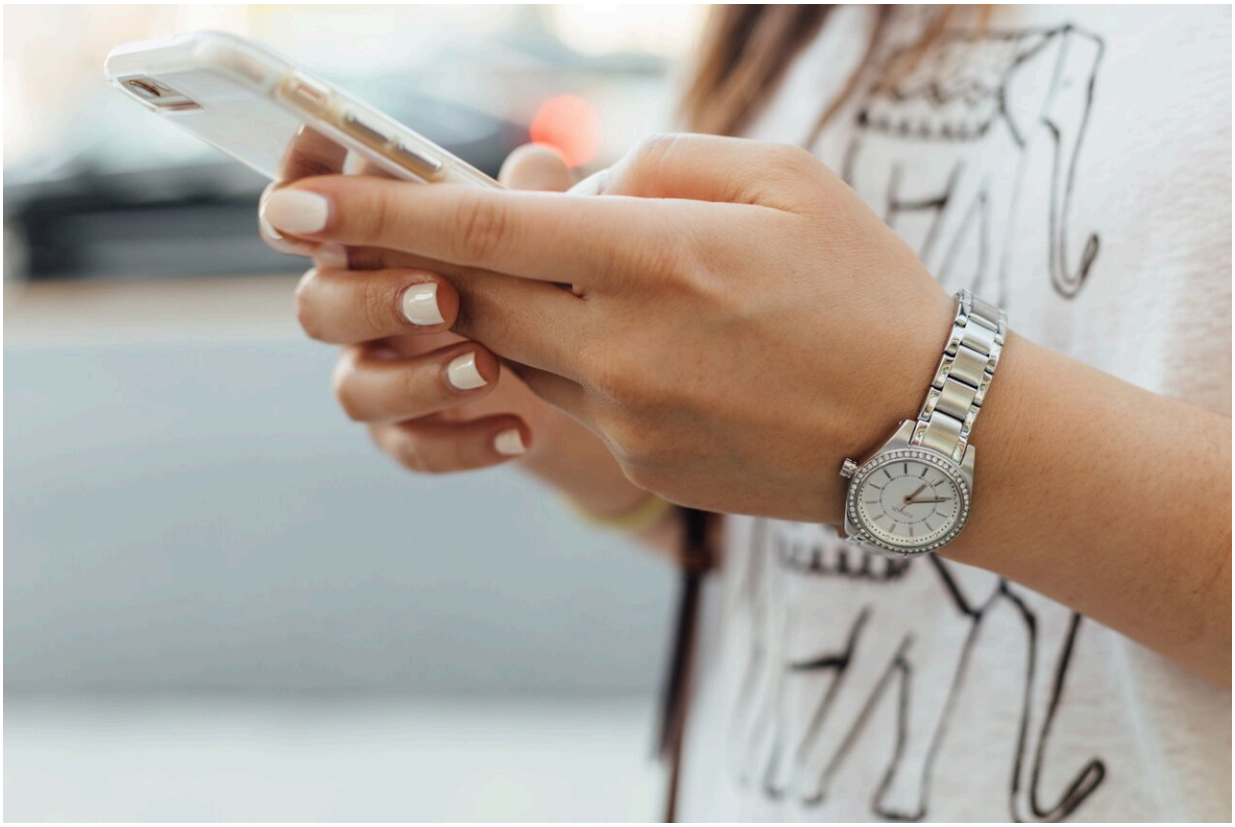
AI 'nudify' sites are being sued for victimizing people. How can we battle deepfake abuse?

August 21 2024, by Nicola Henry



Credit: Unsplash/CC0 Public Domain

Last week, the office of the San Francisco City Attorney issued a landmark lawsuit. It's accusing 16 "nudify" websites of violating United

States laws in relation to non-consensual intimate images and child abuse material.

"Nudify" sites and apps are easy to use. They let anyone upload a photo of a real person to generate a fake but photorealistic image of what they might look like undressed. Within seconds, someone's photo becomes an explicit image.

In the first half of 2024, the 16 websites named in the lawsuit have been visited more than 200 million times. One of the sites says, "imagine wasting time taking her out on dates, when you can just use [redacted site] to get her nudes."

These sites are also advertised on social media. Since the start of this year, there has been a 2,400% increase in advertising of nudify apps or sites on social media.

What can victims do?

Even if the images look fake, deepfake abuse can cause significant harm . It can damage a person's reputation and career prospects. It can have detrimental mental and physical health effects, including social isolation, self-harm and a loss of trust in others.

Many victims don't even know their images have been created or shared. If they do, they might successfully report the content to mainstream platforms, but struggle to get it removed from private personal devices or from "rogue" websites that have few protections in place.

Victims can make a report to a digital platform if fake, non-consensual intimate images of them are shared without their consent.

If they're in Australia, or if the perpetrator is based in Australia, the

victim can report to the eSafety Commissioner, who can work on their behalf to have the content taken down.

What can digital platforms do?

Digital platforms have policies prohibiting the non-consensual sharing of sexualized deepfakes. But the policies are not always consistently enforced.

Although most nudify apps have been removed from app stores, some are still around. Some "only" let users create near-nude images—say, in a bikini or underwear.

Tech companies can do a lot to stop the spread. Social media, video-sharing platforms and porn sites can ban or remove nudify ads. They can block keywords, such as "undress" or "nudify," as well as issue warnings to people using these search terms.

More broadly, technology companies can use tools to detect fake images. Companies behind the development of AI image-generator tools need to incorporate "guardrails" to prevent the creation of harmful or illegal content.

Watermarking and labeling of synthetic and AI-generated content are important—but not very effective once images have been shared. Digital hashing can also prevent the future sharing of non-consensual content.

Some platforms already use such tools to address deepfake abuse. They're part of the solution, but we shouldn't rely on them to fix the problem.

Search engines play a role, too. They can reduce the visibility of nudify and non-consensual deepfake sites. Last month, Google announced

several measures on deepfake abuse. When someone reports non-consensual explicit deepfakes, Google can prevent the content appearing in search results and remove duplicate images.

Governments can also introduce laws and regulatory frameworks to address deepfake abuse. This can include blocking access to nudify and deepfake sites, although VPNs can bypass blocked sites.

What does the law say?

In Australia, there are criminal laws on the non-consensual sharing of intimate images, or making threats to share intimate images against adults.

There are also federal offenses for accessing, transmitting, soliciting or possessing child abuse material. This includes fictional or fake images, including drawings, cartoons or images generated using AI.

Under Australian state and territory laws, an "intimate image" of an adult is defined broadly to include digitally altered or manipulated images. Currently, it's only a crime to share or make threats to share non-consensual, synthetic, intimate images. An exception is Victoria, where there's a separate criminal offense for producing intimate images, including digitally created ones.

In June, a bill was introduced to amend federal laws to create a standalone offense for the non-consensual sharing of private sexual material. The maximum prison sentence would be six years. The bill expressly mentions it's irrelevant whether the photos, videos or audio depicting the person are in "unaltered form" or have been "created or altered using technology."

The bill also includes two aggravated offenses, including a maximum of

seven years' imprisonment if the person who shared such images also created or altered them.

Laws are helpful, but can't fully solve the problem. Law enforcement often have limited resources for investigation. Working across jurisdictions, particularly in other countries, can also be difficult. For victim-survivors, pursuing the criminal justice path can take a further emotional toll.

Another option are civil remedies under the federal Online Safety Act. Administered by the eSafety Commissioner, civil penalties include formal warnings and hefty fines for users and tech companies that share or threaten to share non-consensual images.

We must improve our digital literacy

It's getting increasingly difficult to tell real and fake images apart. Even when images look "fake" or are labeled as such, people can still be led to believe they are real.

Investing in digital literacy is therefore crucial. Digital literacy means fostering critical thinking skills so people can assess and challenge misinformation.

Other measures include raising awareness on the harms of deepfake abuse and better education on respectful relationships and sexuality. Another one to tackle is porn literacy to improve critical thinking on the subject that isn't just focused on "unrealistic expectations."

Perpetrators who engage in deepfake abuse, tech developers who enable the tools, and tech companies that allow its spread must all be held accountable. But detecting, preventing and responding to this abuse will ultimately involve creative solutions across the board.

Provided by The Conversation