Expert discusses how deepfakes are being used

August 1 2024, by Jodi Heckel

Digital humanities librarian Mary Ton is surrounded by some infamous deepfakes and images created by generative AI from recent years, as well as a magic lantern projector, which was one of the earliest ways to share images with a larger audience. Ton spoke about the challenges and benefits of deepfakes and how to spot them. Credit: Fred Zwicky

Digital tools can alter images of ourselves and others in ways that are more convincing and harder to detect than ever before. They can create spectacular special effects in movies. But they are also used with the images or voices of people without their consent, or to create propaganda with the intent to fool people.

Mary Ton is a professor and the digital humanities librarian for the University Library, and her expertise includes the use of artificial intelligence in the humanities. She talked with News Bureau arts and humanities editor Jodi Heckel about the use of deepfakes.

What exactly are deepfakes, and how are they created?

Deepfakes get their name from deep learning, a method in computer science that's used to study patterns in large quantities of images, sound and video. These patterns are represented through mathematical equations that are so complex that they resemble the dense networks of neurons in the human brain, making these algorithms artificially intelligent.

When given images of humans, computers look for patterns in pixels, the basic building blocks of a digital image. For sound, computers transform sound waves into visual representations, then use image recognition techniques to identify patterns in much the same way that they approach pictures. Deepfakes apply what they've learned about how humans look and sound to create new arrangements of pixels that mimic the human face and voice.

Manipulating images is nothing new. From the highly stylized statues of Egyptian pharaoh Akhenaten to chemically altered Victorian photos of women's waistlines, we've been using visual representations to project

our ideal selves. Digital technologies introduce an unprecedented level of sophistication into these idealized images, making it difficult to determine how much an image has been modified.

How are deepfakes used in positive and sinister ways in the entertainment industry?

At their best, deepfakes help us communicate. If you've watched "Indiana Jones and the Dial of Destiny," you've seen how special effects artists can use deepfakes to digitally de-age actors, and Ryan Reynold's beefy doppelganger in "Free Guy" is a hilarious example of how they can superimpose the head of one actor onto the body of another.

There are YouTube channels devoted to deepfaked celebrities doing everyday things, like this one of Keanu Reeves hyping himself up to make a phone call. Deepfake technologies also have been used extensively in the dubbing process to match the actor's lip movements to the audio track, creating a more seamless experience across languages.

The Center for Innovation in Teaching and Learning at Illinois has been exploring educational applications for deepfakes by helping faculty create AI avatars with HeyGen.

At their worst, deepfakes undermine performers' livelihoods and well-being. The voice-acting industry has been particularly hard hit because companies are using AI-generated narration instead of hiring actors. Even worse, actresses and Twitch streamers are finding out that their likenesses have been used to create pornographic content without their consent.

While deepfakes are opening up new creative possibilities, these applications are compromising opportunities for artists to contribute to

the entertainment industry on their terms.

What do you expect to see regarding the use of political deepfakes during this presidential election cycle?

Voters across the political spectrum need to be especially vigilant this election season when they encounter images, video and sound recordings of candidates. We've already seen two prominent examples of how deepfakes have been used to misinform and mislead voters.

During the primaries, a deepfaked Joe Biden called New Hampshire residents to discourage them from voting, and the BBC reported that AI-generated images of Donald Trump were being used by political advocacy groups to court Black voters.

Social media provides an international platform to share deepfaked content quickly and anonymously, so be especially wary of the content that you're encountering on Facebook, X, Tik Tok and Instagram.

At the same time, we're seeing legislators adopt AI to promote conversations about accessibility and inclusion. Rep. Jennifer Wexton of Virginia delivered an address using an AI-generated version of her voice to foreground how adaptive technologies support people with neurological disorders like hers. I think that we're going to see more candidates address AI and suggest policies to address its impact on political discourse, art and privacy.

Is there any regulation of deepfake images or videos? What are some of the issues with regulating them?

There's not yet a unified approach to addressing the spread of deepfaked

content, but we're seeing increased momentum within the EU and the U.S. to label images and video created with generative AI tools. So far, this movement has been led by Meta through Instagram and Facebook, as well as prominent news agencies. However, it's difficult to create firm guidelines for implementing a strategy for labeling content consistently because photo manipulation tools like Photoshop are so common.

To complicate matters further, training AI on copyrighted content is legal. The courts have tended to characterize this as a fair use of the material because it transforms the content into an algorithmic representation of patterns. Access to content can be limited by licensing agreements and terms of service.

If you are a content creator, it's especially important for you to review a website's terms of service to see if they permit data mining. If the site does, then anything you post can be used to train generative AI tools. Some sites like DeviantArt and Facebook allow you to opt out. We're also seeing the rise of tools to protect artists and their work. Programs like Glaze and AntiFake introduce noise into images and audio to disrupt the deep learning process, making protected content more difficult to deepfake.

How can someone detect a fake image or video?

The best detector is you! Automated detection is difficult because these tools rely on the same pattern recognition techniques that are used to create deepfakes in the first place. I recommend using the SILL method to evaluate content: stop, investigate, look and listen.

- Stop: Deepfakes are designed to make you react. Pausing before you re-share or act on content gives you time to evaluate the content.
- Investigate the source: Using reverse image searches like Google

Lens or TinEye can help you identify the original source of an image.

- Look for imperfections: Deepfakes tend to be too perfect, so keep an eye out for things like blur, camera movement and shadows. Limited use of hand gestures or an unexpected number of fingers can be signs of an AI-generated image.
- Listen: Deepfaked voices often sound monotone, lacking the rise and fall in pitch of human speech.

Provided by University of Illinois at Urbana-Champaign