As deepfakes become more convincing, the threat to truth grows

theneweconomy.com/technology/as-deepfakes-become-more-convincing-the-threat-to-truth-grows

A still from a real video of US President Barack Obama. Suwajanakorn's demonstration involved mapping Barack Obama's voice onto four different videos of the US President

In April 2013, the official Twitter account for the Associated Press (AP) posted a tweet that wiped $136.5bn off the S&P 500 index in the space of just three minutes. It read: "Breaking: two explosions in the White House and Barack Obama is injured." Several other news outlets began reporting the event, causing further disruption to the Dow Jones Industrial Average, as well as to bond and commodity markets.

There was just one problem: the explosions never happened. AP's Twitter account had been hacked by a collective known as the Syrian Electronic Army. Although the tweet was swiftly removed and the news debunked, its brief visibility was enough to wreak havoc on the financial markets. Losses were swiftly recovered, but the incident served as a reminder of the effect misinformation can have on the financial and corporate spheres.

> Facial-mapping software can create a convincing video of someone doing or saying something that never actually happened

With the world still struggling to get to grips with the subtle intricacies of social media and the internet, a new threat is emerging that could deepen humanity's descent into a post-truth age: AI researchers are developing algorithms and software solutions to create fake videos that are almost indistinguishable from the real thing. These hyper-realistic 'deepfakes' can be used to undermine democracy, damage investor confidence, incite violence and extort money. They pose a huge risk to public trust in politics and business – and tackling the problem won't be easy.

A false alarm

Creating convincing imitations of static images has been possible for decades. In 1920, Arthur Conan Doyle, the creator of Sherlock Holmes, was so taken in by hoax images of fairies that he wrote an article asserting the existence of the supernatural creatures. And when Adobe Photoshop was released in 1990, image manipulation was introduced to the masses.

Creating fake videos capable of deceiving the viewer has always been more difficult: the human mind is finely tuned to identify an imperfect replica of a person's likeness. It's a phenomenon that has been dubbed 'the uncanny valley' and usually conjures up feelings of unease or revulsion in the observer.

However, technological advances are allowing fake videos to escape the uncanny valley. During a TED Talk in April, computer scientist Supasorn Suwajanakorn explained how deep learning is being used to create audiovisual works that look and sound authentic. By training a neural network to analyse static images and existing moving clips of an individual, facial-mapping software can create a convincing video of someone doing or saying something that never actually happened.

https://youtu.be/o2DDU4g0PRo

Suwajanakorn’s demonstration involved mapping Barack Obama’s voice onto four different videos of the US President. In this instance, the aim was not to deceive the audience, but it is easy to see how this technology could be used as a political weapon. According to Robert Chesney, Associate Dean for Academic Affairs at the University of Texas School of Law, deepfakes will exacerbate two existing threats to democracy.

“First, we already have a problem with disinformation campaigns designed to stoke divisions, favour one side or undermine support for particular policies,” Chesney told *The New Economy*. “These campaigns become more effective insofar as the credibility of their falsehoods and frauds improve. Second, there is always a risk of a highly focused attempt to tip an election by releasing false but damaging information close to the vote, with the limited time window for debunking creating more space for falsehoods to flourish.”

The enhanced credibility and difficulty of detection, combined with the special salience of supposed video or audio evidence, make the rise of deepfakes an especially dangerous proposition. In 1964, 77 percent of Americans trusted the government; today, this figure stands at just 24 percent. Deepfakes are sure to result in this number falling further still.

Public image
To date, deepfake software has largely been limited to superimposing the faces of celebrities onto the bodies of pornographic actors. This is undoubtedly unethical, and websites like Reddit have acted swiftly to ban this sort of content, but it also represents just the tip of the iceberg for deepfake misuse. Businesses could face all sorts of risks once deepfake technology is in the hands of rivals, disgruntled customers or embittered employees.

Public trust in the US Government:

77%

1964

24%

2018

“Reputational sabotage is the most obvious risk to business: just as someone might try to sink an electoral candidate, one might try to harm a business through a well-crafted deepfake purporting to reveal something terrible (say, a key leadership figure making racist statements),” Chesney explained. “Blackmail is a possibility as well.”

The software used for deepfakes relies on a large number of inputs – the more existing photos it can use to train itself, the more convincing the deepfake. For individuals in the public eye, this is a frightening prospect.

A quick Google image search for the likes of Donald Trump or Elon Musk will provide more than enough content for neural networks to work with – this means CEOs at the biggest companies are most at risk of being targeted. But as deepfake technologies improve, the number of images required to create an accurate video is likely to fall.

There are already several programs that have automated the process of creating deepfakes. One of the most popular, FakeApp, is available for download on desktop PCs and makes it easy to generate fake videos if a user has enough images of their subject to work with. Just as anyone with a social media account or personal blog can spread misinformation today, it will soon be possible to create deepfake videos with nothing more than a mobile phone and the right app.

Spot the difference

For Chesney, finding a way to prevent the spread of deepfake videos, particularly as they become more realistic, is the “million-dollar question”. A number of technology start-ups are already working to create digital watermarks and find other solutions to the problems posed by fake video technology.

However, these efforts could turn out to be self-defeating: the deep-learning techniques used to spot face-swap videos could, in turn, be used to improve the quality of deepfakes.

“It will not be easy to stop deepfakes,” Chesney admitted. “Education will help, but it will also hurt us in an offsetting way. If we all become inclined to scepticism when presented with audio or video evidence, this will be a boon to liars who hope to escape accountability for legitimate evidence of their wrongdoing.”

The wealthiest individuals may choose to pay for ‘authenticity trails’ or ‘lifelogs’ that can disprove false claims, but the cost to privacy would likely be as great – if not greater – than the financial one. Meanwhile, those less comfortable with having their every action recorded will need to rely on media platforms – including the likes of Facebook and Google – identifying fake content before it is posted online.

It is often said that a lie can travel halfway around the world while the truth is putting on its shoes. Even if it is possible to quickly debunk deepfake videos, it might not be quick enough to prevent a political or corporate disaster from taking place.

Related topics: Artificial Intelligence, deepfake