Using Fake Video Technology To Perpetrate Intimate Partner Abuse

Domestic Violence Advisory

Authors: Adam Dodge, Laura's House & Erica Johnstone, Ridder, Costa & Johnstone LLP[1]

1. What is a deepfake?

There are a variety of techniques that can be used to create videos and other content that misrepresent people and events. One that has come to public attention recently is colloquially referred to as "deepfake" technology, named after a Reddit user who helped popularize it. The technology uses an artificial intelligence method called deep learning to recognize and swap faces in pictures and videos. The technique begins by analyzing a large number of photos or a video of someone's face, training an artificial intelligence algorithm to manipulate that face, and then using that algorithm to map the face onto a person in a video. Although this technique may have legitimate uses, it can also be used to perpetrate intimate partner abuse, by making it appear as though one's partner was in, for example, a pornographic video that they were not in fact in.

"Deep fake technology leverages machine learning techniques to manufacture facts about the world. It manipulates video and audio so individuals appear doing and saying things they never did or said."[2] This Advisory offers direction on how to address the problems deepfake videos can cause in the context of intimate partner abuse. The authors want victims and their supporters to know *everything* that can be done to combat face-swapped videos in California family court when the facts present as intimate partner abuse.

2. How can a face-swapped video be used to perpetrate domestic violence?

Unlawful conduct may take place at the point of photo capture and/or video distribution.

Unlawful capture of images by stalking,[3] surveillance, hacking,[4] force or threats: To create a face-swapped video of reasonable quality today, the perpetrator needs at least a few hundred

---

[1] Adam Dodge is the Legal Director at Laura's House and Public Policy Regional Representative with the California Partnership to End Domestic Violence. Erica Johnstone is a law partner at Ridder, Costa & Johnstone LLP, and cofounder of Without My Consent, a 501(c)(3) nonprofit that develops educational materials to empower victims of digital abuse to seek justice across the United States.

[2] *See* Danielle Citron and Robert Chesney's forthcoming law review article building on their Lawfare post: Danielle Citron & Robert Chesney, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?* LAWFARE BLOG (Feb. 21, 2018), https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy.

[3] California Penal Code section 646.9 defines stalking as: "willfully, maliciously, and repeatedly follows or willfully and maliciously harasses another person and who makes a credible threat with the intent to place that person in reasonable fear for his or her safety, or the safety of his or her immediate family…." The definition of a "credible threat" is also defined in the stalking statute. CAL. PENAL CODE § 646.9(g). Cyber-stalking, the repeated unwanted harassment of an individual through the use of digital media is covered by California's stalking statute. Johnette Jauron, *Navigating Snark-Infested Waters: When and How to Prosecute Online Harassment*, 31 CDAA Prosecutor's Brief no. 4 at 51-59.

[4] The Computer Fraud and Abuse Act (18 U.S.C. § 1030) and California Penal Code Section 502 make it a crime to knowingly access and without permission a computer, data, or computer system. Such conduct is also civilly actionable. CAL. PENAL CODE § 502(e).

photos of the victim.[5] To obtain the photos, a perpetrator may initiate, continue, or increase surreptitious surveillance and documentation of the target. Or, a perpetrator may obtain the photos by force, threats, harassment, and intimidation.

Unlawful distribution of face-swapped video: Unlawful conduct may also take place at the point of distribution because face-swapped video distribution is abuse that can be enjoined under California's Domestic Violence Prevention Act.

Under California Family Code section 6203, a restraining order can issue for: (1) intentionally or recklessly causing or attempting to cause bodily injury; (2) sexual assaulting; (3) placing a person in reasonable apprehension of imminent serious bodily injury to that person or another; or (4) any Family Code § 6320 behavior that has been or could be enjoined.[6] Family Code section 6320(a) behavior includes: molesting, attacking, striking, stalking, threatening, sexually assaulting, battering, credibly impersonating another person, falsely personating, harassing, making annoying telephone calls (including but not limited to annoying telephone calls), destroying personal property, contacting directly or indirectly by mail or otherwise, coming within a specified distance, or disturbing the peace of the other party.[7]

The distribution of face-swapped video as California Family Code § 6320 credible impersonation,[8] false personation,[9] stalking,[10] harassment,[11] and others. There are a variety of theories under which the distribution of fake videos may be defined as "abuse" for purposes of a restraining order application. By way of example, "credible impersonation" is abuse under the Domestic Violence Prevention Act. California Penal Code section 528.5 makes it a misdemeanor to credibly impersonate someone without their consent "through or on an Internet Web site or by other electronic means for purposes of harming, intimidating, threatening, or defrauding another

---

[5] *See* Samantha Cole, *People Are Using AI to Create Fake Porn of Their Friends and Classmates*, MOTHERBOARD (Jan. 26, 2018), https://motherboard.vice.com/en_us/article/ev5eba/ai-fake-porn-of-friends-deepfakes.

[6] CAL. FAMILY CODE § 6203.

[7] CAL. FAMILY CODE § 6320.

[8] Section 528.5 of the Penal Code makes it a misdemeanor to credibly impersonate someone without their consent "through or on an Internet Web site or by other electronic means for purposes of harming, intimidating, threatening, or defrauding another person." CAL. PENAL CODE § 528.5.

[9] When a suspect represents himself or herself as another person and does something that makes that person "become liable to any suit or prosecution, or to pay any sum of money, or to incur any charge, forfeiture, or penalty, or whereby any benefit might accrue to the party personating, or to any other person," they have committed the felony crime of false personation under Penal Code section 529.

[10] California Penal Code section 646.9 defines stalking as: "willfully, maliciously, and repeatedly follows or willfully and maliciously harasses another person and who makes a credible threat with the intent to place that person in reasonable fear for his or her safety, or the safety of his or her immediate family…." The definition of a "credible threat" is also defined in the stalking statute. CAL. PENAL CODE § 646.9(g). Cyber-stalking, the repeated unwanted harassment of an individual through the use of digital media is covered by California's stalking statute. Johnette Jauron, *Navigating Snark-Infested Waters: When and How to Prosecute Online Harassment*, 31 CDAA Prosecutor's Brief no. 4 at 51-59.

[11] California Penal Code section 653.2 defines harassment as engaging in a "knowing and willful course of conduct directed at a specific person that seriously alarms, annoys, torments, or terrorizes that person, and that serves no legitimate purpose." CAL. PENAL CODE § 653.2(a).

person."[12] So, for example, if the perpetrator opened email accounts or created social media profiles in the victim's name as part of the overall strategy to defame, harass, and stalk the victim through face-swapped video, then that conduct would be section 6320 credible impersonation abuse under California's Domestic Violence Prevention Act.

False personation is a separate offense under California's Domestic Violence Prevention Act. California Penal Code section 529 makes it a crime for a "person [to] falsely personate[] another in either his or her private or official capacity, and in that assumed character" do "any other act whereby, if done by the person falsely personated, he might, in any event, become liable to any suit or prosecution, or to pay any sum of money, or to incur any charge, forfeiture, or penalty, or whereby any benefit might accrue to the party personating, or to any other person." So, for example, if the perpetrator made it appear that the victim in the face-swapped video was offering to engage in prostitution, then that would be section 6320 false personation abuse under California's Domestic Violence Prevention Act.

Identity theft laws have been used successfully against nonconsensual porn perpetrators. *See, e.g.*, *People v. Bollaert,* 248 Cal. App. 4th 699, 732 (2016) (affirming judgment of the Superior Court of San Diego County, No. SCD252338, when the defendant, the operator of revenge porn site, U Got Posted, was convicted of six counts of extortion and 21 counts of identity theft in connection with activities related to posting more than 10,000 private, sexually explicit images of people without consent). There are factual scenarios contemplated by face-swapped video that might be simplest to handle via credible impersonation and false personation laws.

The distribution of face-swapped video as California Family Code § 6320 disturbing the peace. "[D]isturbing the peace" is abuse under the Domestic Violence Prevention Act.[13] "Disturbing the peace," within the meaning of Family Code § 6320, means "conduct that destroys the mental or emotional calm of the other party."[14] There are many ways in which one can disturb the peace of the other party. One way is through accessing, reading and publicly disclosing the person's confidential emails.[15] Another way is by downloading and disclosing or threatening to disclose another person's text messages containing "intimate details of [the parties'] lives."[16] Another way is through publicly disclosing the person's private images.[17]

The distribution of face-swapped porn also constitutes "disturbing the peace." It destroys the mental or emotional calm of the other party by hijacking the victim's identity without their permission to distribute believable videos of the victim doing and saying things (s)he never did

---

[12] CAL. PENAL CODE § 528.5. For purposes of this section, "electronic means" shall include opening an e-mail account or an account or profile on a social networking Internet Web site in another person's name. CAL. PENAL CODE § 528.5(c).

[13] CAL. FAMILY CODE § 6203(4) ("abuse" includes any behavior that could be enjoined pursuant to CAL. FAMILY CODE § 6320).

[14] *In re Marriage of Nadkarni*, 173 Cal. App. 4th 1483, 1497 (Cal. App. 6th Dist. 2009).

[15] *See, e.g.*, *id*.

[16] *In re Marriage of Evilsizor & Sweeney*, 237 Cal. App. 4th 1416 (Cal. App. 1st Dist. 2015).

[17] *See, e.g.*, *Michaels v. Internet Entertainment Group*, 5 F. Supp. 2d 823 (C.D. Cal. 1998) (issuing a preliminary injunction prohibiting the publication, distribution or other dissemination of a sexually explicit videotape of entertainers Pamela Anderson and Bret Michaels on multiple grounds, including privacy); *Mischa Barton v. Jon Zacharias,* Case No. BQ057336 (03/14/2017) (ex-boyfriend ordered to stay clear of petitioner and prohibited from selling, distributing, giving away or showing any of the videos or photos).

to spread falsities about the victim.[18] If disturbing the peace abuse can be perpetrated through the disclosure of true but confidential information and images (nonconsensual porn) of the victim, then it can also be perpetrated through the dissemination of fake and falsified information and images of the victim (face-swapped porn).

Describing the abuse: California Judicial Council Form DV-100, Request for Domestic Violence Restraining Order, Section 27, asks the petitioner to describe the abuse and "describe any injuries".[19] This is an opportunity for the petitioner to describe the impact on his/her emotional state, the damage to personal and professional reputation, whether he/she fears for safety, and any other consequences regarding emotional state, relationships, education and/or employment. It's helpful if the petitioner can include metrics quantifying the number of urls and page views. Without My Consent has published a Sample DV-100 Form completed for nonconsensual porn restraining order at WithoutMyConsent.org » Resources » Something Can Be Done Guide » Form DV-100.

Understanding the harm: Face-swapped porn inflicts the harm of sexual objectification without consent.[20] Like nonconsensual porn, face-swapped porn violates the partner's expectation that all aspects of sexual activity should be founded on consent.[21]

Compromising images and videos can damage an ex-partner's reputation, rendering them "unemployable, undateable and potentially at physical risk."[22] A single intimate image can quickly dominate the first several pages of search engine results for the victim's name, as well as being emailed or otherwise exhibited to the victim's family, employers, prospective employers,[23] co-workers, and peers.[24] Victims have been fired from their jobs, expelled from their schools, and forced to move from their homes.[25] They have been threatened with sexual assault, stalked,

---

[18] *See* Danielle Citron & Robert Chesney, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?* LAWFARE BLOG (Feb. 21, 2018) (describing "the cheap and easy fabrication of content that hijacks one's identity—voice, face, body"), https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy.

[19] Judicial Council of California, Form DV-100 (rev. July 1, 2016), http://www.courts.ca.gov/documents/dv100.pdf.

[20] Madeline Buxton, *The Deep, Dark World Of Fake Porn*, REFINERY 29 (Feb. 7, 2018) (quoting Mary Anne Franks), https://www.refinery29.com/2018/02/190051/deepfakes-ai-assisted-fake-porn.

[21] As one commentator recently put it, "revenge porn is patently offensive because the distributor's conduct, in context, offends the fundamental principle of consent in sexual relationships." Cynthia Barmore, *Criminalization in Context: Involuntariness, Obscenity, and the First Amendment*, 67 STAN. L. REV. 447, 463 (2015).

[22] Erica Goode, *Getting revenge porn removed*, THE SYDNEY MORNING HERALD (Sept. 30, 2013) (quoting Danielle Citron), https://www.smh.com.au/technology/getting-revenge-porn-removed-20130930-2un2z.

[23] According to a 2009 Microsoft study, nearly 80 percent of employers consult search engines to collect information about job applicants, and of the U.S. recruiters and HR professionals surveyed, 70% say they have rejected candidates based on information they found online. Reasons for not interviewing and hiring applicants may include concerns about their "lifestyle," "inappropriate" online comments, and "unsuitable" photographs, videos, and other information about them. Cross-Tab, *Online reputation in a connected world*. (Jan. 2010), http://go.microsoft.com/?linkid=9709510 (last visited Mar. 23, 2018). Victims may apply for jobs but if search results show that a number of pornographic websites have published what appears to be the victims' nude photos and videos, then the employer is probably going to reject the application.

[24] *See* Mary Anne Franks, *Revenge Porn Reform: A View from the Front Lines*, 69 FLA. L. REV. (forthcoming Spring 2018) (describing the harm with citations).

[25] *Id.*

and harassed.[26] Victims have developed post-traumatic stress disorders, depression, anxiety, agoraphobia, and difficulty maintaining intimate relationships. Some victims have committed suicide.[27]

A fake video that causes an audience to believe that a partner was featured in revenge porn can cause precisely the same kind of reputation, privacy, and property harms, and can rob people of their potential for the rest of their lives

When perpetrated on an intimate partner, that partner is likely to get a same day restraining order in California family court.[28]

3. How easy is it to create a face-swapped video?

The technology is available to everyone. Software tools like FakeApp are specifically designed to allow people without a technical background or programming experience to create deepfakes. The process is straightforward:

1. Download a program like FakeApp.
2. Ensure your computer has a reasonably powerful graphics card. A consumer-grade gaming-capable graphics card (GTX 1060 with 6GB of VRAM) requires about 10-20 hours of model training.
3. Identify the video to be used.
4. Collect, scrape or take hundreds of photos (or pull images from video content) of the victim to be superimposed into the video.
5. Feed the photos into FakeApp and run the program.

Following the steps above, a New York Times reporter used FakeApp to create a semi-realistic deepfake video of his face grafted onto the actor Chris Pratt's body in a video using 1,841 photos of himself.[29] The first try was unconvincing. But, when the reporter tried again and enlisted the help of a redditor who had a more intuitive understanding of blending facial features and source footage, the result produced very realistic footage of the reporter as Jimmy Kimmel.

Currently, the quality of deepfakes varies. They range from completely garbled to nearly flawless. However, a motivated perpetrator, with sufficient photo and video content of their target, can simply hire someone with the expertise to create a quality face-swapped video.

4. Is this currently a problem for survivors of domestic violence?

There's every indication that it will become a problem that practicing lawyers and legal service organizations will encounter. Although there are legitimate uses for deepfakes, any dual-use

---

[26] *Id.*
[27] *Id.*
[28] *See* Without My Consent, *Something Can Be Done Guide,* WITHOUT MY CONSENT (2017), *available at* http://withoutmyconsent.org/resources/download-guide.
[29] 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.

technology that is used by a significant number of people is unfortunately likely to be used in the context of domestic violence. This technology is becoming increasingly popular. For example, one Reddit group (before it was shut down) had amassed 100,000 users. The New York Times has reported that FakeApp has been downloaded 120,000 times and counting since January 2018.[30]

People are discussing, and in some cases creating fake porn videos of people they know in real life, like their exes, without their permission. They are openly discussing how to use face-swapping videos to perpetrate blackmail and domestic abuse. Below are examples from two prominent online communities.[31]

> Reddit: In a deepfakes subreddit, one user posted: *"Hi. I want to make a porn video with my ex-girlfirend. But I don't have any high-quality video with her, but I have a lot of good photos. If I can do something using only a photos? P.S. Sorry for my English."*
>
> Discord: Users were trading deepfake tips and how it would be a useful blackmail tool. One user said that they made a "pretty good" video of a girl they went to high school with, using around 380 pictures scraped from her Instagram and Facebook accounts. They joked this was made easier because all the Instagram accounts scraped used quality cameras.

5. What is being done to address this problem?

A. Law

California family court: California has a thorough framework in place for handling digital abuse cases, and intimate partner abuse victim enjoy particularly robust protection under California's Domestic Violence Prevention Act.

What about state and federal court? We agree. Face-swapped pornography is not just a domestic violence issue. A defendant could also be sued in state or federal court using a variety of legal theories, including copyright infringement (based on photos of the victim's face); defamation, false light and related claims (based on falsity and harm to reputation); violation of publicity rights (based on use of name/likeness without permission), and a variety of other claims. Defendants may have defenses if the face-swapped video is newsworthy, constitutes fine art, or is in certain other protected categories, but these are unlikely to arise in the context of most intimate partner disputes involving non-celebrities.

B. Market

In part, the job of policing this content has fallen on companies who host the servers and platforms where videos are viewed and hosted. Online platforms Reddit, Discord, and Pornhub

---

[30] *Id.*

[31] Samantha Cole, *People Are Using AI to Create Fake Porn of their Friends and Classmates*, MOTHERBOARD (Jan. 26, 2018), https://motherboard.vice.com/en_us/article/ev5eba/ai-fake-porn-of-friends-deepfakes.

for example, have announced that they will not allow deepfakes and have started deleting them from the platform.[32] News articles and public pressure influence companies to make these types of business decisions.

C. Technology

Solutions to identify and detect face-swapped videos are being discussed and, in some cases, utilized. For example, the website Gfyfcat (200 million daily users), is already using machine learning algorithms to detect such videos on its site.[33] This particular site requires the original video be available online, which wouldn't work if the video were only available on a private Facebook or Instagram account. However, algorithms to detect face-swapped videos may well evolve at a similar rate as the algorithms used to create them.

6. What can victims and advocates do?

Limit public access video and photographic material: To create a believable deepfake, a perpetrator requires video and/or hundreds of photographs of their target. "Open-source tools like Instagram Scraper and the Chrome extension DownAlbum make it easy to pull photos from publicly available Facebook or Instagram accounts and download them all onto your hard drive."[34] A Google image search using the target's name will also locate publically available photos. That data can then be used in connection with a deep learning algorithm to create a face-swapped video. Potential victims can proactively limit access by:

- Making all their social media accounts private.
- Ensuring that any friends or followers with public social media accounts (or who are friends with their abuser) take down photos or videos depicting them and/or make those accounts private.
- Conducting a Google search of their name to assess their video and image digital footprint and take steps to have the footage taken down or removed when possible.
    - Alternatively, one could use the aforementioned apps that scrape social media accounts for photos and videos to take inventory of their digital footprint.

Take Down Requests: If a face-swapped video is published, a victim can make a request to the platform that the offending content be removed. There is not a lot of experience with takedown requests directed to deepfakes, and many platforms do not yet have specific policies directed to them. It is possible that different approaches may work with different platforms. Review their terms and policies, and choose the best approach for each site. Reporting the content under the non-consensual pornography category may be a reasonable approach for sites that already provide a mechanism for that type of content.[35] If you can make a valid claim of copyright

---

[32] Adi Robertson *Reddit bans 'deepfakes' AI porn communities*, THE VERGE (Feb. 7, 2018), https://www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban.
[33] Louise Matsakis *Artificial Intelligence is Now Fighting Fake Porn*, WIRED (Feb. 14, 2018), https://www.wired.com/story/gfycat-artificial-intelligence-deepfakes/.
[34] Samantha Cole, *People Are Using AI to Create Fake Porn of their Friends and Classmates*, MOTHERBOARD (Jan. 26, 2018), https://motherboard.vice.com/en_us/article/ev5eba/ai-fake-porn-of-friends-deepfakes.
[35] *See* Cyber Civil Rights Initiative, *Online Removal Guide* (last visited Mar. 27, 2018), https://www.cybercivilrights.org/online-removal/.

ownership, a copyright takedown is another option. We recommend that you be transparent, honest and polite when requesting that a platform take down any type of content.

Domestic Violence Restraining Order: As illustrated in section two above, pursuant to the Domestic Violence Prevention Act, threats to publish, or actually publishing, a face-swapped video of an intimate partner can qualify as harassment, disturbing the peace and other forms of abuse that entitle the victim to restraining order protection.

Law Enforcement: While criminal laws do not specifically respond to deepfake domestic violence, victims should still contact law enforcement as certain penal code sections (extortion, identity theft, domestic violence, stalking, etc.) may apply and a police report may be helpful evidence in the future. Victims and advocates may find the following guide developed by Without My Consent a helpful place to start: withoutmyconsent.org/resources/download-guide. For example, read the chapter on Evidence Preservation, complete the Evidence Chart, and take that evidence binder to your local helpers, including law enforcement. The stronger your evidence and the more compelling your narrative, the more likely it is that law enforcement will be able to help you.

Civil Litigation: Face-swapped video can give rise to a variety of civil claims, including defamation, false light, copyright infringement, violation of privacy and/or publicity rights, and more. These claims may have statutes of limitation that can forever bar your ability to bring a claim if you delay. We recommend consulting a lawyer to determine your options.

Retain all photos and videos: Victims should retain copies of all original videos and photos in their possession or online as they can be later used to contrast against and contradict a face-swapped video.

New Red Flag – Unwanted Filming or Photos: An intimate partner who is suddenly interested in taking increased photos or video footage of a victim, particularly from multiple angles, may be a deepfake red flag.

Conclusion

California family court self-help staff, clinicians, advocates, and domestic violence restraining order attorneys: When you encounter one of these cases for the first time, please do reach out to the authors of this article. We would love to help connect you with teams who can help you lay the groundwork for a successful resolution.

As an advocate representing a victim with regard to fake video (or other falsified evidence-based abuse) you represent both the individual victim and the public interest in redressing injuries resulting from these malicious torts.

We have a team ready to support this endeavor.