

# Congress Should Not Rush to Regulate Deepfakes

BY **HAYLEY TSUKAYAMA**, **INDIA MCKINNEY**, AND JAMIE WILLIAMS  
JUNE 24, 2019

The House Intelligence Committee held a [hearing](#) earlier this month examining the issue of “deepfakes,” a term coined to describe [images](#) or videos created with a machine learning algorithm that allows people to make false footage that appears real. There is real potential for fake or manipulated images or video to be dangerous or harmful. University of Maryland law school professor Danielle Citron pointed during her hearing testimony to the [horrific story](#) of journalist Rana Ayyub. An online mob spread a false pornographic video featuring Ayyub’s image, forcing her to hide for her own safety. As a society, we must acknowledge the harmful uses of deepfakes and hold the people who produce them accountable for their actions. EFF has [acknowledged the harms of](#) online harassment—including how people use harassment to chill the speech of marginalized people. Yet Congress must tread carefully if it seeks to address the actual problem without censoring lawful and socially valuable speech—such as parodies and satires.

Before Congress drafts legislation to regulate deepfakes, lawmakers should carefully consider what types of content new laws should address, what our current laws already do, and how further legislation will affect free speech and free expression.

## Understanding What Deepfakes Are (and Aren’t)

[Deepfake](#) is a portmanteau of “deep learning” and “fake,” and refers to images or video generated using machine-learning techniques that combine existing images and videos onto source images or videos. The term has been used to describe malicious online content, but technically speaking, the generative techniques underlying

deepfakes can be used to create parodies and satires, in addition to content intended to defame or humiliate.

Deepfakes drew media attention because they can be made with fairly accessible tools at home, marking a shift from a time when creating convincing fake footage required blockbuster special-effects budgets. As [Sarah Cole](#) from Motherboard reported, a common use of deepfakes technology is to create pornographic videos that convincingly spliced the face of one real person (such as a celebrity) onto the body of another—as happened to Ayyub. Deepfake videos can also depict people “saying” things they have not said, such as when Jordan Peele used machine learning to make a video of [Barack Obama](#) delivering a PSA about fake news.

The term has also been used misleadingly to describe other types of altered video, even those without artificial intelligence or machine learning involved. For example, lawmakers at the recent hearing frequently mentioned a widely circulated [video of Speaker Nancy Pelosi](#), which was slowed-down to make her sound drunk and garbled. Lawmakers also mentioned a number of other types of “false” media they wanted to target, including doctored photos, news articles, footage similar to that from *Forrest Gump* (in which Tom Hanks was [spliced into historical footage](#)), and [Milli Vanilli’s lip-syncing](#).

Congress’s confusion over the term “deepfake” is evident in Rep. Yvette Clarke’s new bill, introduced in advance of the hearing, “[The Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act](#),” or the DEEPFAKES Accountability Act. The bill requires mandatory labeling, watermarking, or audio disclosures for all “advanced technological false personation records.” This is defined in the bill as any media that falsely appears to depict speech or conduct of any person engaged in “material activity,” created via any technical means, that a reasonable person would believe to be authentic, and that was created without the consent of the person depicted. Material activity is defined to mean any speech,

conduct, or depiction that “a reasonable person would recognize has a tendency to cause perceptible individual or societal harm.”

There are a few problems with this bill, in addition to its overbroad definition of “deepfakes.”

First, it’s unclear how mandatory labeling and watermarking will solve the real harms that malicious deepfakes are causing. The trolls of the world will likely just not comply, particularly if they don’t live in the United States.

Second, the bill’s breadth and penalties trigger many First Amendment problems. For example, while there is an exception for parodies, satires, and entertainment—so long as a reasonable person would not mistake the “falsified material activity” as authentic—the bill fails to specify who has the burden of proof, which could lead to a chilling effect for creators. And in addition to civil penalties of up to \$150,000 for failure to include a watermark or disclosure, the bill imposes criminal penalties—even without any showing of harm—for violations intended not only to harass, incite violence, interfere in an election, or perpetuate fraud, but also to “humiliate” the person depicted, a vague term which the bill does not define. The First Amendment generally bars criminal laws that impose penalties without any showing of harm.

What’s more, the bill creepily exempts officers and employees of the United States acting in furtherance of public safety or national security.

The Clarke bill underscores a key question that must be answered: from a legal and legislative perspective, what is the difference between a malicious “deepfakes” video and satire, parody, or entertainment? What lawmakers have discussed so far shows they do not know how to make these distinctions. Those concerns deepen when

considering how many lawmakers and experts have suggested addressing the issue by altering the country's most important laws protecting Internet speech.

From a legal and legislative perspective, what is the difference between a malicious “deepfakes” video and satire, parody, or entertainment?

## **Understanding What 230 Is (and Ain't)**

During the hearing, policymakers suggested that limiting the protections provided by Section 230 (47 U.S.C. § 230) would solve all deepfakes problems. Such discussion misrepresents the scope of Section 230's protection, and minimizes the harm that further narrowing the statute would cause individual people. Limiting the most important law for protecting user speech will not solve the nuanced problems that deepfakes present.

[Section 230](#) protects providers and users of “interactive computer services” that republish content created by someone else from being held liable for that third-party speech. For example, social media platforms have protection against lawsuits based on decisions to moderate third-party content, or decisions to transmit content without moderation. Individuals who forward emails, or otherwise transmit content created by others, enjoy the same protections.

Unlike what some people—and [even some politicians](#)—seem to believe, Section 230 does not provide total immunity to the companies or individuals it covers. They can still be held accountable for their own speech. It only protects them from being held liable for things their users say online.

Section 230 also does [not just protect big companies](#). In fact, today, it is especially important for *small* companies without the resources to defend against expensive lawsuits based on speech of their users. The legal protection provided by Section 230 has spurred innovation that has led to the development of a rich and vibrant variety of open platforms that support all kinds of speech: from the important to the mundane, from the mainstream to the controversial, from the popular to the niche. These platforms facilitate the broad and rapid sharing of information, opinions, and ideas critical to a democratic society. The world has changed in many ways, but the online innovation and speech Section 230 enables is more important, not less, than it was two decades ago. New companies must share these same protections to grow and provide the competition that has made the Internet what it is today.

We've already seen how carving exemptions into Section 230, even when lawmakers believe they are narrowly tailored, can go awry. The Allow States and Victims to Fight Online Sex Trafficking Act of 2017, or [FOSTA](#), was passed by Congress for the worthy purpose of fighting sex trafficking. But the law contains language that [criminalizes the protected speech](#) of those who advocate for and provide resources to adult, consensual sex workers. Worse yet, the law actually [hinders efforts](#) to prosecute sex traffickers and aid victims. Given FOSTA's effects, it is reasonable to expect further unintended, but damaging, outcomes from alterations—particularly when attempting to draft legislation addressing something as subjective as the ways platforms should regulate speech.

Without 230 protection, platforms both large and small will be less open for speech of ordinary Internet users. Altering Section 230's language to increase liability for harmful deepfakes will not only sweep up important contributions to the public discourse, like parodies and satires, but it will also implicate a range of other forms of lawful and socially beneficial speech, as platforms censor more and more user speech to avoid any risk of legal liability.

## Why It's Important to Protect Parodies (And Other Speech That Isn't "True")

During the June 7<sup>th</sup> hearing, many members of Congress spoke in the same breath about a deepfakes video featuring Saturday Night Live cast members and propagandists faking information from a government, as if they were equivalent uses of deepfakes technology. But it's important that any regulations target only malicious, harmful deepfakes. We appreciate comments from Reps. Jim Himes and Brad Wenstrup that acknowledge the risk overbroad regulations pose to parody, satire, political commentary, and other socially valuable uses deepfakes technology.

Recently, we saw the important role that satirical deepfakes can play when the very issue of moderating deepfakes was satirized with a deepfake video. Israeli startup Canny AI made a video featuring footage from 2007, altered to depict Facebook chief executive Mark Zuckerberg's [bragging about abusing stolen data](#). The goal? To spark discussion about the platform's policies, by goading Facebook into making a moderation decision about whether to keep it up—as it had with the Pelosi video—or take it down. (Facebook said it would not take it down.)

This is exactly the sort of discourse that should be allowed to take place on platforms, without fear of censorship. Parody and satire are crucial tools to challenge powerful institutions—including Facebook—because of the momentary confusion that arises precisely because something is so outrageous and yet within the realm of possibility.

It can also be notoriously difficult, particularly for computers, to correctly distinguish satires and parodies from malicious content.

It can also be notoriously difficult, particularly for computers, to correctly distinguish satires and parodies from malicious content. This makes it especially [challenging](#) for companies to consistently and clearly enforce their policies. Parodies and satires are already often silenced in error, and such takedowns will only increase if laws are passed that hold platforms liable for failing to takedown deepfakes content.

There's no doubt that deepfake videos have been used in horrible and harmful ways, specifically against marginalized communities. While existing laws provide options for those who suffer harm from online harassment or fraud and to address injuries caused by [tortious uses of this technology](#), the impulse for legislators to “do more” to prevent malicious attacks is understandable. But mandatory watermarks or imposing intermediary liability for third party content is not guaranteed to stop them. Such measures are almost certain to sweep up protected speech as well.

To be clear, companies are allowed to, and should be able to, [moderate content](#) on their platforms. But, the government should not be telling them how to do it. We implore Congress to realize that broadly shifting the liability structure—or increasing the liability that platforms face based on content that they didn't create—leads to de facto government regulation of speech.

*Correction: An earlier version of this post contained a reference to generative adversarial networks. That has been removed to reflect that there are multiple ways to create deepfakes.*

## RELATED UPDATES

**This Texas Bill Would Systematically Silence Anyone Who Dares to Talk About Abortion Pills**

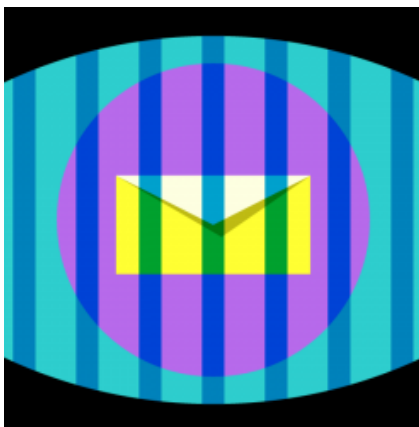


**DEEPLINKS BLOG** BY JENNIFER PINSOFF | MARCH 13, 2023



**DEEPLINKS BLOG** BY JASON KELLEY | MARCH 9, 2023

## **Utah's Governor Should Veto "Social Media Regulations" Bill S.B. 152**



**LEGAL CASE**

## **A.B.O Comix, et al. v. San Mateo County**

**EFF Tells Supreme Court: Trademark Law Doesn't Trump the First Amendment**





**DEEPLINKS BLOG** BY CARA GAGLIANO | MARCH 7, 2023



**LEGAL CASE**

## **Counterman v. Colorado**

**DEEPLINKS BLOG** BY DAVID GREENE | MARCH 2, 2023

## **EFF and Student Press Law Center Urge Supreme Court to Require Government to Show Subjective Intent in Threat Cases**

**EFF Files Amicus Brief to Protect the Speech Rights of Immigrants and Immigrant Rights Advocates**



**LEGAL CASE**

**United States v. Helaman Hansen**



DEEPLINKS BLOG BY JOSH RICHMAN | FEBRUARY 16, 2023

**Section 230 is On Trial. Here's What You Need to Know.**

**Participation in the Fediverse**

