Deepfakes are being used for good—here's how

November 7 2022, by Dominic Lees

In the second season of BBC mystery thriller [The Capture](#), deepfakes threaten the future of democracy and UK national security. In a dystopia set in present day London, hackers use AI to insert these highly realistic false images and videos of people into live news broadcasts to destroy the careers of politicians.

But my team's research has shown how difficult it is to create convincing deepfakes in reality. In fact, technology and creative professionals have started collaborating on solutions to help people spot bogus videos of politicians and celebrities. We stand a decent chance of staying one step ahead of fraudsters.

In my research project, [Virtual Maggie](#), I attempted to use deepfakes to digitally resurrect former UK prime minister Margaret Thatcher for a new drama. After months of work, we were unable to create a virtual Maggie acceptable for broadcast.

Producing convincing deepfakes in [high definition](#) requires top spec hardware, a lot of computer time, and human intervention to fix glitches in the output. This didn't stop me enjoying The Capture, despite knowing that Ben Chanan's drama was not a scenario likely to play out in the near future. Like every good dystopia, it had the seed of something that might one day be possible.

The use of deepfakes since they began in 2017 has been shocking. The majority of deepfakes on the internet are assaults on women, grabbing

facial images without consent and inserting them into pornographic content. Deepfakes expert [Henry Ajder](#) found that 96% of deepfakes found online were pornographic, and 100% of these were video images of women.

The premise of The Capture is grounded in facts. Deepfakes threaten democracy. In the 2019 UK [general election](#), artist [Bill Posters](#) released a provocative video of Boris Johnson saying we should vote for Jeremy Corbyn.

Posters' deepfake was a lot more convincing than the glitchy Russian deepfake showing Ukraine President Volodymyr Zelenskyy asking his troops to surrender. Yet, unlike the Kremlin, the British artist made it obvious his AI Boris was unreal by having "Boris" direct viewers to a website about deepfakes. He aimed to highlight our vulnerability to faked political propaganda.

Deepfakes may not yet be typically convincing enough to fool people. But [creative work](#) typically involves an unwritten agreement between creator and viewers to suspend their disbelief.

A fork in the road

The threat from deepfakes has led to an intensive search for tech solutions. A coalition of companies has formed the [Content Authenticity Initiative](#) (CAI) to provide "a way to evaluate truth in the media presented to us".

It's a promising approach. CAI collaborators and technology companies Truepic and Qualcomm have created a system that embeds the history of an image in its metadata so it can be verified. US photographer [Sara Naomi Lewkowicz](#) has completed an [experimental project](#) with CAI that [embeds source information](#) in her photos.

But creative and technology professionals don't necessarily want to trammel the emerging technology of deepfakes. Researchers at the Massachusetts Institute of Technology Media Lab have been brainstorming ways of putting deepfakes to good use. Some of these are in healthcare and treatment.

Research engineers Kate Glazko and Yiwei Zheng are using deepfakes to help people with aphantasia, the inability to create mental images in your mind. The breakup simulator, under development, aims to use deepfakes to "alleviate the anxiety of difficult conversation through rehearsal".

The most profound positive uses for deepfakes include campaigns for political change. The parents of Joaquin Oliver, killed in a high school shooting in Florida in 2018, used the technology to bring him back in a forceful video calling for gun control.

Getting creative

There are also cultural applications of deepfakes. At the Dali Museum in Florida, a deepfake Salvador Dali welcomes visitors, telling them about himself and his art. Researcher Mihaela Mihailova says this gives visitors "a sense of immediacy, closeness, and personalization". Deepfake Dali even offers you the chance to take a selfie with him.

Deepfakes and AI-generated characters can be educational. In Shanghai, during lockdown, Associate Professor Jiang Fei noticed his students' attention dropped during online lessons. To help them focus better he used an anime version of himself to front his teaching. Jiang Fei said: "The enthusiasm of the students in class, and the improvement of the quality of homework have made obvious progress."

Channel Four used its 2020 alternative Christmas message to entertain

viewers with a deepfaked queen, while making a serious point about not trusting everything we see on video.

A growing network of film producers, researchers and AI technologists in the UK, hosted by the University of Reading and funded by [the Alan Turing Institute](#), is seeking to harness the positive potential of deepfakes in creative screen production. Filmmaker Benjamin Field told the group during a workshop how he used deepfakes to "resurrect" the animator who created Thunderbirds for [Gerry Anderson: A Life Uncharted](#), a documentary about the troubled life of the kids' TV hero.

Field and his co-producer, Anderson's youngest son Jamie, discovered old audio tapes and used deepfakes to construct a "filmed" interview with the famous puppeteer. Field is among a small group of creatives determined to find positives ways of using deepfakes in broadcasting.

Deepfakes and AI characters are part of our future and the above examples show how this could be at least partly positive. But we also need laws to protect people whose images are stolen or abused, and ethical guidelines on how deepfakes are used by filmmakers. Responsible producers have already formed a partnership on AI, and drafted a [code of conduct](#) which could help avert the catastrophic vision of the future that we saw in the The Capture.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation