrounding deepfake pornography laws. As legislation will play a central role in how society deals with this new technology, close attention should be paid to the scholars working on these topics. For example, the risk of deepfake child pornography being produced merits further research (Ratner, 2021). The question of how to define the victims and perpetrators of deepfake pornography also remains unresolved (Delfino, 2019; Meskys et al., 2020). Furthermore, this paper focused solely on pornographic deepfakes and considered them separately from political deepfakes, even though developments regarding, the two, such as legislation and public awareness, are intertwined. Finally, it was outside the scope of this paper toapply all available concepts and theories from Science and Technology studies to the case of deepfake technology. For example, Technofeminism (Wajcman, 2007) or New and Emerging Science and Technology (NEST) Ethics (Swierstra & Rip, 2007) were not applied. Indeed, even within the used risk framework, a focus was set on risk governance. Applying a wider scope of theories and concepts in future research may provide further useful insights.

References

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). *The state of deepfakes: Landscape, threats, and impact.* Deeptrace. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Beaumont-Thomas, B. (2024, January 26). Taylor Swift deepfake pornography sparks renewed calls for US legislation. *The Guardian.* https://www.theguardian.com/music/2024/jan/26/taylor-swift-deepfake-pornography-sparks-renewed-calls-for-us-legislation

Bijker, W. E. (2001). Understanding technological culture through a constructivist view of science, technology, and society. In S. H. Cutcliffe, & C. Mitcham (Eds.), *Visions of STS: Counterpoints in science, technology, and society studies* (pp. 19-34). State University of New York Press.

Bijker, W. E. (2006). The vulnerability of technological culture. In H. Nowotny (Ed.), *Cultures of technology and the quest for innovation* (pp. 52-70). Berghahn Books. https://doi.org/10.1515/9781782389644-006

Bijker, W., Hommels, A., & Mesman, J. (2014) Studying vulnerability in technological cultures. In: A. Hommels, J. Mesman, & W.E. Bijker (Eds.), *Vulnerability in technological cultures: New directions in research and governance* (pp. 1-26). The MIT Press. https://doi.org/10.7551/mitpress/9209.003.0002

Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*(6), 1753-1820. https://doi.org/10.15779/Z38RV0D15J

Delfino, R. A. (2019). Pornographic deepfakes: The case for federal criminalization of revenge porn's next tragic act. *Fordham Law Review*, *88*(3), 887-938. https://ir.lawnet.fordham.edu/flr/vol88/iss3/2

Eggestein, J., & Knapp, K. (2014). Fighting child pornography: A review of legal and technological developments. *Journal of Digital Forensics, Security and Law*, *9*(4), 29-48. https://doi.org/10.15394/jdfsl.2014.1191

Franks, M., & Waldman, A. (2019). Sex, lies, and videotape: Deep fakes and free speech delusions. *Maryland Law Review*, *78*(4), 892-898. https://digitalcommons.law.umaryland.edu/mlr/vol78/iss4/6/

Giles, K., Hartmann, K., & Mustaffa, M. (2019). *The role of deepfakes in malign influence campaigns*. NATO Strategic Communications Centre of Excellence. https://www.stratcomcoe.org/role-deepfakes-malign-influence-campaigns

Gosse, C., & Burkell, J. (2020). Politics and porn: How news media characterizes problems presented by deepfakes. *Critical Studies in Media Communication*, *37*(5), 497-511. https://doi.org/10.1080/15295036.2020.1832697

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 149-152. https://doi.org/10.1089/cyber.2021.29208.jth

Harris, D. (2019). Deepfakes: False pornography is here and the law cannot protect you. *Duke Law & Technology Review*, *17*, 99-128. https://scholarship.law.duke.edu/dltr/vol17/iss1/4

Harris, K. R. (2021). Video on demand: What deepfakes do and how they harm. *Synthese, 199*, 13373-13391. https://doi.org/10.1007/s11229-021-03379-y

Karasavva, V., & Noorbhai, A. (2021). The real threat of deepfake pornography: A review of Canadian policy. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 203-209. https://doi.org/10.1089/cyber.2020.0272

Kerner, C., & Risse, M. (2021). Beyond porn and discreditation: Epistemic promises and perils of deepfake technology in digital lifeworlds. *Moral Philosophy and Politics*, *8*(1), 81-108. https://doi.org/doi:10.1515/mopp-2020-0024

Kleinman, D. L. (2005). *Science and technology in society: From biotechnology to the internet.* Wiley-Blackwell.

Mania, K. (2024). Legal protection of revenge and deepfake porn victims in the European Union: Findings from a comparative legal study. *Trauma, Violence, & Abuse*, *25*(1), 117-129. https://doi.org/10.1177/15248380221143772

Meskys, E., Liaudanskas, A., Kalpokiene, J., & Jurcys, P. (2020). Regulating deep fakes: Legal and ethical considerations. *Journal of Intellectual Property Law & Practice*, *15*(1), 24-31. https://doi.org/10.1093/jiplp/jpz167

Öhman, C. (2020). Introducing the pervert's dilemma: A contribution to the critique of deepfake pornography. *Ethics and Information Technology*, *22*, 133-140. https://doi.org/10.1007/s10676-019-09522-1

Olson, P. (2024, January 29). Can Taylor Swift save humanity from AI's dark side? *Bloomberg*. https://www.bloomberg.com/opinion/articles/2024-01-29/deepfake-porn-demands-taylor-swift-justice

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14*(3), 399–441.

Ratner, C. (2021). When "Sweetie" is not so sweet: Artificial Intelligence and its implications for child pornography. *Family Court Review*, *59*(2), 386-401. https://doi.org/10.1111/fcre.12576

Reid, S. (2021). The deepfake dilemma: Reconciling privacy and first amendment protections. *University of Pennsylvania Journal of Constitutional Law*, *23*(1), 209-238. https://scholarship.law.upenn.edu/jcl/vol23/iss1/5

Spivak, R. (2019). "Deepfakes": The newest way to commit one of the oldest crimes. *Georgetown Law Technology Review*, *3*(2), 339-400. https://georgetownlawtechreview.org/deepfakes-the-newest-way-to-commit-one-of-the-oldest-crimes/GLTR-05-2019/

Swierstra, T., & Rip, A. (2007). Nano-ethics as NEST-ethics: Patterns of moral argumentation about new and emerging science and technology. *NanoEthics*, *1*, 3–20. https://doi.org/10.1007/s11569-007-0005-8

Van Asselt, M. B. A., & Renn, O. (2011). Risk governance. *Journal of Risk Research*, *14*(4), 431-449. https://doi.org/10.1080/13669877.2011.553730

Wajcman, J. (2007). From women and technology to gendered technoscience. *Information, Communication & Society*, *10*(3), 287–298. https://doi.org/10.1080/13691180701409770

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136.

Woolgar, S., & Cooper, G. (1999). Do artefacts have ambivalence? Moses' bridges, Winner's bridges and other urban legends in S&TS. *Social Studies of Science*, *29*(3), 433–449.

Wyatt, S. (1998). *Technology's arrow: Developing information networks for public administration in Britain and the United States* [Doctoral dissertation, Maastricht University]. Datawyse / Universitaire Pers Maastricht. https://doi.org/10.26481/dis.19981105sw

Yousif, N. (2024, January 28). X blocks searches for Taylor Swift after explicit AI images of her go viral. *BBC*. https://www.bbc.com/news/world-us-canada-68123671