



How Deepfakes Are Impacting Society

Image: Actor Jordan Peele and his
deepfake Barack Obama

Fake online video and audio content has become a powerful tool for spreading political misinformation and harming personal reputations. By Morgan Currie

Deepfakes are the result of machine-learning systems that manipulate the content of one piece of media by transferring it to another. The AI ingests video, photographs or audio of a person or object, then learns to mimic its behaviour and output the results onto another target person or object, creating an eerily accurate counterfeit.

The deep-learning software used to make deepfakes has become cheap and accessible, raising questions about the potential for abuse. While there are plenty of examples that are benign and playful – Salvador Dali taking selfies with museum patrons¹, for instance – the origins of the technology show how harmful it can be.

The term first became widely used in 2017, after a Reddit user by the name 'Deepfakes' posted pornographic videos featuring actresses whose faces were digitally altered to resemble female celebrities, such as Scarlett Johansson and Gal Gadot. For many, the videos crossed basic lines governing consent and harassment and showcased a potent new tool for revenge porn². That 'Deepfakes' used Google's free open-source machine-learning software also drove home how easily a hobbyist,

or anyone with an interest in the technology, could masquerade falsehoods as reality.

Since then, other examples of disturbing deepfakes include a video of US House of Representatives Speaker Nancy Pelosi³ altered to make her sound drunk – it circulated widely after Donald Trump posted it and Facebook refused to take it down – and a video by two artists of Facebook CEO Mark Zuckerberg confessing that his company "really owns the future". Eric Adams confessed to using deepfake audio technology to speak Spanish and Mandarin to constituents in robocalls about local events. There's also a whole industry of livestreaming 24/7 deepfakes of influencers in China⁴.

Actor Jordan Peele used a deepfake Barack Obama⁵ to warn of the dangers of deepfakes, highlighting how they can distort reality in ways that could undermine people's faith in trusted media sources and incite toxic behaviour. Some in the intelligence community have warned that foreign governments could spread deepfakes to disrupt⁶ or sway⁷ elections.

Meanwhile, the vast majority of deepfakes are of non-consensual porn⁸, not misinformation, which raises another set of legal and ethical concerns. According to WIRED, 2023 has seen the largest amount of new deepfake online porn⁹ - more than the total of other years combined – and a corresponding rise in non-consensual content in circulation. Victims become the targets of gender-based online harassment, and family members can wind up seeing the images and videos.

Social media companies have started to address the deepfake dilemma – Facebook set up a public contest in 2019 to help it develop models to detect deepfakes and banned them in early 2020¹⁰, in anticipation of the damage that could be done in an election year. Twitter now deletes reported deepfakes and blocks any of their publishers.

Governments are also putting forward laws to curb the technology. California passed a 2019 law banning deepfakes altogether, and in December 2020 the US Congress passed into law the Identifying Outputs of Generative Adversarial Networks Act¹¹.

The entertainment industry has responded coolly to these protections, claiming too much oversight clamps down on free speech rights. In 2018, Walt Disney Company's Vice President of Government Relations, Lisa Pitney, wrote¹² that a proposed New York law that included controls on the use of "digital replicas", would "interfere with the right and ability of companies like ours to tell stories about real people and events. The public has an interest in those stories, and the First Amendment protects those who tell them."

Others feel such legislation is not going far enough. Existing laws put the burden on users to identify deepfakes, exonerating the platforms they circulate on. Social media companies remain exempt from regulations¹³ and no industry-wide standards currently exist, keeping them off the hook for now.

References:

1. <https://newatlas.com/salvador-dali-ai-generated-computer-museum-exhibition/59641/>
2. <https://www.wired.com/story/deepfake-porn-harms-adult-performers-too/>
3. <https://www.youtube.com/watch?v=sD0o5nDJwgA>
4. <https://www.technologyreview.com/2023/09/19/1079832/chinese-ecommerce-deepfakes-livestream-influencers-ai/>
5. <https://www.youtube.com/watch?v=cQ54GDm1eL0>
6. <https://thehill.com/policy/cybersecurity/446611-house-intel-to-examine-deepfake-videos-in-june>
7. <https://www.cnbc.com/2019/10/15/deepfakes-could-be-problem-for-the-2020-election.html>
8. <https://www.vice.com/en/article/7x57v9/most-deepfakes-are-porn-harassment-not-fake-news>
9. <https://www.wired.co.uk/article/deepfake-porn-is-out-of-control>
10. <https://www.vice.com/en/article/8xwqp3/facebook-deepfake-detection-challenge-dataset>
11. <https://www.congress.gov/bill/116th-congress/senate-bill/2904>
12. https://www.rightofpublicityroadmap.com/sites/default/files/pdfs/disney_opposition_letters_a8155b.pdf
13. <https://www.wired.com/story/opinion-californias-anti-deepfake-law-is-far-too-feeble/>

Cite as Morgan Currie (2023). How Deepfakes Are Impacting Society. *The New Real Magazine*, Edition One. pp 77-78. <https://doi.org/10.2218/newreal.9266>