

Inside the Taylor Swift deepfake scandal: 'It's men telling a powerful woman to get back in her box'

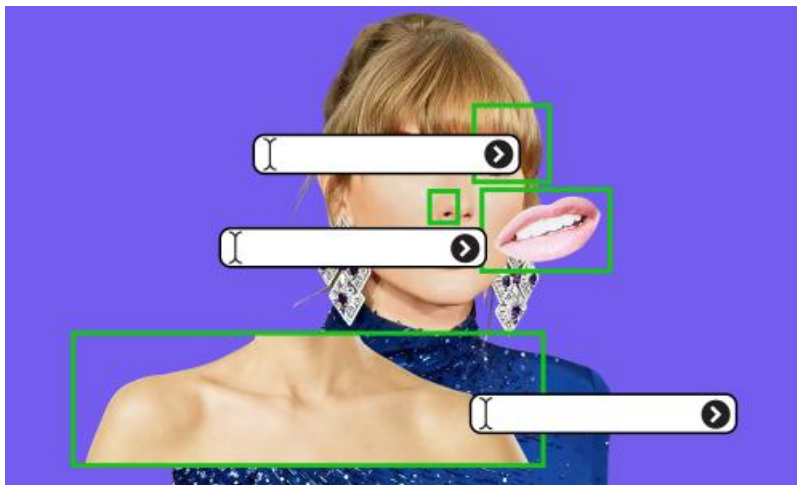
Publication Date: 2024-01-31

Author: Emine Saner

Section: Technology

Tags: Deepfake, Pornography, Women, Taylor Swift, features

Article URL: <https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box>



For almost a whole day last week, deepfake pornographic images of Taylor Swift rapidly spread through X. The social media platform, formerly Twitter, was so slow to react that one image racked up 47m views before it was taken down. It was largely Swift's fans who mobilised and mass-reported the images, and there was a sense of public anger, with even the White House calling it "alarming". X eventually removed the images and blocked searches to the pop star's name on Sunday evening. For women who have been victims of the creation and sharing of nonconsensual deepfake pornography, the events of the past week will have been a horrible reminder of their own abuse, even if they may also hope that the spotlight will force legislators into action. But because the pictures were removed, Swift's experience is far from the norm. Most victims, even those who are famous, are less fortunate. The 17-year-old Marvel actor Xochitl Gomez spoke this month about X failing to remove pornographic deepfakes of her. "This has nothing to do with me. And yet it's on here with my face," she said. Noelle Martin is a survivor of image-based abuse, a term that covers the sharing of nonconsensual sexual images and explicit deepfakes. She first discovered her face was being used in pornographic content 11 years ago. "Everyday women like me will not have millions of people working to protect us and to help take down the content, and we won't have the benefit of big tech companies, where this is facilitated, responding to the abuse," she says. Martin, an activist and researcher at the Tech & Policy Lab at the University of Western Australia, says that at first it was doctored pictures of her, but in the past few years, as generative AI has boomed, it has been videos, which are mostly shared on pornographic sites. "It is sickening, shocking," she says. "I try not to look at it. If I do come across it, it's just ..." She pauses. "I don't even know how to describe it. Just a wash of pain, really." Even if the images aren't particularly realistic, "it's still enough to cause irreparable harm to a person", she says. And good luck trying to get the images removed from the internet. "Takedown and removal is a futile process. It's an uphill battle, and you can never guarantee its complete removal once something's out there." It affects everything, she says, "from your employability to your future earning capacity to your relationships. It's an inescapable form of abuse, that has consequences that operate in perpetuity." Martin has had to mention it at job interviews. "It's something that you have to talk about on first dates. It infringes upon every aspect of your life." When the campaigner and writer Laura Bates published her book *Men Who Hate Women*, an investigation into the cesspits of online misogyny, men would send her images that made it look as if Bates was performing "all kinds of sex acts, including individuals who sent me images of myself changed to make it look like I was giving them oral sex". It's hard for people to understand the impact, she says, even when you know it's not real. "There's something really visceral about seeing an incredibly hyper-realistic image of yourself in somebody's extreme

misogynistic fantasy of you,” she says. “There’s something really degrading about that, very humiliating. It stays with you.” And that image can be shared with potentially millions of people, she adds. Deepfake pornographic images and videos are, says Bates, “absolutely circulated within extremist misogynistic communities”. What was particularly notable about the Swift abuse was “just how far they were allowed to circulate on mainstream social media platforms as well. Even when they then take action and claim to be shutting it down, by that point those images have spread across so many other thousands of forums and websites.” A 2019 study from the cybersecurity company Deeptrace found that 96% of online deepfake video content was of nonconsensual pornography. When the vast majority of AI is being used to create deepfake pornography, she points out, “this isn’t a niche problem”. It is, she says, “just the new way of controlling women. You take somebody like Swift, who is extraordinarily successful and powerful, and it’s a way of putting her back in her box. It’s a way of saying to any woman: it doesn’t matter who you are, how powerful you are – we can reduce you to a sex object and there’s nothing you can do about it.” In that way, it’s nothing new, says Bates, “but it’s the facilitated spread of this particular form of virulent misogyny that should worry us, and how normalised and accepted it is”. We know, says Rani Govender, a senior policy and public affairs officer at the NSPCC, “that this is an issue which is absolutely impacting young people. In the same way that other forms of image-based sexual abuse work, it particularly impacts girls.” There have been cases of children creating explicit deepfake imagery of other children, often using apps that “strip” a subject in a photo. “Then this is being sent around schools and used as a form of sexual harassment and bullying. Fear is a theme that comes up a lot: worrying that people will think it’s real, that it can lead to further sexual harassment and bullying. [There is] worry about what their parents might think.” One 14-year-old girl told the NSPCC’s ChildLine service last year that a group of boys made fake explicit sexual images of her and other girls and sent them to group chats. The boys were excluded from school for a time, but returned, and the girls were told to move on, which they struggled to do. Another girl, 15, said that a stranger had taken photographs from her Instagram account and made fake nudes of her, using her real bedroom as a background. Govender says this kind of material is created by strangers online as part of a grooming process, or can be used to blackmail and threaten children. AI has also been used to generate images of child sexual abuse, which are shared and sold by offenders. Even children who haven’t been targeted are still vulnerable to seeing the proliferation of deepfake pornography. “There’s already a big challenge with how much explicit and pornographic material is easily available to children on social media sites,” says Govender. “If it’s becoming easier to produce and share this material, that’s going to have really negative impacts on children’s views of the seriousness of these images as well.” The campaign My Image My Choice was started by the creators of the 2023 film *Another Body*, which is about an engineering student in the US who sought justice after discovering deepfake pornography of herself. A lot of the media coverage of AI, says the film’s co-director Sophie Compton, “was exclusively focused on threats to democracy and elections, and missing the violence against women angle. What we’ve seen over the last couple of years is the development of this community that was pretty fringe and dark and intense entering the mainstream in a really concerning way.” Women started getting in touch with her: “The number of responses we got was quite overwhelming.” For women who work online particularly, such as YouTubers, many “have basically had to accept that it’s part of the job, that they are going to be deepfaked on a huge scale”. The word deepfake – now used as a catch-all term to describe any digitally manipulated image or video that can look convincingly real – was originally coined to refer to pornography, points out Henry Ajder, a deepfakes and AI expert who has been researching this for years, and has advised the UK government on legislation. In 2017, Reddit forum users were putting female celebrities’ faces into pornographic footage. It was Ajder’s research in 2019 that found that almost all deepfake content was pornographic, and by 2020 he was discovering communities on the messaging platform Telegram “where hundreds of thousands of these images were being generated”. As AI quickly developed, it “changed the game yet again”. People using open-source software – as opposed to AI tools such as Dall-E 3 or Midjourney, which have been trained to prohibit pornographic content – can essentially create what they like, which can include extreme and violent fantasies made real. Swift is not a new target, says Ajder, who remembers explicit footage and images of her circulating five years ago. “What is novel in this case is the way that this content was able to spread on an open, popular social media platform. Most of this stuff prior has been shared in places like 4chan, Discord communities or on dedicated deepfake pornography websites.” Over the past six years, Ajder has spent a lot of time “in pretty dark corners of the internet, observing the characteristics and behaviours, the ways that these people who are creating this interact. It’s safe to assume that the vast, vast majority are men. I think a lot of people targeting celebrities are doing so for sexual gratification. It’s often accompanied by very misogynistic language – it may be sexual gratification, but it’s very much coupled with some pretty awful views about women.” He has seen men targeted, too, particularly in countries where homosexuality is forbidden, but the victims are overwhelmingly women. There have been cases, he says, where images have been created as “revenge porn”. “It’s also been used to target female politicians as a way to try to silence and intimidate them. It really does manifest a lot of the challenges that women already face, but provides a whole new visceral and very potent weapon to dehumanise and objectify.” Is there a financial motive? “Yes and no,” says Ajder. “Some websites have certainly profited, whether that’s through advertising revenue, or through charging [for images].” But with the leaps forward in technology, it has become more accessible than ever. “What previously might have been computationally very intensive and difficult can now be run on a gaming PC or a high-powered laptop.” Ajder believes millions of women and girls have been victims of this. “The amount of people that I now hear from in schools, and workplace contexts, who are falling victim to this is unsurprising, but still incredibly disturbing,” says Ajder. “While it’s sad

that it's taken one of the biggest celebrities in the world to be targeted for people to acknowledge how big a problem this is, my hope is that this can be a catalyst for meaningful legislative change." It should be "very clear", says Ajder, "that if you are creating or sharing or engaging with this kind of content, you are effectively a sex offender. You're committing a sexual offence against another human being." Under the UK's new online safety act, the sharing of nonconsensual deepfake pornographic material is illegal. "I don't think anyone's expecting large numbers of criminal convictions, but technically a lot of the sharing of these images of Taylor Swift would have constituted a criminal offence," says Clare McGlynn, a professor of law at Durham University and an expert in image-based abuse. She and others have been campaigning to change the law on altered images for many years, "but largely we were shouting into the void". For years, she says, the government's line was that the harms of fake images were not significant, "although, of course, they just asserted that without actually speaking to victims. It's a broader issue of online abuse against women and girls not being taken as seriously. People are not understanding that the harms of this can be profound and devastating and are constant and ongoing – it doesn't just happen and you can then try to get over it and move on with your life. It's always likely to be on the internet, always reappearing." McGlynn believes the Online Safety Act is a missed opportunity. "The offence is just about the distribution of an altered image – it's not about its creation." And it lets platforms off too easily. She says draft guidance from Ofcom, the regulator, is "relatively weak and focuses on individual pieces of content", rather than the entire systems that facilitate abuse. "It's not yet taking as strong a position to try and get the platforms to really do something." Social media companies such as Discord will point out they have moderators, while X says it has a "zero tolerance" policy towards posting nonconsensual nudity, although when an image can be viewed tens of millions of times before its removal, that starts to look a little hollow. AI is clearly only going to get better and become more easily available, with concerns about fake news, scams and democracy-shaking disinformation campaigns, but with deepfake pornography, the damage is already being done. "It's somewhat unique, compared to some of the other threats that AI-generated content poses, in that it does not have to be hyper-realistic to still do harm," says Ajder. "It can be clearly fake and still be traumatising and humiliating. It's already very potent." But still it could get worse in ways we have, and haven't, thought of. Ajder is concerned about AI-generated audio, which can replicate someone's voice, and as the pace of developments within virtual reality picks up, so will the possibility of sexual abuse within it. "We've already seen cases where you can quite crudely put the face of someone on to an avatar that you can effectively manipulate however you want, sexually. I worry that the very fast-evolving space of synthetic AI-generated video combined with virtual reality is going to lead to more abuse, particularly of women." We need to get over the idea that because it's online, or that it is labelled as fake, it isn't harmful, says Bates. "People think this isn't violence," she says. "There isn't any accountability for tech companies who are allowing this stuff to proliferate; there isn't any kind of any retribution for allowing this to happen." Whether you're a girl at school, or a woman whose photograph has been copied, or a global pop star, once those images are out there, points out Bates, "it's already too late". • Do you have an opinion on the issues raised in this article? If you would like to submit a response of up to 300 words by email to be considered for publication in our letters section, please [click here](#).