

Taylor Swift deepfakes: New technologies have long been weaponized against women. The solution involves everyone

February 1 2024, by Nicola Henry and Alice Witt



Credit: Markus Winkler from Pexels

Sexually graphic ["deepfake" images of Taylor Swift](#) went viral on social media last week, fueling widespread condemnation from Swifties, the

general public and even the White House.

This problem isn't new. Swift is one of many celebrities and public figures, [mainly women](#), who have fallen victim to deepfake pornography in recent years. High-profile examples garner significant media attention, but the increasingly sophisticated nature of AI means [anyone can now be targeted](#).

While there are grave concerns about the broader implications of deepfakes, it's important to remember the technology isn't the *cause* of abuse. It's just another tool used to enact it.

Deepfakes and other digitally manipulated media

The sexually explicit deepfakes of Swift appeared on multiple [social media platforms](#) last week, including X (formerly Twitter), Instagram, Facebook and Reddit.

Most major platforms have bans on sharing synthetic and digitally manipulated media that cause harm, confusion or deception, including deepfake porn. This includes images created through simpler means such as photo-editing software. Nonetheless, one deepfake depicting Swift [was viewed](#) 47 million times over a 17-hour period before it was removed from X.

There's a [long history](#) of digital technologies, apps and services being used to facilitate gender-based violence, including [sexual harassment](#), [sexual assault](#), domestic or family violence, dating abuse, stalking and monitoring, and hate speech.

As such, our focus should also be on addressing the problematic gender norms and beliefs that often underpin these types of abuse.

The emergence of deepfakes

The origins of deepfakes can be traced to November 2017 [when a Reddit user called "deepfakes"](#) created a forum and video-editing software that allowed users to train their computers to swap the faces of porn actors with the faces of celebrities.

Since then, there's been a massive expansion of dedicated deepfake websites and threads, as well as apps that can create customized deepfakes for free or for a fee.

In the past, creating a convincing deepfake often required extensive time and expertise, a powerful computer and access to multiple images of the person being targeted. Today, almost anyone can make a deepfake—sometimes in a matter of seconds.

The harms of deepfake porn

Not all applications of AI-generated imagery are harmful. You might have seen funny viral deepfakes such as the images of [Pope Francis in a puffer jacket](#). Or if you watch the latest Indiana Jones film, you'll see Harrison Ford "de-aged" by 40 years thanks to AI.

That said, deepfakes are often created for malicious purposes, including disinformation, cyberbullying, child [sexual abuse](#), sexual extortion and other forms of image-based sexual abuse.

A [report published](#) by startup Home Security Heroes estimated there were 95,820 deepfake videos online in 2023, a 550% increase since 2019.

When it comes to deepfake porn, women in particular are

disproportionately targeted. [According to DeepTrace](#), 96% of all deepfakes online [are non-consensual fake videos of women](#). These are mostly (but not exclusively) well-known actors and musicians.

This is concerning but not surprising. [Research](#) shows online abuse disproportionately affects women and girls, particularly Indigenous women, women from migrant backgrounds and lesbian, gay, bisexual, transgender and intersex people.

Public figures in particular face [higher rates of online abuse](#), especially women and gender-diverse people. One study found celebrities [are attributed more blame](#) than non-celebrities for the abuse they receive online, and this abuse is often viewed as less serious.

Research shows image-based abuse can result in [significant harms](#) for victims, including anxiety, depression, suicidal ideation, social isolation and reputational damage. For [public figures](#), deepfakes and other forms of online abuse can similarly result in diminished career prospects, withdrawal from public life and negative mental health outcomes.

In 2016, Australian activist and law reform campaigner [Noelle Martin's](#) photos were taken from social media and superimposed onto pornographic images. Martin reported feeling "physically sick, disgusted, angry, degraded, dehumanized" as a result. Digitally altered and deepfake images of Martin continue to circulate online without her consent.

Responding to deepfake porn

Anyone can be targeted through deepfakes. All that's needed is an image of someone's face. Even professional work images can be used.

Although [law reform](#) alone won't solve this socio-legal problem, it can

signal the issue is being taken seriously. We need laws specifically targeting non-consensual deepfake porn.

In Australia, there are [image-based sexual abuse offenses](#) in every Australian state and territory except Tasmania, as well as at the federal level. However, only some laws specifically mention digitally altered images (including deepfakes).

Technology companies could also do a lot more to proactively detect and moderate deepfake porn. They need to prioritize embedding "[safety by design](#)" approaches into their services from the outset. This could mean:

- designing and testing AI with potential misuses in mind
- using watermarks and other indicators to label content as synthetic
- "nudging" users to refrain from certain behaviors (such as using pop-ups to remind them about the importance of consent).

Research shows there are gaps in public understanding of deepfakes and how to detect them. This further highlights a need for digital literacy and education on the difference between consensual and non-consensual uses of intimate images, and the harms of non-consensual deepfake porn.

Finally, and perhaps most importantly, we need to address the underlying systemic inequalities that contribute to technology-facilitated abuse against women and gender-diverse people. This includes recognizing [deepfake](#) porn for the often-gendered problem it is—for celebrities and non-celebrities alike.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Taylor Swift deepfakes: New technologies have long been weaponized against women. The solution involves everyone (2024, February 1) retrieved 8 May 2024 from <https://techxplore.com/news/2024-02-taylor-swift-deepfakes-technologies-weaponized.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.