TechXplore

Evaluating the risks posed by deepfakes

December 17 2019

A few weeks ago, French charity Solidarité Sida caused a sensation when it published a fake yet highly realistic video of Donald Trump proclaiming "AIDS is over" as part of an awareness-raising campaign. The video in question is what's known as a deepfake, a technique that involves using machine learning to fabricate increasingly realistic images and videos as well as audio and text files.

This use of a deepfake video by a charity highlights the growing prevalence of this phenomenon. While pornography currently accounts for the vast majority of deepfake videos, the technique can also be used to defraud, to defame, to spread fake news or to steal someone's identity.

Evolving threats

In September, EPFL's International Risk Governance Center (IRGC) brought together around 30 experts for an interdisciplinary seminar to discuss this fast-evolving phenomenon and its growing prevalence. The IRGC has today published a report containing valuable insights into the risks associated with deepfakes.

The headline observation is that these risks could potentially cause widespread harm across many areas of life. "Any business organization or activity that relies on documentary evidence is potentially vulnerable," says Aengus Collins, the report's author and deputy director of the IRGC. Deepfakes can cause a great deal of uncertainty and confusion. In a recent case, thieves used deepfaked audio of a chief executive's voice to steal money from a company. On a society-wide scale, a proliferation

of fabricated content could undermine truth and erode public trust, the very cornerstones of democratic debate.

The report provides a framework for categorizing deepfake risks. It highlights three key impacts—reputational harm, fraud and extortion, and the manipulation of decision-making processes—and notes that these impacts can be felt individually, institutionally or across society.

With such a wide range of potential harm from deepfakes, where are risk-governance responses most needed? The experts recommend focusing on the scale and severity of the potential harm, as well as the ability of the "target" to cope with the fallout. For instance, a well-resourced company with established processes will be better able to absorb the impact of a deepfake attack than a private victim of harassment.

Interdependent solutions

In the report, the IRGC sets out 15 recommendations covering a variety of potential responses to deepfakes that could mitigate the risks they present. It also calls for deeper research across the board.

One of the main categories of recommendation is technology, including tools that can verify the provenance of digital content or detect deepfakes. At EPFL, the Multimedia Signal Processing Group (MMSPG) and startup Quantum Integrity are currently developing a deepfake detection solution that could be deployed in 2020. "For any given set of defenses, there will be vulnerabilities that can be exploited," Collins says. "But maintaining and developing technological responses to deepfakes is crucial to deterring most of the misuses."

The report also highlights the need for a greater focus on the legal status of deepfakes, in order to clarify how laws in areas such as defamation,

harassment and copyright apply to synthetic content.

More generally, digital literacy has an important role to play. But Collins cautions that there is a paradox here: "One of the goals of digital literacy in this area is to encourage people not to take digital content at face value. But there also needs to be a positive focus on things like corroboration and the assessment of sources. Otherwise, encouraging people to distrust everything they see risks exacerbating problems related to the erosion of truth and trust."

Wider horizons

While the IRGC report focuses on deepfake risk governance, this research is part of a wider workstream on the risks associated with emerging and converging technologies, which will continue in 2020. "We are currently deciding what our next focus will be," says Collins. "And there is no shortage of candidates. We live in a time when the relationship between technology, risk and public policy is more important than ever."

More information: Forged Authenticity - Governing Deepfake Risks. infoscience.epfl.ch/record/273296/files/Forged%20Authenticity%20Governing%20Deepfake%20Risks.pdf

Provided by Ecole Polytechnique Federale de Lausanne