Deepfakes: Informed digital citizens are the best defense against online manipulation

January 8 2020, by Nadia Naffi

Credit: CC0 Public Domain

More than a decade ago, Internet analyst and new media scholar Clay Shirky said: "The only real way to end spam is to shut down e-mail

communication." Will shutting down the Internet be the only way to end deepfake propaganda in 2020?

Today, anyone can create their own fake news and also break it. Online propaganda is more misleading and manipulative than ever.

Deepfakes, a specific form of disinformation that uses machine-learning algorithms to create audio and video of real people saying and doing things they never said or did, are moving quickly toward being indistinguishable from reality.

Detecting disinformation powered by unethical uses of digital media, big data and artificial intelligence, and their spread through social media, is of the utmost urgency.

Countries must educate and equip their citizens. Educators also face real challenges in helping youth develop eagle eyes for deepfakes. If young people lack confidence in finding and evaluating reliable public information, their motivation for participating in or relying on our democratic structures will be increasingly at risk.

Undermining democracy

It is now possible to generate a video of a person speaking and making ordinary expressions from just a few or even a single image of this person's face. Face swap apps such as FaceApp and lip-sync apps such as Dubsmash are examples of accessible user-friendly basic deepfake tools that people can use without any programming or coding background.

While the use of this technology may enrapture or stun viewers for its expert depictions in entertainment and gaming industries, the sinister face of deepfakes is a serious threat to both people's security and democracy.

Deepfakes' potential to be used as a weapon is alarmingly increasing and many harms can be anticipated based on people's ability to create explicit content without others' consent.

It's expected that people will use deepfakes to cyberbully, destroy reputations, blackmail, spread hate speech, incite violence, disrupt democratic processes, spread disinformation to targeted audiences and to commit cybercrime and frauds.

Deepfake detection

Key players have ventured into finding a response to deepfake threats.

Facebook announced Jan. 6 it "will strengthen its policy toward misleading manipulated videos that have been identified as deepfakes." The company says it will remove manipulated media that's been "edited or synthesized —beyond adjustments for clarity or quality—in ways that aren't apparent to an average person" and if the media is "the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic."

The news follows Facebook's "deepfake challenge," which aims to design new tools that detect manipulated media content. The challenge is supported by Microsoft, a consortium on artificial intelligence and a US$10-million fund.

In late October, Facebook CEO Mark Zuckerberg testified at a U.S. House of Representatives Financial Services Committee hearing in Washington about the company's cryptocurrency plans, where Zuckerberg faced questions about what the company is doing to prevent deepfakes.

The Defense Advanced Research Projects Agency (DARPA) of the U.S.

Department of Defense is working on using specific types of algorithms to assess the integrity of digital visual media.

Some researchers discuss the use of convolutional neural networks —a set of algorithms that loosely replicates the human brain, designed to analyze visual imagery and recognize patterns —to detect the inconsistencies across the multiple frames in deepfakes. Others propose algorithms to detect completely generated faces.

Hani Farid, an expert in digital forensics and one of the leading authorities on detecting fake photos, and his student Shruti Agarwal at University of California, Berkeley are developing a software that uses the subtle characteristics of how a person speaks to distinguish this person from the fake version.

Farid is also collaborating very closely with deepfake pioneer Hao Li to confront the problem of "increasingly seamless off-the-shelf deception."

YouTube nation

What if we wake up tomorrow to a deepfake of Greta Thunberg, *Time* magazine's 2019 Person of the Year, accusing a specific organization to be the major catalyst of climate change? Would any youth be skeptical of the information?

We are living in a digital era when many people expect every answer to be found through a Google search, a YouTube or a Vimeo video or a TED talk. Nearly 100 percent of Canadian youth between 15 to 24 years old use the internet on a daily basis. Most follow news and current affairs through social media platforms such as Facebook, Twitter and Instagram.

In 2017, 90 percent of Canadians aged 18 to 24 were active YouTube

users.

According to Statista, a company that provides market and consumer data, "as of May 2019, more than 500 hours of video were uploaded to YouTube every minute," equating to "approximately 30,000 hours of newly uploaded content per hour." The company reports that between 2014 and 2019 "the number of video content hours uploaded every 60 seconds grew by around 40 percent."

Many of today's 18- to 24-year-old social media users recognize the agendas and algorithms behind the posts that pop up on their walls. In my Ph.D. thesis research, I explored how 42 participants in this age group understood refugees in a contexts where ideas about refugees were deeply influenced by social media propaganda, fake news and disinformation. I found that many craved to become influencers and disrupt public commentary and media-generated messages in ways that resonate with advocacy or activist campaigns today led by youth.

The deepfake phenomenon is a new critical challenge they, and all participants in our democracies, now face.

Education for resilience

In Canada, Journalists for Human Rights announced a new program, funded by Heritage Canada, to train journalists and to enhance "citizen preparedness against online manipulation and misinformation."

Educators can play a key role in fostering youth agency to detect deepfakes and reduce their influence. One challenge is ensuring youth learn critical media literacy skills while they continue to explore valuable resources online and build their capacities and knowledge to participate in democratic structures.

Following steps I have identified in the "Get Ready to Act Against Social Media Propaganda" model —beginning with explaining stances on a controversial issue targeted through social media propaganda —educators can help youth discuss how they perceive and recognize deepfakes. They can explore the content's origins, who it's targeting, the reaction it's trying to achieve and who's behind it.

They can also discuss youth's role and responsibility to respond and stand up to disinformation and potential digital strategies to pursue in this process. A well-equipped generation of digital citizens could be our best bet.

Provided by The Conversation

Citation: Deepfakes: Informed digital citizens are the best defense against online manipulation (2020, January 8) retrieved 13 March 2024 from https://techxplore.com/news/2020-01-deepfakes-digital-citizens-defense-online.html