FACE/OFF: THE DAMAGING IMPACTS OF DEEPFAKES

Anokhy Desai

INTRODUCTION

In April of 2020, ESPN aired an ad for an upcoming documentary, showing a doctored 1998 SportsCenter clip that predicted the release of such a documentary and used the word "lit," joking that viewers "don't even know what that means yet."[1] The ad, with its surprising predictions, was actually a deepfake, or a video doctored with the help of artificial intelligence (AI). While advertisers have graduated to deepfake technology over the older methods like Photoshop to produce memorable content like this, creators with less benign intent have used the technology to create pornography of public figures and celebrities, and even private individuals, without their consent. In fact, nonconsensual pornography comprises 96% of deepfakes on the Internet.[2]

The First Amendment's freedom of speech protections may provide a user with the right to post content online, but using the likeness of someone else without their consent, especially for a harmful purpose like pornography, is a known privacy violation under the EU's General Data Privacy Regulation (GDPR), various U.S. state privacy and biometric laws, and, for public figures, defamation laws.[3] What is worse, deepfake porn publishers are not held accountable because their real identities are usually untraceable, and most victims cannot hold the platforms hosting the content liable due to Section 230 of the Communications Decency Act. In order to provide justice to the private citizens and public figures whose likenesses are used without their consent, Congress can amend § 230, incentivizing such content's removal; platforms can utilize upcoming technical solutions, both offensive and defensive; and victims can look to their state's available privacy protections for legal remedies.

---

[1] Tiffany Hsu, *An ESPN Commercial Hints at Advertising's Deepfake Future*, N.Y. TIMES (Apr. 22, 2020), https://www.nytimes.com/2020/04/22/business/media/espn-kenny-mayne-state-farm-commercial.html.

[2] Henry Ajder et al., *The State of Deepfakes: Landscape, Threats, and Impact*, DEEPTRACE LABS 7 (Sept. 2019), http://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

[3] Natalie A. Prescott, *The Anatomy of Biometric Laws: What U.S. Companies Need to Know in 2020*, NAT'L L. REV. (Jan. 15, 2020), https://www.natlawreview.com/article/anatomy-biometric-laws-what-us-companies-need-to-know-2020; 42 PA. CONS. STAT. § 8316(e)(2); WASH REV. CODE § 63.60.070(1).

A. DEEPFAKES: A STUBBORN PROBLEM

Deepfakes are media content, usually involving face-swapping, that make it look and sometimes sound like a person is doing or saying something they did not. Those curious about deepfakes no longer need a degree in machine learning or technical knowledge of AI. Deepfake technology, named for the deep learning process necessary to create fake audio and videos, is readily available to anyone with a smartphone.[4] Snapchat "Cameos" let users paste their face onto a number of preset videos, and TikTok seems to be following suit.[5] In addition to sharing short videos reminiscent of JibJab e-cards, users have created eerily well-edited movie clips in which Nicolas Cage is superimposed on another actor's face using software like FakeApp.[6]

The term "deepfake" was already known when Nicolas Cage-superimpositions went viral, but became mainstream because of less "safe for work" content. Reddit user and amateur programmer "u/deepfakes" amassed a fan base after posting deepfake videos he created of several famous female faces, including those of Michelle Obama and Gal Gadot, superimposed onto adult actresses.[7] He crossed the Uncanny Valley by feeding publicly-available stock images and interviews of the actresses, as well as existing pornography, into AI software that autonomously maps faces. These programs are open-source and thus readily available to anyone from researchers to hobbyists. Unsurprisingly then, the total number of deepfake videos online totaled 49,000 this June, a 226% increase from last summer.[8] These efforts do not go unnoticed; the top four deepfake porn websites have received more than 134 million views.[9]

The unnamed creator of FakeApp made the deepfake software as a "creative experiment" and was disconcerted to discover the

---

[4] Jan Kietzmann et al., *Deepfakes: Trick or Treat?*, 63(2) BUS. HORIZONS 135, 135-46 (Mar. 2020), https://doi.org/10.1016/j.bushor.2019.11.006.

[5] Josh Constine, *ByteDance & TikTok Have Secretly Built a Deepfakes Maker*, TECHCRUNCH (Jan. 3, 2020), https://techcrunch.com/2020/01/03/tiktok-deepfakes-face-swap/.

[6] The irony was not lost on Twitter users who commented about Cage's role in Face/Off, a movie quite literally involving swapping faces. Sam Hysom, *People Are Using Face-Swapping Tech to Add Nicolas Cage to Random Movies and What is 2018*, MASHABLE (Jan. 31, 2018), https://mashable.com/2018/01/31/nicolas-cage-face-swapping-deepfakes/.

[7] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, VICE (Dec. 11, 2017), https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn.

[8] Matt Burgess, *Porn Sites Still Won't Take Down Nonconsensual Deepfakes*, WIRED (Aug. 30, 2020), https://www.wired.com/story/porn-sites-still-wont-take-down-non-consensual-deepfakes/.

[9] *Id.*

app's primary use.[10] While he does not support the use of his app for nonconsensual pornography, he refused to condemn deepfake technology and AI, instead stating that any technology can be used for harmful purposes. He has continued to improve the app and hopes to have his technology incorporated into low-budget projects that would use his software in place of expensive special effects. Users have already demonstrated the app's impressive capabilities, showing the nearly imperceptible difference between a homemade version of a Star Wars movie scene and that in the theatrical release.[11]

If the creator of the app most commonly used to doctor nonconsensual porn and his fellow computer scientists continue to develop this technology, deepfake porn and the issues arising from it cannot be solved by banning the software outright, as a national ban would not stop the technology's international development or its products domestic availability. The explicit media harms women on both sides of the conversion; the faces used without consent typically belong to women who are public figures or celebrities, and the bodies belong to uncompensated and uncredited adult actresses.

When celebrities' likenesses are not sampled, creators use private citizens' photos taken from sources like Google Images or Facebook profiles. It is true that celebrities knowingly put themselves in the public eye through their roles, interviews, and public-facing social media accounts. However, they understand and consent to the public nature of these appearances. Celebrities can choose to not attend an interview on a late-night show or not take an audition for a role. Private citizens can similarly choose to make their social media accounts private and to un-list their profiles from search engines if they do not consent to having their information and images made public.[12]

The moment a deepfake video is created using the likeness of a nonconsenting woman, that choice is taken away and their right to bodily privacy is fundamentally compromised. Noelle Martin was just another law student until multiple deepfakes of her selfie were published, changing the course of her personal life and career. Now an advocate for victims of similar deepfake-based harassment,

---

[10] Kevin Roose, *Here Come the Fake Videos, Too*, N.Y. TIMES (Mar. 4, 2018), https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html.

[11] Dave Itzkoff, *How 'Rogue One' Brought Back Familiar Faces*, N.Y. TIMES (Dec. 27, 2016), https://www.nytimes.com/2016/12/27/movies/how-rogue-one-brought-back-grand-moff-tarkin.html; Samantha Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, VICE (Jan. 24, 2018), https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley.

[12] FACEBOOK, *What Should I Do If I Don't Want Search Engines to Link to My Facebook Profile*? (last visited Dec. 20, 2020), https://www.facebook.com/help/124518907626945.

Martin speaks internationally about her years-long struggle to repair her reputation, which was ruined despite her never having participated in explicit media.[13] Actress Kristen Bell was also shocked to learn that there was deepfake pornography of her.[14] Her fame spared her reputation, but the lack of consent felt even more violating in the #MeToo era. The fact that these videos exist and get millions of views, even with "fake" in the title or tags, is unsurprising.

Deepfake porn exists to objectify and humiliate women. Therefore, reintroducing the human element by obtaining consent removes most of the allure for many consumers. As with rape, deepfakes are about power, not sex. Boston University law professor Danielle Citron explains that the appropriation of someone's image without their consent may feel like a victimless crime, but realistically amounts to a form of virtual sexual assault.[15]

B. USER PROTECTION: A WORK IN PROGRESS

The technology's development continues, and new U.S. state laws similar to the first anti-deepfake law in New South Wales are slow to catch up. If we cannot remove the vehicle, the issue then becomes two-fold: how can we hold the content creators accountable, and how can we protect the victims?

The unfortunate reality is that deepfake content will remain available online. It is very difficult to hold a deepfake content creator accountable when none of their real information is linked to their account. It is equally difficult to hold content hosts accountable when they are protected by § 230 and thus have no incentive to remove popular, monetizable videos. When Martin contacted porn sites to have her likeness removed from their platforms, she got more silence and threats than removal confirmations. Additionally, even if a video is removed from one site, it will find its way to another. Videos from the now-banned Reddit page "r/deepfakes" have appeared in multiple other forums.[16] When asked how they are

---

[13] Noelle Martin, *Online Predators Spread Fake Porn of Me. Here's How I Fought Back.* TED (Nov. 2017), https://www.ted.com/talks/noelle_martin_online_predators_spread_fake_porn_of_me_here_s_how_i_fought_back.

[14] Claudia Willen, *Kristen Bell Says She Was 'Shocked' to Learn That Her Face Was Used in a Pornographic Deepfake Video*, INSIDER (Jun. 11, 2020), https://www.insider.com/kristen-bell-face-pornographic-deepfake-video-response-2020-6.

[15] EJ Dickson, *TikTok Stars Are Being Turned into Deepfake Porn Without Their Consent*, ROLLING STONE (Oct. 26, 2020), https://www.rollingstone.com/culture/culture-features/tiktok-creators-deepfake-pornography-discord-pornhub-1078859/.

[16] Leo Kelion, *Reddit Bans Deepfake Porn Videos*, BBC (Feb. 7, 2018), https://www.bbc.com/news/technology-42984127; Katyanna Quach, *It Took Us*

addressing this issue, popular porn sites that host of deepfakes like PornHub and xHamster stated that they treat such media "like any other nonconsensual content."[17] Giorgio Patrini, CEO and chief scientist at Sensity, formerly Deeptrace Labs,[18] balked at the companies' responses, saying the "attitude of these websites is that they don't really consider this a problem."[19] xHamster's Vice President explained that the company's moderation process includes removing videos if individuals report their image being used without permission and if the video is a violation of the site's Terms of Use.[20] Despite this assurance, the site's moderators are encouraged to find a way to keep nonconsensual content up.[21] Only when Visa and Mastercard cut ties with PornHub after allegations of unlawful content on the site, including portrayals of the sexual abuse of minors, did the site respond with a real policy change.[22] The site purged over 10 million videos and permanently banned uploads from unverified users.[23]

While deepfake content creators currently face no repercussions, American deepfake victims have some legal recourse available in the state-level privacy laws protecting biometric data and personally identifiable information (PII), which often have language that protects a person's likeness. Some states even have deepfake-specific laws that forbid and punish creating and sharing nonconsensual deepfake content. Affected individuals, especially celebrities, can also consider copyright and defamation claims, but must ultimately keep in mind their low chances of success give the protections § 230 provides to online service providers like porn sites.

1. SECTION 230 OF THE COMMUNICATIONS DECENCY ACT

Congress enacted § 230 of the Communications Decency Act to protect developing technologies from excessive liability claims for third party content on their platforms. The statute states that

---

Less Than 30 Seconds to Find Banned 'Deepfake' AI Smut on the Internet, THE REGISTER (Feb. 9, 2018), https://www.theregister.com/2018/02/09/deepfake_ai/.

[17] Burgess, *supra* note 8.

[18] *See generally* Aider, *supra* note 2.

[19] Burgess, *supra* note 17.

[20] *Id.*

[21] Sebastian Meineck & Yannah Alfering, *We Went Undercover in xHamster's Unpaid Content Moderation Team*, VICE (Oct. 27, 2020), https://www.vice.com/en/article/akdzdp/inside-xhamsters-unpaid-content-moderation-team.

[22] Samantha Cole, *Visa and Mastercard Will Stop Processing Payments to Pornhub*, VICE (Dec. 10, 2020), https://www.vice.com/en/article/7k94be/mastercard-will-stop-processing-payments-to-pornhub.

[23] *Id.*

"interactive computer service [providers]" cannot be treated as the publisher of information posted by a user, and thus cannot be held liable for that content on behalf of the user.[24] While Congress originally intended this rule to foster growth and innovation, § 230 now serves as a shield for many of the largest and most powerful companies in the country, including tech giants like Facebook and taboo yet household names like PornHub. Porn sites such as xHamster use § 230 to avoid liability for deepfake videos, as all content is user-provided. Even though that content is nonconsensual and supposedly monitored, it is not technically illegal to host, and thus the site has no real incentive to remove it.

2. PRIVACY REGULATIONS

The E.U.'s GDPR introduced new terminology that was adopted for the U.S.'s California Consumer Privacy Act (CCPA) and will be kept by its replacement, the California Privacy Rights Act (CPRA). Data subjects are the individuals whose data is collected and processed, data controllers are the entities who make decisions about what to do with subjects' data, and data processors are the entities that process the data based on those decisions. The GDPR protects data subjects by providing them with several rights, a few of which also exist under CCPA, like the right to know about the PII a business collects from them and the right to delete it.[25]

Both the GDPR and CCPA protect user-generated data like images posted to social media, including images that contain the likenesses of people other than the poster.[26] In the U.S., CCPA applies to matters involving biometric information, including voice recordings and faceprints, and can be cited to fine businesses that have not de-identified such data.[27] De-identifying faceprints on deepfake porn would involve blurring the face, which would render deepfake creation moot. This would be a big win for victims, but only California residents are protected by CCPA.[28] Nevada and Maine have enacted consumer privacy legislation this year, but neither include a way for deepfake victims to control the use of their image.[29]

---

[24] 47 U.S.C. § 230(a).

[25] CAL. OFFICE OF THE ATT'Y GEN., *California Consumer Privacy Act (CCPA)* (last visited Dec. 20, 2020), https://oag.ca.gov/privacy/ccpa.

[26] *Id.*; Dan Shewan, *10 Things You Need to Know About the EU General Data Protection Regulation*, WORDSTREAM (Aug, 13, 2019), https://www.wordstream.com/blog/ws/2017/09/28/eu-gdpr.

[27] Cal. Civ. Code §1798.140(b) (2018).

[28] *Id.*

[29] Sarah Rippy, *US State Comprehensive Privacy Law Comparison*, IAPP (Nov. 17, 2020), https://iapp.org/resources/article/state-comparison-table/; NEV. REV. STAT § 603A.040; 35-A ME. REV. STAT c. 94 §9301(1).

3. AVAILABLE REMEDIES

The first state in the country to enact such a law, Illinois enacted the Biometric Information Privacy Act (BIPA) in 2008 to regulate the way biometric information is collected, processed, and stored. Like CCPA, BIPA recognizes faceprints as protected data and can be used to fine noncompliant businesses, like porn sites that do not securely store the deepfake videos on their platforms or take security measures to prevent them from being shared to other platforms. Similarly, Texas, Washington, New York, and Arkansas have enacted biometric data laws.[30] If a deepfake victim resides in a state that has a biometric data law, they may be able to take private action through lawsuits, class actions, or requests to their attorney general's office, but as these laws still consider properly stored data to be "protected" despite thousands of views by other users, these remedies only go so far.

Other adult sites can proactively help victims by using copyright law to prevent the spread of deepfakes on their platforms. PornHub's copyright system uses machine learning to automatically detect when copyrighted videos are uploaded.[31] Sites can then set up an auto-delete feature for copyrighted content and follow a three-strikes policy for uploaders of copyrighted content. Still, sites may face an issue with watermarked content and original deepfakes, so deepfake victims may not find many protections in copyright law.

In defamation case law, a plaintiff must demonstrate a false assertion purporting to be fact that caused them harm through written or verbal communication to a third person, made with negligence. Public figures must meet higher standards to show malice, but any deepfake plaintiff would need to show at minimum that the deepfake creator or poster negligently made a false assertion about them to viewers in order to make a prima facie case, meaning that simply putting "deepfake" in a video's title often defeats this approach.[32] Counterintuitively, celebrities generally have more protection over their image. Deceased celebrities are protected under the Celebrities Rights Act, which reversed *Lugosi v. Universal Pictures* and preserved a celebrity's right of publicity after their death, while living celebrities' publicity rights are protected by laws like the California Civil Code.[33] Whether a deepfake video would count as a false assertion is up to the factfinder, but the

---

[30] Prescott, *supra* note 3.

[31] Samantha Cole, *Facial Recognition for Porn Stars Is a Privacy Nightmare Waiting to Happen*, VICE (Oct. 11, 2017), https://www.vice.com/en/article/a3kmpb/facial-recognition-for-porn-stars-is-a-privacy-nightmare-waiting-to-happen.

[32] N.Y. Times Co. v. Sullivan, 376 U.S. 254, 254 (1964).

[33] CAL. CIV. CODE §3344(a) (1872).

plaintiff can try bringing false light or invasion of privacy claims regarding the use of their digitally manipulated likeness.[34]

Finally, some states have enacted deepfake laws, including Virginia, Texas, and California, the first imposing criminal penalties on deepfake porn creators and propagators, and the last allowing victims to sue for damages.[35]

C. SUGGESTED MITIGATIONS: THE NEXT STEP

Congressional action comes in where existing remedies are lacking. Last year, a national defense bill focusing on deepfakes used for political disinformation activities was signed into law, but stopped short of banning explicit or nonconsensual deepfakes.[36] Congress can, however, amend § 230 to hold platforms liable for nonconsensual posts. GDPR's compliance rate indicates that fines can incentivize businesses to follow the law, and PornHub's video purge after the brief loss of its financial vendors illustrates that financial disincentives can also work, but also that neither fully solves the problem.[37]

Additionally, existing legal frameworks like copyright, defamation, and right of publicity claims may not be helpful to the majority of deepfake victims, who do not find those laws applicable, cannot afford to bring such matters to court, or cannot afford an investigator to identify the real creator. Even with her monetary resources, actress Scarlett Johansson found that trying to protect her image internationally via copyright claims failed, and that sending removal requests did not prevent the media from being re-uploaded on other platforms.[38] Given the broad applicability of theories like the fair use doctrine and creative work exception, identified deepfake uploaders have precedent to argue their works involving

---

[34] David Fink & Sarah Diamond, *Deepfakes: 2020 and Beyond*, LAW.COM (Sept. 3, 2020), https://www.law.com/therecorder/2020/09/03/deepfakes-2020-and-beyond/.

[35] *Id.*; Tom Simonite, *Most Deepfakes Are Porn, and They're Multiplying Fast*, WIRED (Oct. 7, 2019), https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/.

[36] Matthew Ferraro et al., *First Federal Legislation on Deepfakes Signed Into Law*, WILMERHALE (Dec. 23, 2019), https://www.wilmerhale.com/en/insights/client-alerts/20191223-first-federal-legislation-on-deepfakes-signed-into-law.

[37] HELPNET SECURITY, *Despite Potential Fines, GDPR Compliance Rate Remains Low* (Dec. 4, 2019), https://www.helpnetsecurity.com/2019/12/04/gdpr-compliance-rate/.

[38] Drew Harwell, *Scarlett Johansson On Fake AI-Generated Sex Videos*, WASH. POST (Dec. 31, 2018), https://www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/.

celebrities are protected as creative works that are transformative and not plagiarized.[39]

This gap in offensive and defensive remedies can be filled by the tool that created the problem: technology. Researchers have already begun building a deepfake detection system that considers lighting, shadows, and facial movements to flag doctored images and video frames.[40] Once available for use, porn sites can utilize these kinds of software to take down videos of unverified individuals or celebrities in the content. This method would require the second step of adult content creators providing their faces to a verified creator database. This step may seem as invasive as the deepfake itself, but sites like PornHub already require their models to provide such verification to create an account and content.[41] This offensive strategy, combined with the technical defensive strategy of adding a filter to an image file to make it impossible to generate a deepfake from the image, can better protect the thousands of women whose likenesses are used without consent today.[42]

One final issue remains. If a deepfake contains an adult actress's full body and just the face of a third party, does the third party have standing to take legal action? After all, not all deepfakes are created with the most advanced technology, leaving some videos looking very unrealistic. The poor quality of the video and the use of "fake" or "deepfake" in the title can further remove the video from the third party, but the fact that her image was used without her consent is still a problem. Viewers still suspend their belief enough to give deepfake porn videos millions of views, meaning those women are still virtually assaulted. Additionally, it can be difficult for untrained viewers to differentiate between deepfakes and authentic videos; the fantasy is that the viewer is truly watching the third party.

There is no one answer to the issue of deepfake pornography. Deepfake porn causes real harm to women whose likenesses are used for nonconsensual explicit content, and AI-generated deepfake porn raises ethical issues about the objectification of women and security and privacy concerns for the women whose images are stored in the database required for the AI to create the content. With

---

[39] U.S. COPYRIGHT OFFICE, *More Information on Fair Use* (Oct. 2020), https://www.copyright.gov/fair-use/more-info.html.; Hoepker v. Kruger, 200 F. Supp.2d 340, 340 (S.D.N.Y. 2002).

[40] Sheng-Yu Wang et al., *CNN-Generated Images Are Surprisingly Easy to Spot... For Now* (Apr. 4, 2020), https://arxiv.org/abs/1912.11035.

[41] Samantha Cole, *Pornhub Just Purged All Unverified Content From the Platform*, VICE (Dec. 14, 2020), https://www.vice.com/en/article/jgqjjy/pornhub-suspended-all-unverified-videos-content.

[42] Nataniel Ruiz et al., *Disrupting Deepfakes: Adversarial Attacks Against Conditional Image Translation Networks and Facial Manipulation Systems* (Apr. 27, 2020), https://arxiv.org/pdf/2003.01279.pdf.

the available legal avenues, new regulations, and developing technical solutions, however, we can lessen the harm to women.