Verifying images: deepfakes, control, and consent

Emily van der Nagel

Published online: 02 Jun 2020.

Submit your article to this journal ⬀

View related articles ⬀

View Crossmark data ⬀

Routledge
Taylor & Francis Group

Verifying images: deepfakes, control, and consent

Emily van der Nagel

School of Media, Film and Journalism, Monash University, Caulfield, Victoria, Australia

ABSTRACT

Deepfakes are portrayed as 'deceptive media' that amalgamate fact and fiction. But even when they do seem threatening to our narratives of the truth, there is still room for pushback to the problem of deepfakes thwarting control over images of women. In this article, I first examine the general phenomenon of deepfake porn as a misogynistic way to control women's images. I then present some lessons from photographic studies about manipulated images before turning to a case study that directly addresses image consent: the verification system used on the bulletin board thread Reddit Gonewild. I argue that the verification system is a way for women to reclaim their own image as new ways of digitally manipulating images and evidence emerge. By verifying their images, these women are also emphasizing the importance of consent in all forms of porn production and consumption.

Introduction

Since the emergence of deepfakes – face-swap videos produced with artificial intelligence (Paris and Donovan 2019) – the news media discourse has focused on three main themes pertaining to this technology: parodies, media manipulation or 'fake news' (Shao 2019), and pornography. In fact, deepfakes continue a long history of women's images being used to harass, humiliate, and harm them. While parody videos have become a way to showcase deepfake technology (one especially creative example depicts Rowan Atkinson's face moving uncannily in time with Charlize Theron's sensual performance in a perfume advertisement) and manipulated news media are perceived to be a credible threat to democracy, deepfake porn, a kind of image-based sexual abuse, is the most prominent way that this technology has been mobilized. According to a study by deepfake detection company Deeptrace Labs, 96% of deepfake videos are non-consensual porn, mostly of women celebrities (Patrini 2019).

Deepfakes are portrayed as 'deceptive media' (Paris and Donovan 2019) that amalgamate fact and fiction. The ease of creating a deepfake has perhaps been overestimated – journalist Kevin Roose (2018) describes a process that involves using the FakeApp software program and painstakingly taking several hundred photographs of his own face to use as a data source, only to discover that without experience in blending and tweaking a face to match a source video, the result appears glitchy. Commentators may also be

overestimating how often deepfakes are presented as truthful videos: what is spectacular about deepfakes is the illusion that there is an audience who routinely find them credible. According to Milena Popova (2019), many deepfakes are explicitly labelled fakes. This points at the effort expended to create the deepfake, and to contain the video within specific communities that are deliberately seeking out this kind of content, but it also reveals that people who want to see porn videos with celebrity faces superimposed on them are less concerned with the veracity, or the intimacy, of the performance than the control they have over their porn consumption.

Even when they do seem threatening to our narratives of the truth, there is still room for pushback to the problem of deepfakes thwarting control over images of women. In this article, I first examine the general phenomenon of deepfake porn as a misogynistic way to control women's images. I then present some lessons from photography studies about manipulated images before turning to a case study that directly addresses image consent: the verification system used on the bulletin board thread Reddit Gonewild. I argue that the verification system is a way for women to reclaim their own image as new ways of digitally manipulating images and evidence emerge. By verifying their images, these women are also emphasizing the importance of consent in all forms of porn production and consumption.

Deepfakes: a tool to desire and control women

The term 'deepfake' originated from the Reddit username of an account, and then a sub-reddit, dedicated to this kind of content (Burkell and Gosse 2019). It may be a play on the term 'DeepFace', the Facebook facial recognition software which developers claimed in 2014 was close to human-level performance at recognizing faces (Taigman et al. 2014). Deepfake porn is not confined to Reddit, and deepfakes are not inherently abusive. But it is important for this article that Reddit contains both the origin of the term and a potential way to combat its non-consensual aspects: the verification process that I detail later is an enduring feature of one of Reddit's most popular Not Safe For Work (NSFW) or adult content subreddits, Gonewild.

In journalist Christine Lagorio-Chafkin's (2018) book on Reddit, she interviews its founders, Steve Huffman and Alexis Ohanian, to discover that many of their own sensibilities have become foundational to the platform. According to Lagorio-Chafkin, these qualities include intelligence, generosity, a penchant for mischief, and a belief in social justice, free speech, and anonymity. The same qualities led Adrienne Massanari (2017) to call Reddit a 'toxic tech-noculture' with a platform architecture and a governance model that implicitly prioritize the views and desires of young, white, geeky men at the expense of others. As Lagorio-Chafkin puts it, there are competing narratives about Reddit, a source of whimsy and fun as well as hate and harassment. While a small subset of Redditors introduced and celebrated deepfake porn, Reddit's owners eventually made the decision to shut down the deepfakes subreddit and, among other platforms, has banned non-consensual deepfake porn:

> Reddit prohibits the dissemination of images or video depicting any person in a state of nudity or engaged in any act of sexual conduct apparently created or posted without their per-mission, *including depictions that have been faked*. (Reddit 2020; emphasis added)

At the same time, Reddit still hosts a number of other deepfake-oriented subreddits such as r/SFWdeepfakes, which focuses on parody videos, and r/MediaSynthesis, which

broadens the remit to include artificial intelligence-generated and manipulated content, guides on creating it, and image libraries to use as source material. Deepfakes, even deep-fake porn videos, are not violations by default. For example, it is not difficult to imagine enterprising porn performers using deepfake technology to allow their audience to buy porn with their own facial expressions set to an essentially customisable body shape and size. But it is also possible that the existence of consensual or parody videos may be used to belie the main use of deepfakes: as a technology for image-based sexual abuse that produces sexual content without the consent of the subject.

Image-based sexual abuse violates what Danielle Citron (2018) calls 'sexual privacy', causing deep and lasting harm for its subjects by debasing and humiliating them, reducing them to a set of body parts. For Stuart Hargreaves (2018), the main harm in a related practice that violates sexual privacy – 'creepshots', or photographs taken of unaware women in public – is the way it reproduces the male gaze. By treating women's faces as a digital resource to be edited onto sexual bodies by artificial intelligence, this reinforces the idea that women exist as sexual objects. The discussion around creepshots on platforms like Reddit rewards and celebrates the men who take them, which furthers toxic masculinity and creates a broader environment in which women's images are understood as consumable, malleable, and brought into being for the enjoyment and gratification of men. This view of images of women is not something created by deepfakes, but is at least as old as the technology of photography.

Truth, lies, and exploitation: lessons from photography studies

As Jessica Lake (2016) argues, the advent of photography in the 1830s altered the architecture and experience of seeing and being seen – especially for women, who were more likely than men to be photographed and exploited. Lake focuses on consent as she links photography of women in public to the development of privacy law, pinpointing a case from 1900 when a young woman in New York went to court after discovering, with shock and dismay, her own photograph used to advertise a brand of flour. In order to wrest control back from men with cameras, privacy law was developed as a way to assert the right to one's own image. Privacy may now be an established legal concept (Prosser 1960), but it has not stopped men from capturing and interfering with women's images as much as it has labelled this intrusion.

Nearly 90 years after the case of the young woman recognizing her own face on a bag of flour, Adobe Photoshop was introduced with an advertisement that depicted the founder's wife, Jennifer, sitting topless on a beach. Used to demonstrate the potential of digital technology to potential Adobe investors, 'Jennifer in Paradise' is a key image of the digital era, and one that serves as a reminder that women still exist as objects in visual culture, while men remain active subjects (Golding 2019). A manipulated photograph still evokes the 'presumption of veracity' that Susan Sontag ([1971] 2002) wrote about: all photographs have authority, interest, and seductiveness, as they imply a trace of the real, no matter how much they have been edited, and they are always constructions of reality, no matter if they have been presented as the truth.

Images reveal and conceal, argues social media theorist Nathan Jurgenson. In his book *The Social Photo*, Jurgenson (2019) reminds us that photographs are never neutral windows to the world, and cameras are not truth-telling machines. From photography scholarship, we can

conclude that images might not tell the truth, but they do tell us something: what is desirable, for example, or what is worth paying attention to. For example, a deepfake of actor Emma Watson's face set on top of a porn performer's body does not tell us that Emma Watson has performed in porn – but it does tell us that she is considered sexually desirable, and her image is available to the creator or audience of the deepfake. If, as John Berger (1977) put it, every image embodies a way of seeing, then deepfake nudes are a way of seeing women as objects, celebrities as commodified and consumable, and the digital as malleable (for the few that have access to the specific sets of software, skills, and image libraries). Deepfakes might not present accurately filmed porn to us, but the phenomenon tells us a story about power, control, and desire. As José van Dijck argues of digitally manipulated images: 'Our photographs tell us who we want to be' (2008, 70). If who we want to be is a consumer of ethically, consensually made porn, how can we ensure this in a landscape that includes digitally manipulated images? Reddit offered us the term 'deepfake', but on a different corner of the platform, a verification system is used to assure its audience that they are seeing photographs of people who want them posted there.

Consent in image production: Reddit Gonewild's verification system

As an exhibitionist subreddit, and an offshoot of the larger NSFW section of Reddit, Gonewild features mostly women in various stages of undress, from topless selfies to shots of spread open vulva. The subreddit is enormously popular, and photographs posted there are likely to be downloaded and shared, making for a vast potential audience. As most people on Reddit are pseudonymous, this means it would not be difficult to engage in image-based sexual abuse by posting a photograph of an ex-partner, or one taken surreptitiously. To combat posts like this, Gonewild has instituted a verification system. Unlike real-name platforms such as Facebook, Reddit does not ask for official identification. Instead, the system aims to link an account to the photographs posted. Verification involves submitting a photograph depicting the woman holding up a handwritten sign with her username, the date, and the name of the subreddit. Reddit Gonewild's Frequently Asked Questions page deliberately addresses the issue of digital image manipulation when it advises women to:

> crumple the sign up into a ball and then take your pictures with the sign uncrumpled! This creates a lot of random angles in the paper, and convinces the [moderators] and users that the sign was not photoshopped. (xs51 2019)

The technical idea is that flat surfaces are more vulnerable to being edited, but the social implication is that images are often taken without permission and modified in a way that does not fit the poster's original intentions. I argue elsewhere that Reddit Gonewild's specific form of embodied verification exists in opposition to official verification (van der Nagel 2020). Seeing bodies without names or faces attached assert their consent for the Gonewild audience to see their most intimate body parts means the identifiable individual is protected, while their body is linked with their Reddit account and therefore their permission for their images to circulate within that space.

Of course, the system is still vulnerable to abuse. Images posted after the verification shots can be downloaded and recirculated to other places on the internet. But especially within the platform, and the toxic technoculture, of Reddit – in which users may well browse Gonewild and also become interested in deepfakes – verification calls attention

to consent as a key part of sexual encounters, in person or on a social media platform. As verification posts appear as part of the rolling front page, publicly available to the audience rather than privately submitted to moderators, seeing verification posts is part of the Gonewild experience. It reminds the viewer that even when they are faceless and pseudonymous, these are real women insisting upon agency for their own image.

While Reddit Gonewild's verification system does not directly address deepfake porn, it contributes to a culture of consent. While some porn consumers are more intent on seeing the face of a celebrity or known woman they desire in a simulacrum of sexual pleasure – or at least sexual activity – Alan McKee's (2006) work interviewing porn consumers reveals that many appreciate the idea of consensually produced material. Of the 46 Australian porn consumers he interviewed, a key issue they spontaneously raised was that they liked seeing porn performers who looked like they were 'into it'. Focusing on verification as a key part of the process of consuming and sharing digital content can lead to larger cultures of consent. There are various ways to prove the origin of a digital object, such as embedded metadata or a watermark on the image, but Reddit Gonewild's verification system deliberately communicates: 'I took this to be shared in this context'. Verification more broadly provides another way to feel confident that a photograph or video has been published intentionally: if it comes from a social media account with ties to the people in the image or who produced the original version. Accounts can be compromised. But if, for example, a porn video is tweeted out by the porn performer featured, it is more likely to be content that they are happy to share.

Images can be manipulated, and within a culture of objectifying women, some audiences may not value authenticity as much as sexual gratification. But if Reddit can incubate deepfake porn and verified porn, this gives evidence that there is room for discourses and practices of consent – even within a highly gendered subreddit that presumes 'going wild' is the terrain of (mostly) young, white, slender women.

Conclusion

Photographs always tell partial truths. From selecting some objects to appear in the frame while leaving others out to harnessing artificial intelligence to identify and replace faces in porn videos, the editing of images, and even the idea of owning one's own image, has a history entwined with the manipulation and control of women. Paying attention to practices of consent, such as Reddit Gonewild's verification system, is an important way to make visible cultures that emphasize women's agency over their bodies and images.

As an emerging phenomenon, there is still much more research to be done on deepfake porn. Understanding its origins, prevalence, impact, subjects, and creators are all important ways of forming critical perspectives and, following McKee (2006), there is also a particular need to hear the insights of consumers of this kind of media. This is not because the voices of perpetrators of abuse are more important than their victims. Instead, situated in a broader culture of patriarchal heteronormativity, interrogating the demand for, and practice of, desired faces stitched on to available bodies is crucial for our continuing discussions of deepfake porn, and image-based sexual abuse more broadly.

ORCID

Emily van der Nagel 🔘 http://orcid.org/0000-0003-2994-4542

References

Berger, John. 1977. *Ways of Seeing*. London: Penguin.

Burkell, Jacquelyn and Chandell Gosse. 2019. 'Nothing New Here: Emphasizing the Social and Cultural Context of Deepfakes.' *First Monday* 24 (12). https://firstmonday.org/ojs/index.php/fm/article/view/10287/8297.

Citron, Danielle K. 2018. 'Sexual Privacy.' *Yale Law Journal* 128: 1870–1960.

Golding, Dan. 2019. 'Far From Paradise: The Body, the Apparatus and the Image of Contemporary Virtual Reality.' *Convergence: The International Journal of Research Into New Media Technologies* 25 (2): 340–353.

Hargreaves, Stuart. 2018. '"I'm a Creep, I'm a Weirdo": Street Photography in the Service of the Male Gaze.' In *Surveillance, Privacy and Public Space*, edited by Bryce Clayton Newell, Tjerk Timan and Bert-Jaap Koops, 179–198. London: Routledge.

Jurgenson, Nathan. 2019. *The Social Photo: On Photography and Social Media*. London: Verso.

Lagorio-Chafkin, Christine. 2018. *We Are the Nerds: The Birth and Tumultuous Life of Reddit, the Internet's Culture Laboratory*. London: Piatkus.

Lake, Jessica. 2016. *The Face That Launched a Thousand Lawsuits: The American Woman Who Forged a Right to Privacy*. New Haven: Yale University Press.

Massanari, Adrienne. 2017. '#Gamergate and the Fappening: How Reddit's Algorithm, Governance, and Culture Support Toxic Technocultures.' *New Media & Society* 19 (3): 329–346.

McKee, Alan. 2006. 'The Aesthetics of Pornography: The Insights of Consumers.' *Continuum: Journal of Media & Cultural Studies* 20 (4): 523–539.

Paris, Britt and Joan Donovan. 2019. 'Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence.' *Data & Society*. September 18. Accessed March 7, 2020. https://datasociety.net/output/deepfakes-and-cheap-fakes/.

Patrini, Giorgio. 2019. 'Mapping the Deepfake Landscape.' *Deepfake Labs*. October 7. Accessed March 7, 2020. https://deeptracelabs.com/mapping-the-deepfake-landscape/.

Popova, Milena. 2019. 'Reading Out of Context: Pornographic Deepfakes, Celebrity and Intimacy.' *Porn Studies*. doi:10.1080/23268743.2019.1675090.

Prosser, William L. 1960. 'Privacy.' *California Law Review* 48 (3): 383–424.

Reddit. 2020. 'Do Not Post Involuntary Pornography.' *Reddit Help*. Accessed March 7, 2020. https://www.reddithelp.com/en/categories/rules-reporting/account-and-community-restrictions/do-not-post-involuntary-pornography.

Roose, Kevin. 2018. 'Here Come the Fake Videos, Too.' *The New York Times*. March 4. Accessed March 7, 2020. https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html.f.

Shao, Grace. 2019. 'Fake Videos Could Be the Next Big Problem in the 2020 Elections.' *CNBC*. October 15. Accessed March 7, 2020. https://www.cnbc.com/2019/10/15/deepfakes-could-be-problem-for-the-2020-election.html.

Sontag, Susan. [1971] 2002. *On Photography*. London: Penguin.

Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato and Lior Wolf. 2014. 'DeepFace: Closing the Gap to Human-Level Performance in Face Verification.' *Facebook Research*, June 24. Accessed March 7, 2020. https://research.fb.com/publications/deepface-closing-the-gap-to-human-level-performance-in-face-verification/.

van der Nagel, Emily. 2020. 'Embodied Verification: Linking Identities and Bodies on NSFW Reddit.' In *Mediated Interfaces: The Body on Social Media*, edited by Carolina Cambre, Crystal Abidin and Katie Warfield. New York: Bloomsbury.

van Dijck, José. 2008. 'Digital Photography: Communication, Identity, Memory.' *Visual Communication* 7 (1): 57–76.

xs51. 2019. 'The Gonewild FAQ.' *Reddit*. Accessed March 7, 2020. https://www.reddit.com/r/gonewild/wiki/faq.