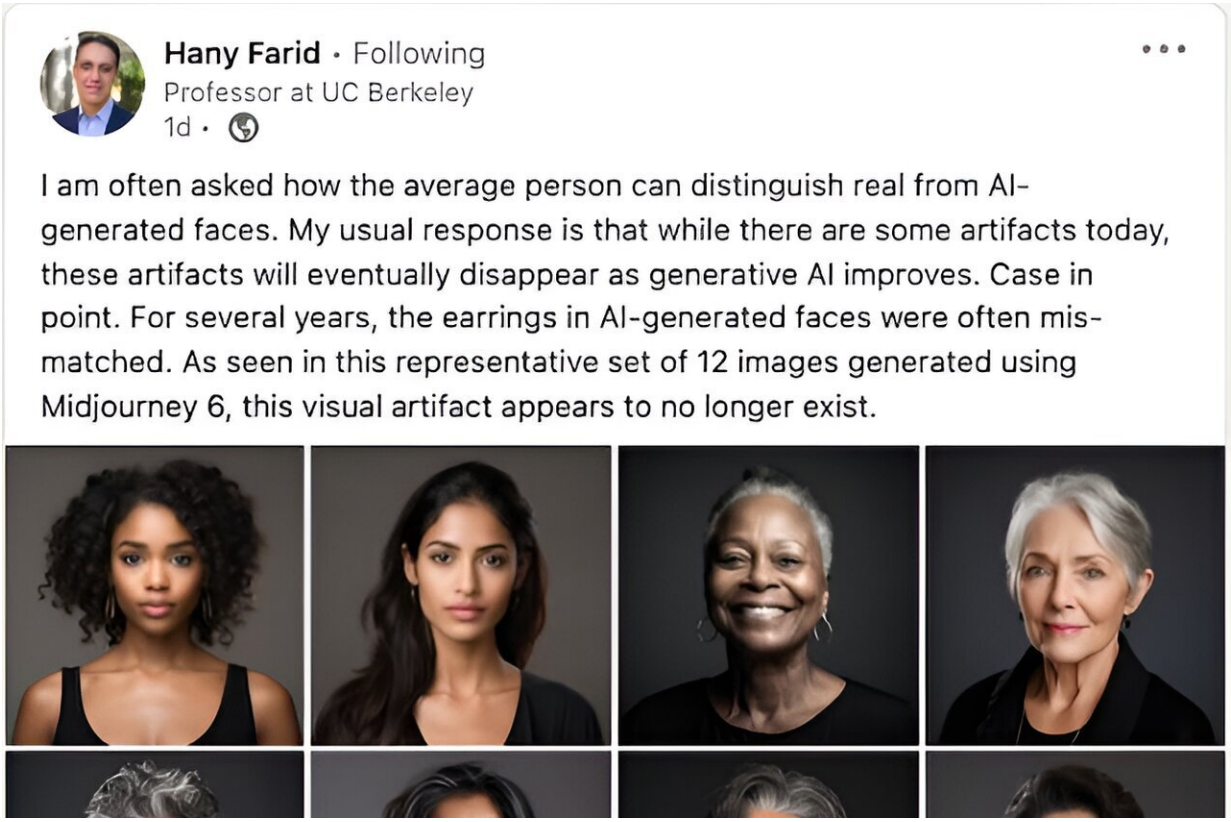
Deepfakes: How to empower youth to fight the threat of misinformation and disinformation

January 29 2024, by Nadia Naffi



Professor Hany Farid from University of California at Berkeley, expert in analyzing digital images and detecting digital manipulation shares, on LinkedIn, an example of the rapid evolution of the technology. Credit: Nadia Naffi

The World Economic Forum's Global Risks Report 2024 has issued a stark warning: misinformation and disinformation, primarily driven by deepfakes, are ranked as the most severe global short-term risks the world faces in the next two years.

In October 2023, the Innovation council of Québec shared the same realization after months of consultations with experts and the public.

This digital deception, which leverages artificial intelligence and, more recently generative AI, to create hyper-realistic fabrications, extends beyond being a technological marvel; it poses a profound societal threat.

In response to the gap in effectively combating deepfakes with technology and legislation alone, a research project led by my team and I sheds light on a vital solution: human intervention through education.

Technological solutions alone are inadequate

Despite ongoing development of deepfake detection tools, these technological solutions are racing to catch up with the rapidly advancing capabilities of deepfake algorithms.

Legal systems and governments are struggling to keep pace with this swift advancement of digital deception.

There is an urgent need for education to adopt a more serious, aggressive and strategic approach in equipping youth to combat this imminent threat.

Political disinformation concerns

The potential for political polarization is particularly alarming.

Nearly three billion people are expected to vote in countries including Bangladesh, India, Indonesia, Mexico, Pakistan, the United Kingdom and the United States within the next two years.

Disinformation campaigns threaten to undermine the legitimacy of newly elected governments.

Deepfakes of prominent figures like Palestinian American supermodel Bella Hadid and others have been manipulated to falsify their political statements, exemplifying the technology's capacity to sway public opinion and skew political narratives.

A deepfake of Greta Thunberg advocating for "vegan grenades" highlights the nefarious use of this technology.

Meta's unveiling of an AI assistant featuring celebrities' likenesses raises concerns about misuse and spreading disinformation.

Financial fraud, pornographic harms

Deepfake videos are also, unsurprisingly, being leveraged to commit financial fraud.

The popular YouTuber MrBeast was impersonated in a deepfake scam on TikTok, falsely promising an iPhone 15 giveaway that led to financial deceit.

These incidents highlight vulnerability to sophisticated AI-driven frauds and scams targeting people of all ages.

Deepfake pornography represents a grave concern for young people and adults alike, where individuals' faces are non-consensually superimposed onto explicit content. Sexually explicit deepfake images of Taylor Swift

spread on social media before platforms took them down. One was viewed over 45 million times.

Policy and technology approaches

Meta's policy now mandates political advertisers to disclose any AI manipulation in ads, a move mirrored by Google.

Neil Zhang, a Ph.D. student at the University of Rochester, is developing detection tools for audio deepfakes, including advanced algorithms and watermarking techniques.

The U.S. has introduced several acts: the Deepfakes Accountability Act of 2023, the No AI FRAUD Act safeguarding identities against AI misuse and the Preventing Deepfakes of Intimate Images Act targeting non-consensual pornographic deepfakes.

In Canada, legislators have proposed Bill C-27 and the Artificial Intelligence and Data Act (AIDA) which emphasize AI transparency and data privacy.

The United Kingdom adopted its Online Safety Bill. The EU recently announced a provisional deal surrounding its AI Act; the EU's AI Liability Directive addresses broader online safety and AI regulation issues.

The Indian government announced plans to draft regulations targeting deepfakes.

These measures reflect growing global commitments to curbing the pernicious effects of deepfakes. However, these efforts are insufficient to contain, let alone stop, the proliferation of deepfake dissemination.

Research study with youth

Research I have conducted with colleagues, funded by the Social Sciences and Humanities Research Council (SSHRC) and Canadian Heritage, unveils how empowering youth with digital agency can be a force against the rising tide of disinformation fueled by deepfake and artificial intelligence technologies.

Our study focused on how youth perceive the impact of deepfakes on critical issues and their own process of constructing knowledge in digital contexts. We explored their capacity and willingness to effectively counterbalance disinformation.

The study brought together Canadian university students, aged 18 to 24, for a series of hands-on workshops, in-depth individual interviews and focus group discussions.

Participants created deepfakes, gaining a firsthand understanding of easy access to and use of this technology and its potential for misuse. This experiential learning proved invaluable in demystifying how easily deepfakes are generated.

Participants initially perceived deepfakes as an uncontrollable and inevitable part of the digital landscape.

Through engagement and discussion, they went from being passive deepfake bystanders to developing a deeper realization of their grave threat. Critically, they also developed a sense of responsibility in preventing and mitigating deepfakes' spread, and a readiness to counter deepfakes.

Students shared recommendations for concrete actions, including urging educational systems to empower youth and help them recognize their

actions can make a difference. This includes:

- teaching the detrimental effects of disinformation on society;
- providing spaces for youth to reflect on and challenge societal norms, inform them about social media policies and outlining permissible and prohibited content;
- training students in recognizing deepfakes through exposure to the technology behind them;
- encouraging involvement in meaningful causes while staying alert to disinformation and guiding youth in respectfully and productively countering disinformation.

Multifaceted strategy needed

Based on our research and the participants' recommendations, we propose a multifaceted strategy to counter the proliferation of deepfakes.

Deepfake education needs to be integrated into educational curricula, along with nurturing critical thinking and digital agency in our youth. Youth need to be encouraged in active, yet safe, well-informed and strategic, participation in the fight against malicious deepfakes in digital spaces.

We emphasize the importance of hands-on collaborative learning experiences. We also advocate for an interdisciplinary educational approach that marries technology, psychology, media studies and ethics to fully grasp the implications of deepfakes.

The human element

Our research underscores a crucial realization: The human element,

particularly the role of education, is indispensable in the fight against deepfakes. We cannot rely solely on technology and legal fixes.

By equipping younger generations, but also every single member of our society, with the skills to critically analyze and challenge disinformation, we are nurturing a digitally literate society resilient enough to withstand the manipulative power of deepfakes.

To do so, we must equip people to understand they have roles and agency in safeguarding the integrity of our digital world.

Provided by The Conversation

Citation: Deepfakes: How to empower youth to fight the threat of misinformation and disinformation (2024, January 29) retrieved 29 April 2024 from https://techxplore.com/news/2024-01-deepfakes-empower-youth-threat-misinformation.html