
Non-consensual Images & Videos and Consent in Social Media

Jane Im

University of Michigan
Ann Arbor, U.S.
imjane@umich.edu

ABSTRACT

This is a position paper that questions the current culture of "consent" in social media. This question stems from the recent surge of problematic non-consensual images and videos spreading online including social media, such as creepshots and deepfakes. I first provide an overview of the increase in non-consensual images/videos in social media, as well as social media's current approaches to overcome them. Considering the current approaches' limitations of being post hoc, this paper suggests the need to explore and rethink about the platforms' designs regarding consent.

KEYWORDS

social media; consent; deepfake; creepshot

ACM Reference Format:

Jane Im. 2018. Non-consensual Images & Videos and Consent in Social Media . In *CHI '19: ACM CHI Conference on Human Factors in Computing Systems, May 04–09, 2019, Glasgow, UK*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '19, May 04–09, 2019, Glasgow, UK

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

MOTIVATION: PROBLEMATIC NON-CONSENSUAL IMAGES/VIDEOS IN SOCIAL MEDIA

Problematic non-consensual images and videos, especially of females, are increasing in social media platforms. Here I define problematic non-consensual images/videos as those that are uploaded online without the consent of the person in the picture and contain sensitive information of the person such as nudity. In this paper the term is used broadly enough to encompass *creepshots*, *deepfakes*, and *images/videos posted during/after sexual assaults*.

Creepshots Creepshots are known as images of women, taken without their consent. They are typically taken in public and close-up and revealing [3]. It is well known that Tumblr has many groups where people share creepshots [3]. Recently Tumblr announced the ban of "adult content, including explicit sexual content and nudity (with some exceptions)" ¹. Reddit had a popular subreddit called "r/Creepshots" which was later banned after much controversy in 2012 [1]. Regardless of the recent bans, it is alarmingly easy to upload creepshots online due to devices like mini spy cameras [17].

Deepfakes Deepfakes are artifacts of combining existing images/videos onto source images/videos using a machine learning technique called generative adversarial network (GAN) [14]. It is theoretically possible for deepfakes to be consensual if it involves subjects that know and agree on their faces being swapped. However, at least now many deepfakes seem non-consensual, mostly targeted towards creating altered porn. They are also dangerous because they easily manipulate reality and beliefs, as well as skew information [5]. Reddit had a subreddit called "r/deepfakes" which was the main hub for sharing deepfakes related to involuntary porn and had over 90,000 subscribers. The subreddit was later banned in 2018 [12].

Photo/video posted during/after sexual assaults The situation gets worse when the platforms are used to distribute such images/videos during or after a sexual assault [10]. Instagram was used in a case of sexual assault with the offender posting a photo of the victim [8].

IMPACT ON VICTIMS' LIVES

The impact of the images and videos are tremendous on the lives of the victims. Although there is not yet a paper that comprehensively shows the physical or psychological state of creepshots' victims, some of them have been reported to be afraid of being blamed and being looked upon as "cheap", not even opening up to their close ones about the experience [2, 17]. Victims have reported they feel they have no control over themselves after seeing the images and videos. There have been multiple cases where the victims committed suicides when the photos/videos are related to sexual assaults [4, 9].

CURRENT APPROACHES IN SOCIAL MEDIA PLATFORMS

There are various fields that are relevant to solving this problem of non-consensual images/videos, such as law and education. For instance, it is clear that we need education to foster a culture where

¹<https://staff.tumblr.com/post/180758987165/a-better-more-positive-tumblr>

people are aware that such acts are indeed harmful to others [10]. We also need to enforce law that prevent people from posting such images/videos [16]. In this paper however, I focus on the current approaches within social media platforms which are all *post hoc*. In the next section I suggest a way that is not post hoc as it aims to prevent such postings in the long run.

Banning Subcommunities One approach that the social platforms are using is banning the subcommunities where such postings are frequent. For instance, Reddit banned "r/deepfakes", a subreddit where AI-powered porn using celebrities' faces were rampant [12]. Reddit also has previously banned "r/Creepshots" [13]. One question is whether these bans would work in the long term since the banned subcommunities' active users may have moved to other platforms.

Deploying Machine Learning Models Another main technique is using machine learning models which many platforms including Tumblr and Facebook use [6, 15]. However, machine learning models are currently imperfect in cases such as when having to differentiate porn and nudity for art and politics [7].

Using Human Workers On the other hand, social platforms also hire human workers that go through images/videos and delete them manually. These workers are put into labor that require hours and hours of looking at psychologically damaging content [11].

RETHINKING ABOUT CONSENT IN SOCIAL MEDIA

As we have seen in the previous section, current approaches are post hoc and each has its own limitation. Suggesting a different direction, I would like to explore whether there is a way to redesign social platforms' features so that they can nudge people to rethink about non-consensual images/videos in the long run. That is, I argue that a more broader question should be asked about the *overall culture of consent in social media*, not just focusing on immediate post hoc steps such as banning subcommunities. For instance in Facebook, it is very easy to tag a person on a photo in the default settings (without asking him/her). A user has to actually change one's settings in order to *prevent* others from tagging oneself. If such designs were different in a way that *asks the user about other people's consent*, would people have thought about consent in social media any differently?

I am not arguing that this problem of non-consensual images/videos can be merely solved through redesign of social platforms. Since technology itself is not the only reason for such images/videos to be posted or shared, this problem requires a mix of various fields including law, politics, education, and technology in order for it to be solved. However, I believe there is a role for HCI and CSCW by exploring whether social platforms' design can be improved in a way to nudge people to rethink about posting non-consensual images/videos. As Powell has written, "*some of the additional harm of the Werribee DVD might have been prevented if young people were engaged as proactive and ethical bystanders*" [10].

FUTURE STEPS

This project is currently in a preliminary stage and I would like to receive feedback from fellow researchers that have approached sensitive topics. My expertise is in system-building and therefore would like to focus on coming up with insights on how to design and build social platforms in a way where the idea of consent is deeply embedded.

REFERENCES

- [1] Adrian Chen. 2012. Unmasking Reddit's Violentacrez, The Biggest Troll on the Web. Retrieved Feb 10, 2019 from <https://gawker.com/5950981/unmasking-reddits-violentacrez-the-biggest-troll-on-the-web>
- [2] Marcus Costello. 2018. Man photographed 'upskirting' likely still out there violating women. Retrieved Feb 12, 2019 from <https://www.sbs.com.au/news/the-feed/man-photographed-upskirting-likely-still-out-there-violating-women>
- [3] Joseph Cox. 2018. Tumblr Has a Massive Creepshots Problem. Retrieved Feb 10, 2019 from https://motherboard.vice.com/en_us/article/paxevg/tumblr-creepshots
- [4] Alexa Dodge. 2016. Digitizing rape culture: Online sexual violence and the power of the digital photograph. *Crime, media, culture* 12, 1 (2016), 65–82.
- [5] Douglas Harris. 2019. Deepfakes: False Pornography Is Here and the Law Cannot Protect You. *Duke Law & Technology Review* 17, 1 (2019), 99–127.
- [6] Daniel Lowd. 2018. Can Facebook use AI to fight online abuse? Retrieved Feb 12, 2019 from <https://theconversation.com/can-facebook-use-ai-to-fight-online-abuse-95203>
- [7] Rachel Metz. 2019. We tested Tumblr's ban on porn. It needs work. Retrieved Feb 10, 2019 from <https://www.cnn.com/2019/01/02/tech/ai-porn-moderation/index.html>
- [8] Anna North. 2017. 5 years ago, a girl was raped in Steubenville. Now her rapist's father shot a judge. Retrieved Feb 10, 2019 from <https://www.vox.com/identities/2017/8/22/16185174/steubenville-shooting-judge-nathaniel-richmond>
- [9] Kelly Oliver. 2016. Rape as spectator sport and creepshot entertainment: Social media and the valorization of lack of consent. ASJ.
- [10] Anastasia Powell. 2010. Configuring consent: Emerging technologies, unauthorized sexual images and sexual assault. *Australian & New Zealand journal of criminology* 43, 1 (2010), 76–90.
- [11] Sarah T Roberts. 2014. *Behind the screen: The hidden digital labor of commercial content moderation*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.
- [12] Adi Robertson. 2018. Reddit bans 'deepfakes' AI porn communities. Retrieved Feb 10, 2019 from <https://www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban>
- [13] Jessica Roy. 2012. Reddit Bans Creepshots and Creepsquad, the Subreddits That Fueled Its 'War' With Gawker. Retrieved Feb 10, 2019 from <https://observer.com/2012/10/reddit-bans-creepshots-and-creepsquad-the-subreddits-that-fueled-its-war-with-gawker>
- [14] Oscar Schwartz. 2018. You thought fake news was bad? Deep fakes are where truth goes to die. Retrieved Feb 10, 2019 from <https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>
- [15] Bijan Stephen. 2019. Porn: you know it when you see it, but can a computer? Retrieved Feb 12, 2019 from <https://www.theverge.com/2019/1/30/18202474/tumblr-porn-ai-nudity-artificial-intelligence-machine-learning>
- [16] Marc Tran. 2015. Combatting gender privilege and recognizing a woman's right to privacy in public spaces: Arguments to criminalize catcalling and creepshots. *Hastings Women's LJ* 26 (2015), 185.

- [17] Kelly Wynne. 2018. Toilets In Seoul Will Be Checked Every Day To Combat Spy-Cam Porn. Retrieved Feb 10, 2019 from <https://www.newsweek.com/toilets-seoul-will-be-checked-everyday-combat-spy-cam-porn-1101391>