So you've been scammed by a deepfake. What can you do?

February 26 2024, by Jeannie Marie Paterson

Credit: Pixabay/CC0 Public Domain

Earlier this month, a Hong Kong company lost HK$200 million (A$40 million) in a deepfake scam. An employee transferred funds following a video conference call with scammers who looked and sounded like

senior company officials.

Generative AI tools can create image, video and voice replicas of real people saying and doing things they never would have done. And these tools are becoming increasingly easy to access and use.

This can perpetuate intimate image abuse (including things like "revenge porn") and disrupt democratic processes. Currently, many jurisdictions are grappling with how to regulate AI deepfakes.

But if you've been a victim of a deepfake scam, can you obtain compensation or redress for your losses? The legislation hasn't caught up yet.

Who is responsible?

In most cases of deepfake fraud, scammers will avoid trying to fool banks and security systems, instead opting for so-called "push payment" frauds where victims are tricked into directing their bank to pay the fraudster.

So, if you're seeking a remedy, there are at least four possible targets:

1. the fraudster (who will often have disappeared)
2. the social media platform that hosted the fake
3. any bank that paid out the money on the instructions of the victim of the fraud
4. the provider of the AI tool that created the fake.

The quick answer is that once the fraudster vanishes, it is currently unclear whether you have a right to a remedy from any of these other parties (though that may change in the future).

Let's see why.

The social media platform

In principle, you could seek damages from a social media platform if it hosted a deepfake used to defraud you. But there are hurdles to overcome.

Platforms typically frame themselves as mere conduits of content—which means they are not legally responsible for the content. In the United States, platforms are explicitly shielded from this kind of liability. However, no such protection exists in most other common law countries, including Australia.

The Australian Competition and Consumer Commission (ACCC) is taking Meta (Facebook's parent company) to court. They are testing the possibility of making digital platforms directly liable for deepfake crypto scams if they actively target the ads to possible victims.

The ACCC is also arguing Meta should be liable as an accessory to the scam—for failing to remove the misleading ads promptly once notified of the problem.

At the very least, platforms should be responsible for promptly removing deepfake content used for fraudulent purposes. They may already claim to be doing this, but it might soon become a legal obligation.

The bank

In Australia, the legal obligations of whether a bank has to reimburse you in the case of a deepfake scam aren't settled.

This was recently considered by the United Kingdom's Supreme Court, in a case likely to be influential in Australia. It suggests banks don't have a duty to refuse a customer's payment instructions where the recipient is suspected to be a (deepfake) fraudster, even if they have a general duty to act promptly once the scam is discovered.

That said, the UK is introducing a mandatory scheme that requires banks to reimburse victims of push payment fraud, at least in certain circumstances.

In Australia, the ACCC and others have presented proposals for a similar scheme, though none exists at this stage.

The AI tool provider

The providers of generative AI tools are currently not legally obliged to make their tools unusable for fraud or deception. In law, there is no duty of care to the world at large to prevent someone else's fraud.

However, providers of generative AI do have an opportunity to use technology to reduce the likelihood of deepfakes. Like banks and social media platforms, they may soon be required to do this, at least in some jurisdictions.

The recently proposed EU AI Act obligates the providers of generative AI tools to design these tools in a way that allows the synthetic/fake content to be detected.

Currently, it's proposed this could work through digital watermarking, although its effectiveness is still being debated. Other measures include prompt limits, digital ID to verify a person's identity, and further education about the signs of deepfakes.

Can we stop deepfake fraud altogether?

None of these legal or technical guardrails are likely to be entirely effective in stemming the tide of deepfake fraud, scams or deception—especially as generative AI technology keeps advancing.

However, the response doesn't need to be perfect: slowing down AI-generated fakes and frauds can still reduce harm. We also need to pressure platforms, banks, and tech providers to stay on top of the risks.

So while you might never be able to completely prevent yourself from being the victim of a deepfake scam, with all these new legal and technical developments, you might soon be able to seek compensation if things go wrong.

Provided by The Conversation

Citation: So you've been scammed by a deepfake. What can you do? (2024, February 26) retrieved 27 April 2024 from https://techxplore.com/news/2024-02-youve-scammed-deepfake.html