

DEEPTRUST ALLIANCE



THE STATE OF DEEPFAKE PORN

PRESENTED BY DEEPTRUST ALLIANCE

CONTENTS



02

WORD FROM
THE CEO

03

THE DARK SIDE OF
DEEPPAKES

04

WHY IS THIS A
PROBLEM?

06

REGULATION +
LEGISLATION

07

PLATFORMS +
DISTRIBUTION

09

OUR SOLUTION

10

REFERENCES

A WORD FROM THE CEO

Deepfakes burst into the collective imagination in 2018, portending serious consequences for our society. Clearly, the dangers of deepfakes must be confronted – and speed is paramount.

We founded DeepTrust Alliance to help business, government and civil society adopt breakthrough technologies to mitigate the impact of the growing cybersecurity threat to virtually any internet transaction. As a community of technologists and technology-optimists, DeepTrust Alliance is fundamentally hopeful about the potential of AI technologies to transform and empower human communication in positive ways.

However, we view the unchecked development and dissemination of deepfakes, particularly pornographic ones, as an insidious enemy to women's rights national security, and the digital economy,

DeepTrust Alliance is committed to raising awareness, convening stakeholders, building open solutions and driving comprehensive adoption to counteract the dangerous and disruptive emergence of an all too powerful global threat.

Kathryn Harrison

KATHRYN HARRISON
FOUNDER AND CEO





THE DARK SIDE OF DEEPAKES

WHAT WOULD YOU DO IF...

the above image suddenly popped up on your feed, but the woman pictured was your mother, wife, sister, or daughter? Pornographic deepfakes are a new form of misogynistic gender-based violence, and are by far the largest issue that deepfakes present in our world today.

Deepfakes, or manipulated media created using powerful AI technology, have captured the public's imagination. Yet, much of the fear around their use focuses on hypothetical applications in election or news intervention. Although it's certainly true that deepfakes already cause widespread harm, the victims are rarely governments; rather, they're women.

96% of deepfakes available online are pornographic, and often feature actresses and musicians [1]. But as the technology improves and becomes increasingly widely available, the threat to everyday women grows.

"Nothing can stop someone from cutting and pasting my image or anyone else's onto a different body and making it look as eerily realistic as desired." [6]

-Scarlett Johansson

Anyone with even a single picture online can become a victim of non-consensual deepfake porn. Regardless of who you are, the practice is humiliating and exploitative; Kristen Bell and Scarlett Johansson are victims of deepfake porn, and have expressed this in their work to raise awareness of the issue.

WHY IS THIS A PROBLEM?

Many creators and distributors of deepfake porn believe it's a victimless crime because the original source content was uploaded intentionally and widely available. This could not be further from the truth.

First, deepfake porn is created with malicious or harmful intent to shame, embarrass, or psychologically impact women. Second, pornography actresses' work is used without consent or compensation. Finally, digital pornography has wide distribution and persistence, so it cannot be easily shut down or eliminated. This problem is only expected to get worse as the technology to build deepfakes becomes more accessible, cheaper, and harder to detect.

Women, minorities, and other under-represented groups tend to be the initial focus of malicious fakes, but there is also great risk that this technology will threaten those in politics as well.

Let's return to the opening question for a moment. Imagine if someone you loved became a victim of deepfake pornography. How would you react?

You would, of course, immediately attempt to remove it and seek justice. However, as we will explain, you'd likely find that platforms and the law fail to offer your loved one complete — or even reasonable — protection from this harm.



EXTORTION + HARASSMENT

Rana Ayyub, a journalist and outspoken critic of the Hindu Nationalist movement in India, had her likeness incorporated into a porn video using deepfake technology. The video was distributed widely across India through Facebook and WhatsApp groups, leading to death threats. Local law enforcement refused to assist until the United Nations intervened.

96%

of deepfakes are pornographic [1]

100%

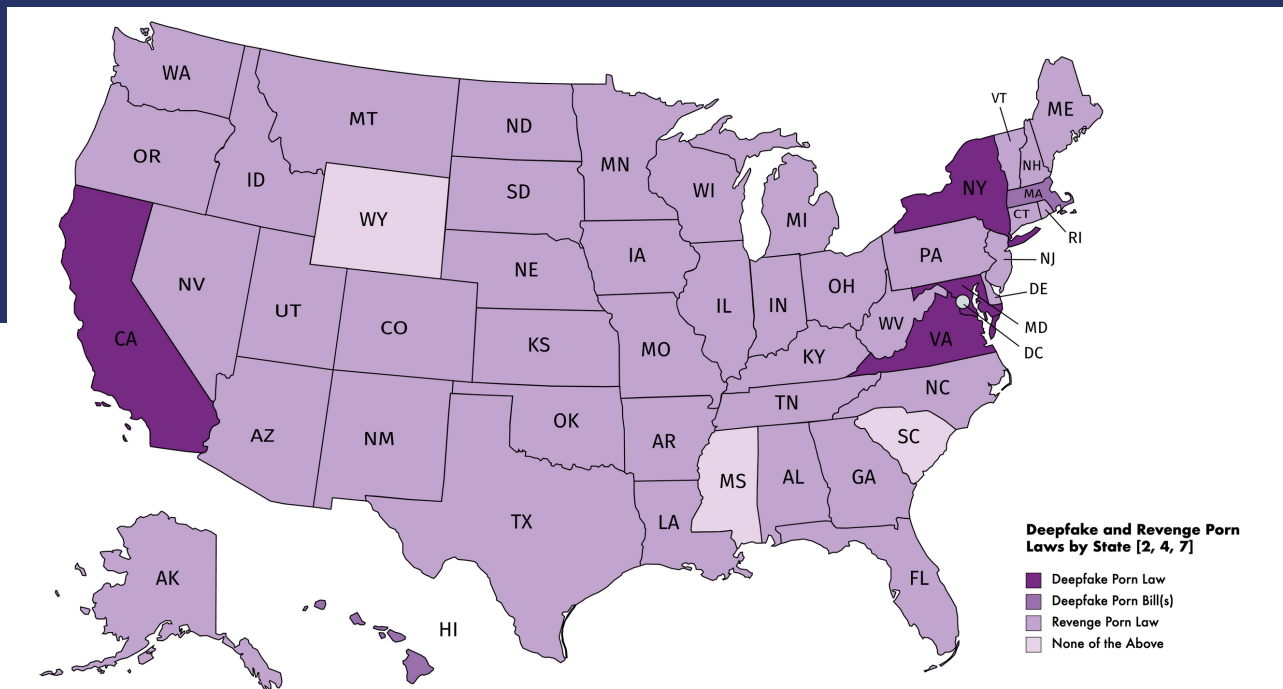
target women [1]

REGULATION + LEGISLATION

What effective legal and policy-based solutions exist to address the threat of malicious deepfakes? The answer may disappoint you.

The United States’ legal and regulatory landscape remains largely unwritten. Several laws focus on the emerging threat of deepfakes as political weapons, and ignore the current impact of pornographic deepfakes. Just four states have laws criminalizing deepfake porn: Virginia (Va. Code Ann. § 18.2-386.2), California (Cal. Civ. Code § 1708.86), New York (N.Y. Civ. Rights § 52-c(1)(a) as amended by S5959D) and Maryland (Md. Crim. L. § 11-208), which recently expanded its child porn law to encompass computer-generated images that are “indistinguishable from an actual and identifiable child,” (M. Ferraro, personal communication, November 19, 2020) [4]. No federal law exists at this time.

Therefore, in most states, victims must seek justice under the complicated and often inadequate patchwork of laws that define related crimes, such as revenge porn, copyright violations, and defamation [2]. This process is lengthy, expensive, and ultimately not guaranteed to apply to deepfakes. Even if such cases are won, they can’t reverse the humiliation or stop future replication and distribution of the video. There are no perfect solutions or protections available to the average citizen.



PLATFORMS + DISTRIBUTION

With little governmental or legal pressure to change, platforms have failed to act to protect women from being victims of non-consensual deepfake porn.

The worst offenders are dedicated deepfake porn websites, followed by peer-to-peer messaging apps, such as Telegram, which use hidden copies of deepfake nude apps to create porn for users to purchase. This activity can be very difficult to control as it moves between groups and under different names. Harms are even further amplified when these images and videos make rounds on large social media platforms such as Reddit and Twitter.

Much like the government, mainstream platforms are still determining how to handle deepfakes. Deepfakes alone do not violate most use policies, but explicit content does, making deepfake porn subject to removal from such sites. However, given the widespread availability of this content, problems persist in the enforcement of use policies, particularly in the discovery, reporting, and removal of content. As platforms struggle to decide how to move forward with moderating deepfakes, this presents an important opportunity to shape the culture of norms and ethics around their creation and sharing.

Additionally, even if platforms aren't directly responsible for the generation of content, they should have some incentive to control it. Many Instagram and TikTok influencers are underage, and a recent investigation revealed over two dozen examples of prominent TikTok stars featured in deepfake porn [3].

It's not too late to turn this around. The threat of malicious political deepfakes has already caught the Senate's attention as shown by the recent unanimous approval of the Identifying Outputs of General Adversarial Networks (IOGAN) Act [5] last year. We've also seen social media platforms rapidly develop effective new strategies to address election misinformation. The next step is to turn our attention to the existing threat of pornographic deepfakes, because there's much more to be done.



134M

views across the top four
dedicated deepfake porn
websites [1]

14K

videos of deepfake porn are
available online today [1]

94%

are on dedicated deepfake porn
websites [1]

6%

are on mainstream porn
websites [1]

OUR SOLUTION

Complex problems require collaborative solutions, which is why it's time to bring together important stakeholders and create industry-wide technology standards, best practices, and tools capable of addressing the problem of deepfake porn. No single entity has total control of digital content, and rivals must work together to build the human and technical interventions required to counter nefarious online actors. Through the collaboration of experts across fields, from computer vision, to AI, to forensics, and those focused on the human-side of technology use (legislators, educators, activists, media literacy groups) we can win this digital arms race. The possibilities are endless, but we imagine this could take the form of an oath for technologists, tech standards for forensic research, expectations that companies must meet to avoid contributing to the problem, and legislation that dismantles the financial infrastructure that funds the distribution of non-consensual deepfake porn.

Addressing deepfake porn not only protects women, but it opens up worlds of creativity. A few beneficial applications of deepfakes include in art, museums, Hollywood, education, media literacy, advocacy, and entertainment.

"I wish that the internet were a little bit more responsible and a little bit kinder." [6]
-Kristen Bell

DeepTrust Alliance is the only global non-profit confronting the threat of malicious deepfakes, particularly deepfake porn, through the development of best practices, policy, and ethics. Because no single person, company, or country can overcome the threat of deepfakes alone, we convene a network of tech, government, security and industry organizations working to overcome the challenges of deepfakes. Just as we have worked to confront the problems of misinformation, disinformation, and manipulated media, we turn to the problem of deepfake porn, which, until now, no organization has focused on exclusively. Join us as we bridge the gap between technologists, policy makers, and business leaders to solve the largest problem of deepfakes today, or donate to help us build the playbooks to stop this kind of harm and ensure the technology is developed and distributed in a way that does NOT harm women.

REFERENCES



1. The State of Deepfakes: Landscape, Threats, and Impact, Henry Ajder, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen, September 2019.
2. 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/cgi/viewcontent.cgi?article=5640&context=flr>
3. Dickson, E. J. (2020, October 26). TikTok Stars Are Being Turned Into Deepfake Porn Without Their Consent. *RollingStone*. <https://www.rollingstone.com/culture/culture-features/tiktok-creators-deepfake-pornography-discord-pornhub-1078859/>
4. Ferraro, M., & Tompros, L. (2020, December 17). New York's Right to Publicity and Deepfakes Law Breaks New Ground [Legislative Alert]. *WilmerHale*.
<https://www.wilmerhale.com/en/insights/client-alerts/20201217-new-yorks-right-to-publicity-and-deepfakes-law-breaks-new-ground>
5. Muncaster, P. US Senate Approves New Deepfake Bill. *Infosecurity Magazine*. <https://www.infosecurity-magazine.com/news/us-senate-approves-new-deepfake/>
6. Willen, C. (2020, June 11). Kristen Bell says she was 'shocked' to learn that her face was used in a pornographic deepfake video. *Insider*. <https://www.insider.com/kristen-bell-face-pornographic-deepfake-video-response-2020-6>
7. Ballotpedia. (n.d.). Nonconsensual pornography (revenge porn) laws in the United States. In G. Pallay (Ed.), *Ballotpedia*. Retrieved March 8, 2021, from [https://ballotpedia.org/Nonconsensual_pornography_\(revenge_porn\)_laws_in_the_United_States](https://ballotpedia.org/Nonconsensual_pornography_(revenge_porn)_laws_in_the_United_States)

ABOUT DEEPTRUST ALLIANCE

The Deeptrust Alliance is a nonprofit membership organization bringing breakthrough technology to fight the malicious use of deepfakes and disinformation.

Thanks and credit for this report goes to authors Kathryn Harrison and Amelia Leopold, with contributions from David Toomey, Sara Aros, Aviv Ovayda and Jane Lyons.