WHY DEEPFAKES POSE A SERIOUS THREAT TO BUSINESSES

AND WHAT TO DO ABOUT IT

FINSBURY
GLOVER
HERING

Digital Financial Communications: Deepfakes

The recent entrance of deepfake technology has made digital impersonation more realistic and convincing than ever. Businesses are likely to be among the first entities targeted with deepfakes. Here’s what companies can do to head off serious risks and mitigate damage.

What are deepfakes?

Deepfakes refer to media that has been altered by artificial intelligence (AI) to make it appear that a person is doing or saying something that, in fact, was never done or said. The technique leverages deep learning algorithms to superimpose hyper-realistic face images of a target person onto another person‘s body. Source material to create a deepfake of a CEO, for instance, can be taken from earnings calls, TV interviews, YouTube videos, TED talks and other recordings. With proper post-processing, the resulting videos can be nearly undetectable. A powerful demonstration of what deepfakes can look like is this video of British Prime Minister Boris Johnson produced by UK-based artist Bill Posters in collaboration with the research organization Future Advocacy.

While most deepfakes are created for entertainment purposes, there are growing concerns about a possible weaponization of the underlying technology for malicious ends. A recent study by researchers at the University College London, published in *Crime Science*, even ranked deepfakes as the most serious AI crime threat. And Moody’s Investor Service said that companies’ businesses and credit quality are threatened as advancing technology makes it easier to create deepfake videos and images designed to damage their reputations. Possible dangers are obvious: imagine a deepfake video showing a CEO of a listed company accepting a bribe or confessing financial fraud. Such deepfake attacks could ruin the reputations of executives and companies, and destroy irrevocable trust and shareholder value within minutes.

Here are five potential threat scenarios:

- **Scenario 1: Phishing attacks**
  Deepfakes have the obvious potential to be used for fraud purposes, to pretend to be someone else to fool targets into divulging sensitive information. This could include the synthetic impersonation of a colleague or client over Microsoft Teams, asking for details on a project or access to a database.

- **Scenario 2: Transaction scams**
  Then there's the issue of payment fraud. By now, we're all familiar with the type of fraud in which a senior executive appears to send an email to an unsuspecting employee that requests an urgent money transfer. Meanwhile, cybercriminals have fooled a UK-based energy firm into making a $243,000 wire transfer using a state-of-the-art AI-powered deepfake of its CEO's voice. There are likely to be many other such instances that go unreported.

- **Scenario 3: Blackmail and extortion**
  Deepfakes could also be part of extortion campaigns, in which a company is offered a choice to either pay a fee to stop a deepfake from being made public or suffer the consequences. When faced with compromising material, many company executives may find it easier to pay ransom than to prove to the public that a video is fake. In such an extortion scenario, the authenticity of the deepfake is almost irrelevant, as the damage can be caused instantaneously.

- **Scenario 4: Stock manipulation**
  A well-timed deepfake in the financial results season, for instance showing your CEO announcing a major operating loss (or the termination of a line of business), could send your stock prices plummeting. Even if the deepfake in question is later proven to be fake, ruthless shortsellers could cash in on those stock swings.

- **Scenario 5: Fabricated regulatory action**
  Deepfake campaigns could not only be targeted at individual actors or companies, but also at entire markets. Imagine, for example, a video clip of a fabricated central bank meeting depicting officials discussing liquidity problems or future interest rate changes, or a news anchor or stock market expert announcing a market crash.

While scientists are developing forensic solutions that can detect deepfakes automatically, there are to date no definitive countermeasures against them. And social media platforms, including YouTube, have yet to prove they have technology effective enough to stop deepfake content from being published and spread.

Here are five powerful antidotes that companies can take:

- **Foster greater awareness**
  Threat-aware employees are your organization's first line of defense. Given the novelty of the threat posed by deepfakes, companies should sensitize and train their employees and managers – especially the media monitoring staff – how to identify fake footage. Subtle imperfections to look for are: unnatural eye blinking, blurry face borders, artificially-looking skin, slow speech, and unusual intonation.

- **Plan your response**
  In case of a deepfake incident, companies need to be able to prove that the material that has appeared is fake. Commercial tools or independent experts can help. Once you've got the proof, however, your communications department needs to have a plan in place for spreading the word quickly and effectively. Since standard crisis processes will only be applicable here to a very limited extent, crisis manuals should incorporate processes that are specifically designed for deepfake scenarios.

- **Establish contact with social networks**
  Well-made deepfake videos can easily go viral and reach millions of users within minutes when injected into the social media bloodstream. That's why it's important to react as quickly as possible when a deepfake appears that threatens your business. As the standard reporting processes for fake content on social media usually take far too long, companies need to know who to contact in an emergency to have the content removed quickly.

- **Keep track of public appearances**
  As mentioned above, it is absolutely crucial to achieve interpretive dominance over the fake content that has emerged. Alternatively, exposing the false nature of the video can be accomplished by demonstrating what really was (and wasn't) said at the original event from which the deepfake was created. We therefore recommend tracking and filming every public appearance of your company's most important executives – and archiving the recordings (verified by tokens, trust seals or blockchain based technology).

- **Expand insurance coverage**
  More and more companies are taking precautions to protect themselves against cyber-attacks. While cyber insurance policies do seem to be growing broader in terms of coverage these days, financial loss resulting from deepfake-induced reputational harm will most likely not be included in most contracts. Risk managers should take this as an opportunity to review the individual policy terms and conditions before it is too late.

At present, deepfakes already pose a serious threat to individual and organizational reputations. However, as the technology to create them will become cheaper and even more sophisticated in the future, the problem is only likely to worsen. We believe that companies should consider this emerging possibility of synthetic media attacks and act now. Failing to do so may leave them exposed to the risk of losing irretrievable corporate reputation.