ALM | LAW.COM

CORPORATE COUNSEL

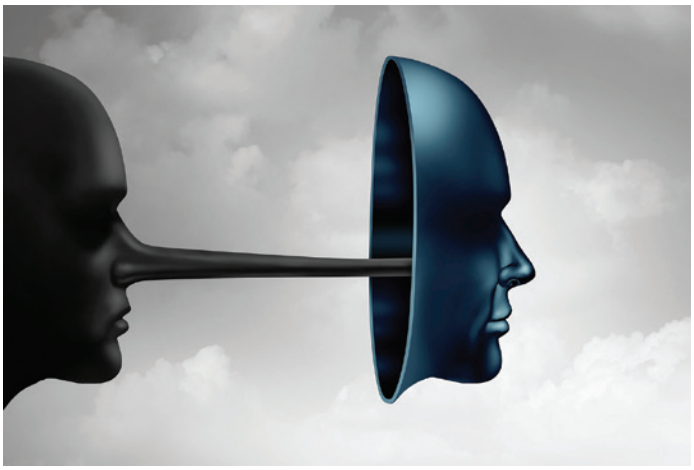
Deep Fakes Causing Real Harms: Time to Take Action

By Harry Valetk
March 1, 2024

"Deep fake" technology has unleashed new powers to distort reality in ways and on a scale that is not yet fully understood.

Powered by artificial intelligence and machine learning techniques that are exponentially advancing their sophistication, deep fakes are increasingly difficult to detect. The result are attacks that distort reality more frequently and convincingly than many in the boardroom may realize, and are wreaking havoc via audio and video of real people saying and doing things they never said or did. Left unchecked, deep fakes will increasingly cause marketplace disruptions, inflict individual and corporate reputational harm, and undermine our fundamental understanding of truth.

In February, for example, a Hong Kong-based clerk working for an undisclosed multinational firm was duped into transferring a total of HK$200 million (approximately $25 million) to unauthorized fraudsters. According to published reports, the clerk made a total of 15 transactions after joining a video conference where the other participants resembled the clerk's own coworkers, but were actually AI-generated deep fakes. Authorities speculate that fraudsters downloaded videos of the clerk's co-workers in advance, and then used AI-powered technology to add fake voices and imagery to use in the video conference.



Credit: freshidea/Adobe Stock

In this context, in-house counsel must ask, what are the legal and business considerations companies managing through such unprecedented risks? Consider the following.

1. Define What Deep Fake Means for Your Organization

Digital impersonations or doctored imagery is nothing new. Deep fake technology capable of escaping detection via machine learning

algorithms that insert faces and voices into video and audio recordings of actual individuals, however, is a new and formidable threat. A good first step, therefore, is to define "deep fake" within your organization's existing policy framework, industry requirements, and risk management nomenclature.

Legal frameworks here are still evolving to address risks and potential harms associated with deep fakes. In the U.S., state laws have only begun to define and proscribe certain uses of deep fake imagery. In Minnesota, for example, the use of deep fakes is only restricted in the context of sexual acts and elections. In the context of online platforms, moreover, Minnesota also criminalizes disseminating or entering into a contract or agreement to disseminate an ad that includes a deep fake—meaning that online platforms that accept an online political ad that includes a deep fake could also be in violation of the law. The term "deep fake" is broadly defined in Minnesota as any video recording, motion-picture film, sound recording, electronic image, or photograph … (1) that is so realistic that a reasonable person would believe it depicts speech or conduct of an individual; and (2) the production of which was substantially dependent upon technical means, rather than the ability of another individual to physically or verbally impersonate such individual.

2. Detection and Risk Mitigation

Deep fake expertise is essential. Procure marketplace tools to detect deep fakes, and assess what risks deep fakes may pose to your business. No one-stop-shop solution may yet exist, and that means you may need to leverage several third party service providers and products to monitor for potentially harmful deep fakes impersonating your senior executives, social influencers, or other key individuals. Once a potential deep fake is detected, separate expertise may be required to establish the authenticity or altered nature of a video or audio message for evidentiary purposes.

And while any organization could be targeted, pay heightened attention to possible deep fakes if your company is about to go public, launch a new product, or is engaged in a corporate transaction. Timing is everything with deep fakes, especially if the ultimate goal is to manipulate public opinion, manipulate stock prices, or disrupt a transaction. Deep fake videos, for example, could feature executives falsely admitting criminal conduct, taking bribes, displaying racism, or any other objectionable audio or video footage designed to skew information or manipulate public opinion.

For this reason, wherever possible, pause before you act. In a world where seeing is no longer necessarily believing, approval protocols and safety checks may be appropriate before reacting to unusual messages or payment instructions. Finally, consider doing your investigation subject to attorney-client privilege may offer advantages if subsequent criminal or civil proceedings take place.

3. Build Deep Fakes Into Your Security Incident Response Plan

Ownership is poorly understood within organizations in the context of deep fakes. Common unresolved questions include, who is responsible for reporting, handling, and remediating? What policy defines a deep fake, and

what process explains what must happen before such a situation is considered resolved? Is it appropriate to contact law enforcement? If so, who should contact law enforcement, and when?

Consider, therefore, baking deep fake scenarios into your existing security incident response plan. This may make sense because similar multi-stakeholder teams must be assembled once detection has been confirmed. And while the outcomes and resolutions may differ from a cybersecurity incident, similar disciplines, tools, and protocols may be used to manage deep fake takedowns, alert senior executives and stakeholders, and notify law enforcement or supervisory authorities.

4. Counter Dissemination Strategy

Long gone are the days when the public's attention is exclusively in the hands of trusted journalists or media outlets that can be managed with court-ordered injunctions or other legal process. Ours is a world where alternative outlets and social media platforms serve as fertile grounds for disseminating deep fakes instantaneously and virally to the indiscriminate masses. Understanding the disruptive object of deep fakes can serve as the basis for an effective counteroffensive. Proactively tracking potentially harmful deep fakes, tracing potential sources of the attack, and counter messaging with accurate and truthful information is key to debunking lies. But in order to counter the filter bubbles aggravating the spread of false narratives, internal and external messaging experts must be retained and deployed in advance.

5. Post-Incident Training, Evidence Preservation, and Public Awareness

Finally, treat each incident as a learning opportunity to assemble a list of lessons learned. Raise internal and external awareness of this emerging enterprise risk. Managing deep fake incidents presents unique challenges. Understanding areas where your team could have performed better will lead to better performance in the future. Preserve evidence of deep fake attacks against your organization, wherever possible, and look for trends. Also look for appropriate ways to share evidence with law enforcement and other regulatory partners. Explore if your existing insurance policies cover damages resulting from a deep fake attack, including costs associated with an investigation and subsequent fallout.

More broadly, spark internal and external conversations about risks associated with deep fakes. A greater public awareness is needed about the harmful and inflammatory impact deep fake audio and video imagery can have on our society, and the risks associated with reputational attacks on our public and private institutions.

Harry A. Valetk *is a partner at Loeb & Loeb's privacy, security, and data innovations practice group based in New York. He focuses his practice on delivering commercially practical advice to companies of every size on privacy, cybersecurity incident response, and deployment of generative artificial intelligence (AI) and data-driven platforms. He can be reached at hvaletk@loeb.com.*