THE INVISIBLE THREAT OF DIGITAL FACE THEFT FOR WOMEN

Ketevan Khuturidze

Revised by: Elena Tsatsua

Abstract

The presentation aims to shed light on the alarming rise of deepfake technology and its implications, particularly focusing on its abuse in identity theft and particular targeting of women. With the rapid advancement of artificial intelligence, deepfakes have become a tool for distorting individuals' faces and voices, often without their consent, to promote offensive content or ideas. This phenomenon not only poses a significant threat to individuals' privacy and security but also undermines women's ability to control their digital footprint.

Introduction:

Every year, we mature as a society that has built, adopted, and coexisted with various technologies. Throughout these years, we have faced the consequences that were intentional and beneficial, or unintended and detrimental. Each time, an opportunity arises to adopt our socio-cultural laws and politics to fit the characteristics of our society. Technology exposes many truths, many of them being ugly. However, no other technological advance has exposed the deep-seated misogyny and patriarchy in our society as effectively as the evolvement of AI technology.

When concerns about AI are expressed, most may think of the futuristic threat of machines taking over the world. However, we are currently facing an immediate and significant threat from AI: deepfakes - synthetic media generated by artificial intelligence (AI) to depict a person's likeness. A combination of the phrases "deep learning" and "fake", this artificial intelligence technology can stitch the image of anyone's face to a video or a photo that they never participated in. When done well, these realistic videos or photos can be quite convincing, making a puppet out of the person featured in them.

Issues and concern

Much of the public concern about deepfakes is centered around fears about them being used to undermine democracy, spread disinformation, and disrupt the political scene. Most deepfake videos of political figures have been used for parody or to educate the public about the role that deepfakes could play in spreading misinformation and disinformation. During the 2019 UK election, Boris Johnson appeared in a deepfake by social enterprise Future Advocacy, in which he endorsed his opponent; in 2020, Britain's Channel 4 created an alternative Christmas message from the Queen in which she made uncharacteristic comments about her family and her position.

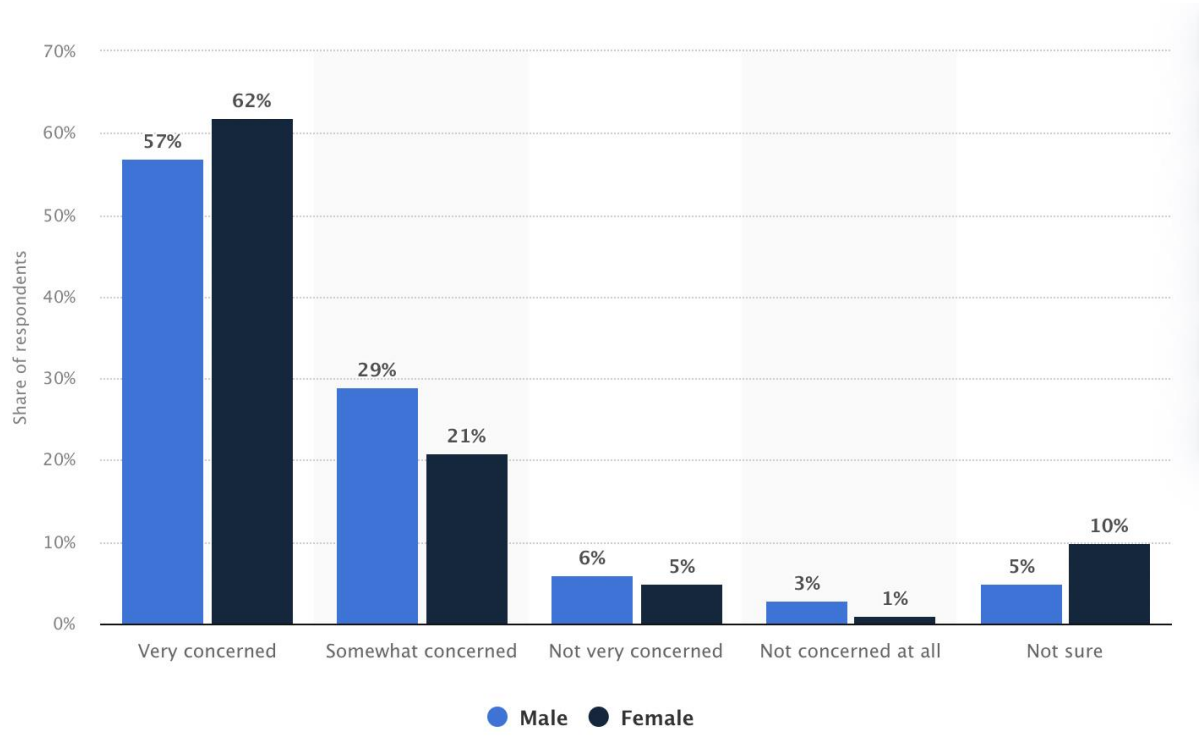
Both videos revealed onscreen that they were deepfakes and that their purpose was to educate the public of the potential misuse of deepfake technology. However, a recent report 'Mapping the Deepfake Landscape' by Sensity AI, an American organization that monitors the number of deepfakes online, reveals that less than 5% of deepfake videos online contain content that can impact the political sphere. Of the thousands of celebrities, public figures and everyday people who had deepfakes made of them, only 35 of these individuals were American

politicians. While this fact is indeed concerning and warrants action, we must also shift our attention to the larger context. More than 95% of deepfake content online is actually x-rated in nature and 99% of those were of women and girls. This is a form of gendered non-consensual deepfake pornography, where women's faces are digitally imposed onto nude or sexual images. This fact, unfortunately, reveals the deep misogyny and gender discrimination embedded within our society.

Gender scholars point out that what underlies this problem is men, masculinities, and unequal gender-sexual power relations in the work cultures of some tech organizations. AI being modeled according to current cultural and structural realities where several unsustainable ideals of masculinity dominate, it not only reflects these inequities, but also reinforces them. AI development has mostly been led by rich white men from Western countries, which has led to sexism, gender bias, and racism being incorporated in AI systems. This matter reflects a deeper issue rooted in disproportionate power dynamics and the continuing dominance of men and masculinities across societies. Therefore, it is unrealistic to discuss AI without addressing how certain men and masculinities figure in organizing around AI.

Research and consequences

Research indicates that while male respondents in the United States are concerned about deepfake videos of themselves being shared online, only 13% of the 86% who are worried think these videos could be of a sexual nature. In contrast, nearly all female respondents see this as a highly likely possibility. This highlights the significant differences in how men and women perceive the risks associated with deepfakes, further emphasizing the unequal gender dynamics ingrained in the development and impact of AI.



Given these underlying issues and disparities, the implications of AI and deepfake technology for women are profound and far-reaching. What is especially alarming is the sheer scale and scope of this phenomenon. One of the contributing factors for this fact is the open-source nature of the Machine Learning community, which has made advanced AI tools more accessible to

the general public. As a result, numerous websites now exist that can easily create deepfakes, allowing anyone with internet access to manipulate and fabricate realistic videos.

What can we do

The access is simplified further by creating chat-bots that generate deepfakes. Sensity discovered a massive operation using AI to create nude images of women and underage girls, primarily on the encrypted messaging app Telegram. Users sent images to the bot, which then generated fake nudes by superimposing them onto the original photos. Sensity's report highlights that deepfakes are mainly being used to harass, humiliate, and extort women. A poll indicated most users wanted fake nudes of women they knew personally, often stealing images from social media or private communications. Over 680,000 women had their images stolen, using an open-sourced deepfake software called "DeepNude". "It's horrifying and shocking to see yourself depicted in a way that you didn't consent to. It is violating. It is dehumanizing. And the reality is that I know, this could impact my employability. This could impact my relationships and mental health" said one of the victims. "DeepNude" went offline after some time as its servers couldn't handle the traffic brought to the app.

This widespread availability exacerbates the risks for women, as these tools can be used to produce non-consensual pornographic content, harass, and undermine the credibility of women on an unprecedented scale. The threats are very real and millions of women knowingly or unknowingly are victims. In May 2024, a Channel 4 News analysis found that nearly 4000 celebrities were listed on the most-visited deepfake websites. In January 2024, Taylor Swift became the latest high-profile target of nonconsensual deepfake images as a collection of sexually explicit, AI-generated images of Swift began going viral across several social media platforms including X and Meta, reaching over 47 million views in 19 hrs.

While celebrities are the focus of deepfakes, it is becoming more common for female public figures of all sorts to be targeted. In some cases, these videos have been expressly created as a tool of harassment. Rana Ayyub, a journalist in India who spoke out against the government's response to the rape of an eight-year-old girl, was the subject of a deepfake video made as part of a coordinated online hate campaign. Noelle Martin, a young woman in Australia who has been advocating about the issue of image-based sexual abuse, also became the subject of manufactured sexual images and deepfaked video. More recently, in June 2024, deepfake pornographic images of around 50 schoolgirls were distributed on the web, created by a teenager using artificial intelligence apps.

How to regulate

The popularization of deepfake technology, while promoting innovation, also underscores the urgent need for stricter regulations and ethical guidelines to protect against its misuse. It's clear that deepfake technology is rapidly hurtling out of control and outpacing the development of necessary legislation to monitor it. Although pornographic deepfakes first emerged in 2017 on Reddit, an American platform, there is still no federal legislation that specifically regulates deepfakes. However, US politicians have called for new laws to criminalize the creation of deepfake images, after the Taylor Swift incident. Other attempts have been made, like The UK Online Safety Act and the AI Act of EU, but they both fail to prevent the creation of explicit deepfakes and only regulate some cases.

Given that deepfakes are based on AI in the first place, some look to AI as a solution to harmful deepfake applications. For instance, researchers have built sophisticated deepfake detection

systems that assess lighting, shadows, facial movements, and other features in order to flag images that are fabricated. Another innovative defensive approach is to add a filter to an image file that makes it impossible to use that image to generate a deepfake. Also, a handful of startups have emerged that offer software to defend against deepfakes, including Truepic and Deeptrace.

Conclusion

As governments and researchers try to examine how to tackle the damaging use of deepfakes, it is critical that they pay attention to the people who are most harmed by deepfakes and recognize the systemic inequalities at fault. In the conversation about responding to deepfakes, non-consensual sexual deepfakes of women should not be a side issue, but at the very center of the discussion and action.

References

1. Statista: U.S. Adults' Fear of AI and Deepfakes by Gender
2. GOV.UK: A Guide to the Online Safety Bill
3. European Commission: Regulatory Framework on AI
4. https://www.taylorfrancis.com/books/edit/10.4324/9781003193579/routledge-handbook-men-masculinities-organizations-jeff-hearn-kadri-aavik-david-collinson-anika-thym?_ga=302483088.1718205760&_gl=1*1mobxg9*_ga*MzAyNDgzMDg4LjE3MTgyMDU3NjA.*_ga_0HYE8YG0M6*MTcxODUyOTkxMy4yLjAuMTcxODUyOTkyMi41MS4wLjA.*_gcl_au*NjYyMzA2MDQwLjE3MTgyMDU3NTk.
5. Routledge: Handbook on Men, Masculinities and Organizations

`