# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

> *"This is developing more rapidly than I thought. Soon, it's going to get to the point where there is no way that we can actually detect [deepfakes] anymore, so we have to look at other types of solutions."*
>
> Hao Li
> Deepfake Pioneer & Associate Professor

Novel digital technologies make it increasingly difficult to distinguish between real and fake media. One of the most recent developments contributing to the problem is the emergence of deepfakes which are hyper-realistic videos that apply artificial intelligence (AI) to depict someone say and do things that never happened. Coupled with the reach and speed of social media, convincing deepfakes can quickly reach millions of people and have negative impacts on our society. While scholarly research on the topic is sparse, this study analyzes 84 publicly available online news articles to examine what deepfakes are and who produces them, what the benefits and threats of deepfake technology are, what examples of deepfakes there are, and how to combat deepfakes. The results suggest that while deepfakes are a significant threat to our society, political system and business, they can be combatted via legislation and regulation, corporate policies and voluntary action, education and training, as well as the development of technology for deepfake detection, content authentication, and deepfake prevention. The study provides a comprehensive review of deepfakes and provides cybersecurity and AI entrepreneurs with business opportunities in fighting against media forgeries and fake news.

## Introduction

In recent years, fake news has become an issue that is a threat to public discourse, human society, and democracy (Borges et al., 2018; Qayyum et al., 2019). Fake news refers to fictitious news style content that is fabricated to deceive the public (Aldwairi & Alwahedi, 2018; Jang & Kim, 2018). False information spreads quickly through social media, where it can impact millions of users (Figueira & Oliveira, 2017). Presently, one out of five Internet users get their news via YouTube, second only to Facebook (Anderson, 2018). This rise in popularity of video highlights the need for tools to confirm media and news content authenticity, as novel technologies allow convincing manipulation of video (Anderson, 2018). Given the ease in obtaining and spreading misinformation through social media platforms, it is increasingly hard to know what to trust, which results in harmful consequences for informed decision making, among other things (Borges et al., 2018; Britt et al., 2019). Indeed, today we live in what some have called a "post-truth" era, which is characterized by digital disinformation and information warfare led by malevolent actors running false information campaigns to manipulate public opinion (Anderson, 2018; Qayyum et al., 2019; Zannettou et al., 2019).

Recent technological advancements have made it easy to create what are now called "deepfakes", hyper-realistic videos using face swaps that leave little trace of manipulation (Chawla, 2019). Deepfakes are the product of artificial intelligence (AI) applications that merge, combine, replace, and superimpose images and video clips to create fake videos that appear authentic (Maras & Alexandrou, 2018). Deepfake technology can generate, for example, a humorous, pornographic, or political video of a person saying anything, without the consent of the person whose image and voice is involved (Day, 2018; Fletcher, 2018). The game-changing factor of deepfakes is the scope, scale, and sophistication of the technology involved, as almost anyone with a computer can fabricate fake videos that are practically indistinguishable from authentic media (Fletcher, 2018). While early examples of deepfakes focused on political leaders, actresses, comedians, and entertainers having their faces weaved into porn videos (Hasan & Salah, 2019), deepfakes in the future will likely be more and more used for revenge porn, bullying, fake video evidence in courts, political sabotage, terrorist propaganda, blackmail, market manipulation, and fake news (Maras & Alexandrou, 2019).

While spreading false information is easy, correcting the record and combating deepfakes are harder (De

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

keersmaecker & Roets, 2017). In order to fight against deepfakes, we need to understand deepfakes, the reasons for their existence, and the technology behind them. However, scholarly research has only recently begun to address digital disinformation in social media (Anderson, 2018). As deepfakes only surfaced on the Internet in 2017, scholarly literature on the topic is sparse. Hence, this study aims to discuss what deepfakes are and who produces them, what the benefits and threats of deepfake technology are, some examples of current deepfakes, and how to combat them. In so doing, the study analyzes a number of news articles on deepfakes drawn from news media websites. The study contributes to the nascent literatures of fake news and deepfakes both by providing a comprehensive review of deepfakes, as well as rooting the emerging topic into an academic debate that also identifies options for politicians, journalists, entrepreneurs, and others to combat deepfakes.

The article is organized as follows. After the introduction, the study explains data collection and news article analysis. The study then puts forward four sections that review deepfakes, what the potential benefits of deepfake technology are, who the actors involved in producing deepfakes are, and the threats of deepfakes to our societies, political systems, and businesses. Thereafter, two sections provide examples of deepfakes and discuss four feasible mechanisms to combat deepfakes. Finally, the study concludes with implications, limitations, and suggestions for future research.

## Method

This study relies on the emerging scholarly literature and publicly available news articles on deepfakes. A total of 84 articles from 11 news companies' websites were collected in August 2019 for the purpose of conducting empirical analysis on how the news media has discussed deepfakes. All articles focused on deepfakes, were written in English and were published in 2018-2019. They were found through Google News search, using keywords "deepfake", "deep fake", and the corresponding plural forms. Once an article was found, a similar search was performed using the news website's own search option to find more articles by that particular media source. The focus of the selected news media ranged from general daily news to concentration on business or technology news. The dataset includes 2 to 16 news articles on deepfakes from each news company. The articles were coded with a short identifier for citing purposes, then analyzed via content analysis with focus on what deepfakes are, who produces them, what the benefits and threats of deepfake technology are, some current examples of deepfakes, and how to combat them. Table 1 in the appendix shows the news articles, their authors, news companies, and publication dates; the article titles are shortened due to space limitations.

## What are Deepfakes?

A combination of "deep learning" and "fake", deepfakes are hyper-realistic videos digitally manipulated to depict people saying and doing things that never actually happened (CNN03; FRB04). Deepfakes rely on neural networks that analyze large sets of data samples to learn to mimic a person's facial expressions, mannerisms, voice, and inflections (CBS02; PCM10). The process involves feeding footage of two people into a deep learning algorithm to train it to swap faces (PCM01). In other words, deepfakes use facial mapping technology and AI that swaps the face of a person on a video into the face of another person (FOX09; PCM03). Deepfakes surfaced to publicity in 2017 when a Reddit user posted videos showing celebrities in compromising sexual situations (FRB01; FRB08; USAT03). Deepfakes are difficult to detect, as they use real footage, can have authentic-sounding audio, and are optimized to spread on social media quickly (FRB05; WP01). Thus, many viewers assume that the video they are looking at is genuine (CNET01; CNN10).

Deepfakes target social media platforms, where conspiracies, rumors, and misinformation spread easily, as users tend to go with the crowd (CNET05; FOX06). At the same time, an ongoing 'infopocalypse' pushes people to think they cannot trust any information unless it comes from their social networks, including family members, close friends or relatives, and supports the opinions they already hold (CNN06). In fact, many people are open to anything that confirms their existing views even if they suspect it may be fake (GRD09). Cheap fakes, that is, low-quality videos with slightly doctored real content, are already everywhere because low-priced hardware such as efficient graphical processing units are widely available (CBS01; CNN08). Software for crafting high-quality, realistic deepfakes for disinformation is increasingly available as open source (FOX05; FT02; PCM04). This enables users with little technical skills and without any artistic expertise to near-perfectly edit videos, swap faces, alter expressions, and synthesize speech (CNET08; GRD10).

As for technology, deepfakes are the product of Generative Adversarial Networks (GANs), namely two artificial neural networks working together to create

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

real-looking media (CNN03). These two networks called 'the generator' and 'the discriminator' are trained on the same dataset of images, videos, or sounds (GRD03). The first then tries to create new samples that are good enough to trick the second network, which works to determine whether the new media it sees is real (FBR07). That way, they drive each other to improve (PCM05). A GAN can look at thousands of photos of a person, and produce a new portrait that approximates those photos without being an exact copy of any one of them (GRD07). In the near future, GANs will be trained on less information and be able to swap heads, whole bodies, and voices (GRD08; USAT01). Although deepfakes usually require a large number of images to create a realistic forgery, researchers have already developed a technique to generate a fake video by feeding it only one photo such as a selfie (CBS03; CNET07).

**The Benefits of Deepfake Technology**

Deepfake technology also has positive uses in many industries, including movies, educational media and digital communications, games and entertainment, social media and healthcare, material science, and various business fields, such as fashion and e-commerce (FRB04).

The film industry can benefit from deepfake technology in multiple ways. For example, it can help in making digital voices for actors who lost theirs due to disease, or for updating film footage instead of reshooting it (FRB01; PCM10). Movie makers will be able to recreate classic scenes in movies, create new movies starring long-dead actors, make use of special effects and advanced face editing in post-production, and improve amateur videos to professional quality (FOX05; GRD07). Deepfake technology also allows for automatic and realistic voice dubbing for movies in any language (PCM09; USAT04), thus allowing diverse audiences to better enjoy films and educational media. A 2019 global malaria awareness campaign featuring David Beckham broke down language barriers through an educational ad that used visual and voice-altering technology to make him appear multilingual (USAT03). Similarly, deepfake technology can break the language barrier on video conference calls by translating speech and simultaneously altering facial and mouth movements to improve eye-contact and make everyone appear to be speaking the same language (CNET05; FRB03; FT03).

The technology behind deepfakes enables multiplayer games and virtual chat worlds with increased telepresence (CNET07), natural-sounding and -looking smart assistants (PCM09) and digital doubles of people. This helps to develop better human relationships and interaction online (CBS03; FRB02). Similarly, the technology can have positive uses in the social and medical fields. Deepfakes can help people deal with the loss of loved ones by digitally bringing a deceased friend "back to life", and thereby potentially aiding a grieving loved one to say goodbye to her (FOX05; PCM10). Further, it can digitally recreate an amputee's limb or allow transgender people to better see themselves as a preferred gender (USAT04). Deepfake technology can even help people with Alzheimer's interact with a younger face they may remember (FOX05). Scientists are also exploring the use of GANs to detect abnormalities in X-rays (CNET04) and their potential in creating virtual chemical molecules to speed up materials science and medical discoveries (GRD03).

Businesses are interested in the potential of brand-applicable deepfake technology, as it can transform e-commerce and advertising in significant ways (FRB02). For example, brands can contract supermodels who are not really supermodels, and show fashion outfits on a variety of models with different skin tones, heights, and weights (FRB07). Further, deepfakes allow for superpersonal content that turns consumers themselves into models; the technology enables virtual fitting to preview how an outfit would look on them before purchasing and can generate targeted fashion ads that vary depending on time, weather, and viewer (FRB02; FRB07). An obvious potential use is being able to quickly try on clothes online; the technology not only allows people to create digital clones of themselves and have these personal avatars travel with them across e-stores, but also to try on a bridal gown or suit in digital form and then virtually experience a wedding venue (FRB02). Also, AI can provide unique artificial voices that differentiate companies and products to make branding distinction easier (PCM10).

**Who Produces Deepfakes?**

There are at least four major types of deepfake producers: 1) communities of deepfake hobbyists, 2) political players such as foreign governments, and various activists, 3) other malevolent actors such as fraudsters, and 4) legitimate actors, such as television companies.

Individuals in deepfake hobby communities are difficult to track down (FRB06). After the introduction of celebrity porn deepfakes to Reddit by one user in late 2017, it only took a few months for a newly founded deepfake hobbyist community to reach 90,000 members

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

(CBS01; GRD08). Many hobbyists focus on porn-related deepfakes (USAT01), while others place famous actors in films in which they never appeared to produce comic effects (GRD05). Overall, hobbyists tend to see AI-crafted videos as a new form of online humor, and contribution to the development of such technology as solving an intellectual puzzle, rather than as a way to trick or threaten people (CNN07; GRD05). Their deepfakes are meant to be entertaining, funny, or politically satirical, and can help with gaining followers on social media (FOX01). Some hobbyists may be looking for more concrete personal benefits, such as raising awareness about the potential of deepfake technology in order to get deepfake-related paid work, for example, with music videos or television shows (GRD02). Thus, hobbyists and legitimate actors such as television companies may collaborate with each other.

While meme-like deepfakes by hobbyists can entertain online users, more malicious actors are also involved. Various political players, including political agitators, hacktivists, terrorists, and foreign states can use deepfakes in disinformation campaigns to manipulate public opinion and undermine confidence in a given country's institutions (CBS01; CBS02). In these times of hybrid warfare, deepfakes are weaponized disinformation aimed at interfering with elections and sowing civil unrest (CNET12). We may anticipate more and more domestic and foreign state-funded Internet "troll farms" that use AI to create and deliver political fake videos tailored to social media users' specific biases (CNN06). Deepfakes are also increasingly being deployed by fraudsters for the purpose of conducting market and stock manipulation, and various other financial crimes (USAT03). Criminals have already used AI-generated fake audios to impersonate an executive on the phone asking for an urgent cash transfer (CNN01; FT01). In the future, video calls will also be able to be faked in real-time. Visual materials required to produce impersonations of executives are often available on the Internet. Deepfake technology can make use of visual and audio impersonations of executives from, for example, TED Talk videos available on YouTube (WP01).

**The Possible Threats of Deepfakes**

Deepfakes are a major threat to our society, political system, and business because they 1) put pressure on journalists struggling to filter real from fake news, 2) threaten national security by disseminating propaganda and interfering in elections, 3) hamper citizen trust toward information by authorities, and, 4) raise cybersecurity issues for people and organizations.

It is highly probably that the journalism industry is going to have to face a massive consumer trust issue due to deepfakes (USAT01). Deepfakes pose a greater threat than "traditional" fake news because they are harder to spot and people are inclined to believe the fake is real (CNN06). The technology allows the production of seemingly legitimate news videos that place the reputation of journalists and the media at risk (USAT01). Also, winning the race to access video footage shot by the witness of an incident can provide competitive advantage to a news media outlet, while danger rises if the offered footage is fake. During the spike in tensions between India and Pakistan in 2019, Reuters found 30 fake videos on the incident; mostly old videos from other events posted with new captions (DD02). Misattributed video footage such as a real protest march or violent skirmish captioned to suggest it happened somewhere else is a growing problem, and will be augmented by the rise of deepfakes (WP01). While looking for eyewitness videos about the mass shooting in Christchurch, New Zealand, Reuters came across a video which claimed to show the moment a suspect was shot dead by police. However, they quickly discovered it was from a different incident in the U.S.A., and the suspect in the Christchurch shooting was not killed (DD02).

The intelligence community is concerned that deepfakes will be used to threaten national security by disseminating political propaganda and disrupting election campaigns (CNET07; CNN10). U.S. intelligence officials have repeatedly warned about the threat of foreign meddling in American politics, especially in the lead-up to elections (CNN02; CNET04). Putting words in someone's mouth on a video that goes viral is a powerful weapon in today's disinformation wars, as such altered videos can easily skew voter opinion (USAT02; WP02). A foreign intelligence agency could produce a deepfake a video of a politician using a racial epithet or taking a bribe, a presidential candidate confessing complicity in a crime, or warning another country of an upcoming war, a government official in a seemingly compromising situation, or admitting a secret plan to carry out a conspiracy, or U.S. soldiers committing war crimes such as killing civilians overseas (CBS02; CNN06; FOX06). While such faked videos would likely cause domestic unrest, riots, and disruptions in elections, other nation states could even choose to act out their foreign policies based on unreality, leading to international conflicts (CBS03).

Deepfakes are likely to hamper digital literacy and citizens' trust toward authority-provided information, as fake videos showing government officials saying

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

things that never happened make people doubt authorities (CNET11; FOX10). Indeed, people nowadays are increasingly affected by AI-generated spam, and by fake news that builds on bigoted text, faked videos, and a plethora of conspiracy theories (GRD06). Nonetheless, the most damaging aspect of deepfakes may not be disinformation per se, but rather how constant contact with misinformation leads people to feel that much information, including video, simply cannot be trusted, thereby resulting in a phenomenon termed as "information apocalypse" or "reality apathy" (CNN01; GRD07). Further, people may even dismiss genuine footage as fake (CBS02), simply because they have become entrenched in the notion that anything they do not want to believe must be fake (CNET05). In other words, the greatest threat is not that people will be deceived, but that they will come to regard everything as deception (GRD07).

Cybersecurity issues constitute another threat imposed by deepfakes. The corporate world has already expressed interest in protecting themselves against viral frauds, as deepfakes could be used for market and stock manipulation, for example, by showing a chief executive saying racist or misogynistic slurs, announcing a fake merger, making false statements of financial losses or bankruptcy, or portraying them as if committing a crime (CNN02; FRB04; WP01). Further, deepfaked porn or product announcements could be used for brand sabotage, blackmail, or to embarrass management (FRB06; PCM03). In addition, deepfake technology enables real-time digital impersonation of an executive, for example, to ask an employee to perform an urgent cash transfer or provide confidential information (CNN01; FT01; PCM03). Further, deepfake technology can create a fraudulent identity and, in live-stream videos, convert an adult face into a child's or younger person's face, raising concerns about the use of the technology by child predators (FOX06). Lastly, deepfakes can contribute to the spread of malicious scripts. Recently, researchers found that a website devoted to deepfakes used its visitors' computers to mine cryptocurrencies. Deepfake hobbyists may in this way become targets of 'cryptojacking' because they are likely to have powerful computers (CNET16).

**Current Examples of Deepfakes**

Most deepfakes today on social platforms like YouTube or Facebook can be seen as harmless fun or artistic works using dead or alive public figures. But there are also examples from the dark side of deepfakes, namely celebrity and revenge porn, as well as attempts at political and non-political influencing.

Many deepfakes focus on celebrities, politicians, and corporate leaders because the internet is packed with source photos and videos of them from which to build the large image stockpiles required to train an AI deepfake system (CNET08; PCM03). The majority of such deepfakes are goofs, pranks, and funny memes with comedic or satirical effect (CNET07; DD01). A deepfake might show, for example, Nicolas Cage acting in movies in which he has never starred in, such as Indiana Jones or Terminator 2 (CNET05; PCM10). Some intriguing examples of deepfakes include a video that replaces Alden Ehrenreich with young Harrison Ford in clips taken from Solo: A Star Wars Story, and a video of actor Bill Hader appearing on Late Night with David Letterman. While Hader talks about Tom Cruise, his face morphs into Cruise's (CNET01; FRB06). Some deepfakes show dead celebrities such as the band Queen's ex-vocalist Freddie Mercury's face imposed on that of actor Rami Malek's, along with the Russian mystic Grigori Rasputin singing Beyonce's powerful ballad "Halo" (FOX02). An art museum in the U.S. has used the technology to bring Salvador Dali back to life to greet visitors (DD01), and another AI system makes anyone dance like a prima ballerina by imposing a real dancer's moves onto a target person's body, thereby generating a video that shows the target as a professional dancer (CNET14; PCM05).

Examples of harmful deepfakes, however, are also popping up increasingly (FOX04). Deepfake technology enables celebrity and revenge porn, that is, involuntary pornography using images of celebrities and non-celebrities, which are shared on social networks without their consent (CNET07; CNET15). Thus, celebrities such as Scarlett Johansson have been featured on deepfaked adult movies, in which their faces have been superimposed over porn stars' faces (CNET08; PCM03). In the political scene, a 2018 deepfake created by Hollywood filmmaker Jordan Peele featured former US President Obama discussing the dangers of fake news and mocking the current president Trump (CBS01; CNN06). In 2019, an altered video of American politician Nancy Pelosi went viral and had massive outreach; the video was slowed down to make her sound intoxicated (CNET01; FRB06). In a 2018 deepfake video, Donald Trump offered advice to the people of Belgium about climate change. The video was created by a Belgian political party "sp.a" in order to attract people to sign an online petition calling on the Belgian government to take more urgent climate action. The video provoked outrage about the American president meddling in a foreign country with Belgium's climate policy (GRD07). In 2019, the U.S. Democratic Party deepfaked its own chairman Tom Perez to highlight the

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

potential threat of deepfakes to the 2020 election (CNN01).

While these are examples of limited political influencing, other deepfakes may have more lasting impact. In Central Africa in 2018, a deepfake of Gabon's long-unseen president Ali Bongo, who was believed in poor health or dead, was cited as the trigger for an unsuccessful coup by the Gabonese military. And in Malaysia, a viral clip deepfake of a man's confession to having sex with a local cabinet minister caused political controversy (WP01). Also non-political individuals have been used for creating deepfakes. In June 2019, a high-quality deepfake by two British artists featuring Facebook CEO Mark Zuckerberg racked up millions of views (CBS01). The video falsely portrays Zuckerberg giving respect to Spectre, a fictional evil organization from the James Bond series that teaches him how to take total control of billions of peoples' confidential data, and thus own their future (CNN04; FOX03; FRB05). Using news footage, deepfake technology, and a voice actor, the video was meant to show how technology can be used to manipulate data (CNN05).

**Methods to Combat Deepfakes**

The reviewed news articles suggest that there are four ways to combat deepfakes: 1) legislation and regulation, 2) corporate policies and voluntary action, 3) education and training, and 4) anti-deepfake technology that includes deepfake detection, content authentication, and deepfake prevention.

Legislation and regulation are both obvious means against deepfakes. At present, deepfakes are not specifically addressed by civil or criminal laws, although legal experts have suggested adapting current laws to cover libel, defamation, identity fraud, or impersonating a government official using deepfakes (FT02; WP01). Virginia state law against revenge porn recently made distributing "falsely created" images and videos a misdemeanor, and thus expanded the law to include deepfakes (CNET03). That said, the increasing sophistication of AI technologies calls for new types of laws and regulatory frameworks (GRD03). For example, deepfakes raise concerns about privacy and copyright, as the visual depictions of people on deepfake videos are not exact copies of any existing material, but rather new representations generated by AI (CNET03; GRD07). Thus, regulators must navigate a difficult legal landscape around free-speech and ownership laws to properly regulate the use of deepfake technology (FRB06).

On the other hand, an appropriate legal solution to the proliferation of harmful deepfakes would not be a complete ban on the technology, which would be unethical (USAT04). While new laws can be introduced to prevent deepfakes, they also need mechanisms of enforcement (FRB09). Today's social media firms enjoy broad immunity for the content that users post on their platforms (WP02). One legislative option could be to walk back social media firms' legal immunity from the content their users post, thus making not only users but also the platforms more responsible for posted material (CNET09). Nonetheless, legislation has had little effect on malevolent actors such as foreign states and terrorists, that may run massive disinformation campaigns against other states on social media platforms.

Corporate policies and voluntary action may provide more effective tools against deepfakes. For example, politicians can commit not to use illicit digital campaign tactics or spread disinformation such as deepfakes in their election campaigns (WP04). Social media companies need to enforce ethics and turn away from the fact that divisive content getting pushed to the top of the feed is financially a win because it maximizes engagement time for advertisements (GRD01). While few social media firms have policies yet about deepfakes, they should collaborate to prevent their platforms from being weaponized for disinformation, and thus proactively enforce transparent, shared policies to block and remove deepfakes (CNET10; FOX06; GRD04). Presently, many companies do not remove disputed content, rather they downrank it to make it more difficult to find, by being less prominent in users' news feeds (CNN04; FOX02; FOX03). On the other hand, the increase in hate speech, fake news, and disinformation polluting digital platforms has led some firms to take more action, such as suspending user accounts and investing in quicker detection technology (CNET03; CNN05; CNN06). Reddit and Pornhub have banned deepfake porn and other non-consensual pornography, and act upon users' flagging of such material (CNET15; FRB10; PCM07). Facebook cuts off any content identified as false or misleading by third-party fact-checkers from running ads and making money; the company works with over 50 fact-checking organizations, academics, experts, and policymakers to find new solutions (CNET06; CNET09; CNET11). Instagram's algorithms will not recommend people view content that is marked as "false" by fact checkers (CNN04). Among news media companies, Wall Street Journal and Reuters have formed corporate teams to help and train their reporters to identify fake content,

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

and to adopt detection techniques and tools such as cross-referencing location on Google maps and reverse image searching (DD01; DD02; CNN01).

Education and training are crucial for combatting deepfakes. Despite considerable news coverage and concerns presented by authorities, the threat of deep fakes has not yet been reckoned with by the public (FRB08). In general, there is a need to raise public awareness about AI's potential for misuse (FOX01). Whereas deepfakes provide cyber criminals new tools for social engineering, companies and organisations need to be on high alert and to establish cyber resilience plans (FT01). Governments, regulators, and individuals need to comprehend that video, contrary to appearances, may not provide an accurate representation of what happened, and know which perceptual cues can help to identify deepfakes (USAT01; WP01). It is recommended that critical thinking and digital literacy be taught in schools as these traits contribute to children's ability to spot fake news and interact more respectfully with each other online.

These skills likewise should also be promoted among the older, less technology-savvy population (GRD02; FOX06). The reason for this is that people need to be able to critically assess the authenticity and social context of a video they may wish to consume, as well as the trustworthiness of its source (that is, who shared the video and what does that say), in order to understand the video's real intent. It is also important to remember that quality is not an indicator of a video's authenticity (FOX04; FRB01). Also, people need to understand that as the technology develops, fewer photographs of real faces will be required to create deepfakes and that nobody is immune (FRB06). Anyone who posts a single selfie or a video capturing 30 frames per second on a social networking site is at risk of being deepfaked (USAT03). While the best method is keeping photos and videos off the internet, having obstructions such as waving hand in front of a face in a photo or on a video can provide some protection (CNET08). Companies, governments, and authorities using facial recognition technology and storing vast amounts of facial data for security and verification purposes, need to address the threat of identity theft if such data were to be leaked (FRB06).

Anti-deepfake technology provides perhaps the most varied set of tools to 1) detect deepfakes, 2) authenticate content, and 3) prevent content from being used to produce deepfakes. Overall, the problems of technology to authenticate content and identify fakes is doing it at scale, and the fact that there are far more available research resources and people working on developing technology to create deepfakes than on technology to detect them (CBS02; WP02). For instance, users upload 500 hours of content per minute on YouTube. Twitter struggles with 8 million accounts a week that attempt to spread content through manipulative tactics (PCM02; WP01). This creates massive challenges for technologies to go through all of the posted material in a short time. Further, deepfake developers tend to use results from published deepfake research to improve their technology and get around new detection systems (CNN06). For example, researchers found that early deepfake methods failed to mimic the rate at which a person blinks; whereas recent programs have fixed the lack of blinking or unnatural blinking after the findings were published (CNN03; GRD05). While funding for deepfake detection development mainly comes from national security agencies such as The Defense Advanced Research Projects Agency (DARPA), there are significant business opportunities for private cybersecurity companies to produce solutions for deepfake detection, build trusted platforms, weed out illicit bots, and fight against fraud and digital pollution (CBS03; FT01; FT02). However, the development of anti-deepfake technology alone is not enough. Organizations must also adopt these technologies; for example, the government in any given country can be modernized to face and help protect its citizens against deepfakes (WP03).

Media forensic experts have suggested subtle indicators to detect deepfakes, including a range of imperfections such as face wobble, shimmer and distortion; waviness in a person's movements; inconsistencies with speech and mouth movements; abnormal movements of fixed objects such as a microphone stand; inconsistencies in lighting, reflections and shadows; blurred edges; angles and blurring of facial features; lack of breathing; unnatural eye direction; missing facial features such as a known mole on a cheek; softness and weight of clothing and hair; overly smooth skin; missing hair and teeth details; misalignment in face symmetry; inconsistencies in pixel levels; and strange behavior of an individual doing something implausible (CNET08; CNET14; CNN09; GRD05; USAT03; USAT04; WP01). While it is getting more and more difficult for people to distinguish between a real video and a fake, AI can be instrumental in detecting deepfakes (CBS02; FRB01). For example, AI algorithms can analyze Photo Response Non-Uniformity (PRNU) patterns in footage, that is, imperfections unique to the light sensor of specific camera models, or biometric data such as blood flow indicated by subtle changes that occur on a person's face in a video (CNN06; GRD07; USAT01). New fake-

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

detection algorithms are based on mammalian auditory systems, for example, the ways mice detect inconsistencies and subtle mistakes in audio, which are often ignored by humans (CNET02). AI can either look at videos on a frame-by-frame basis to track signs of forgery, or review the entire video at once to examine soft biometric signatures, including inconsistencies in the authenticated relationships between head movements, speech patterns, and facial expressions such as smiling, to determine if the video has been manipulated (CNN03; FOX07). The latter method can be tailored for individuals, such as high-profile politicians who are likely to be deepfaked (PCM01).

The problem with deepfakes is not only about proving something is false, but also about proving that an object is authentic (FT02). Authentication of video is especially important to news media companies who have to determine authenticity of a video spreading in a trustless environment, in which details of the video's creator, origin, and distribution may be hard to trace (WP01). Proposed solutions to authenticate content range from digital watermarks to digital forensic techniques (FOX06; GRD01). It would be ideal to create a "truth layer", an automated system across the internet to provide a fake vs. authentic measure of videos; that way, every video posted to a social media site would go through an authentication process (CBS03; USAT04). For example, software embedded in smartphone cameras can create a digital fingerprint at the moment of a film's recording. Upon footage playback, its watermark can be compared with the original fingerprint to check for a match, and provide the viewer with a score that indicates the likelihood of tampering (GRD05). Indeed, digital watermarking such as hashing can provide a video file with a short string of numbers that is lost if the video is manipulated (FOX04; FRB04). It can also provide an authenticated alibi for public figures, given that they constantly record where they are and what they are doing (GRD03). Support for video authenticity is also provided by mapping its provenance, that is, whether the video came from a reputable source originally, and how it has since travelled online (FT01). Blockchain technology can help in verifying the origins and distribution of videos by creating and storing digital signatures in a ledger that is almost impossible to manipulate (CNN06). Social media platforms, news media companies and other online actors should then promote the videos that are verified as authentic over non-verified videos (USAT01). Nonetheless, there will always be people who choose not to believe a verification tool, and rather still have a desire to consume and endorse fake media (USAT01).

Finally, technology can prevent the creation of deepfakes by inserting "noise" into photos or videos. This noise is imperceptible to the human eye, but prevents the visual material from being used in automated deepfake software (USAT04). One could also wear specifically designed 3D-printed glasses to evade facial recognition by tricking deepfake software into misclassifying the wearer. This technology could help likely targets such as politicians, celebrities and executives to prevent deepfakes being made of them (FT01). Also, researchers who are developing GAN technologies can design and put proper safeguards in place so that their technologies become more difficult to misuse for disinformation purposes (FOX06). Similar to the cybersecurity domain in general, the first step towards a solution is understanding the problem and its ability to affect us. Only then does it become possible to develop and implement technical solutions that can solve the challenges (FRB04). That said, none of the technological solutions can entirely remove the risk of deepfakes, and technological solutionism (that every problem has a technological solution) may even disorientate the discussion away from more existential questions of why deepfakes exist and what other threats AI can impose (GRD01; GRD03; GRD04). The most efficient ways to combat deepfakes from spreading therefore involve a mixture of legal, educational, and sociotechnical advances (USAT01).

**Discussion and Conclusion**

This study reviewed and analyzed 84 recent public news articles on deepfakes in order to enable a better understanding of what deepfakes are and who produces them, the benefits and threats of deepfake technology, examples of current deepfakes, and how to combat them. In so doing, the study found that deepfakes are hyper-realistic videos digitally manipulated to depict people saying and doing things that never happened. Deepfakes are created using AI, that is, Generative Adversarial Networks (GANs) that pit discriminative and generative algorithms against one another to fine-tune performance with every repetition, and thereby produce a convincing fake (Fletcher, 2018; Spivak, 2019). These fakes of real people are often highly viral and tend to spread quickly through social media platforms, thus making them an efficient tool for disinformation.

The findings of this study offer several contributions to the emerging body of scholarly literature on deepfakes (see Anderson, 2018; Qayyum et al., 2019; Zannettou et al., 2019). Previous research (Fletcher, 2018) argues that

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

deepfakes benefit from, 1) a citizenry increasingly reliant upon commercial media platforms to absorb, process, and communicate information, 2) a heated political context where false narratives are easily spread and easily believed online, and 3) the appearance of powerful AI algorithms capable of manufacturing seemingly real videos. Our findings support these arguments by specifying that such commercial platforms consist of both news media platforms and a range of social media platforms, and that deepfakes are not only fed by a heated political context, but also the current social context due to the so-called information apocalypse, which makes people cease trusting information that does not come from their personal social networks and is inconsistent with their prior beliefs, a phenomenon addressed in previous literature (Britt et al., 2019; Hamborg et al., 2018; Zannettou et al., 2019). The increase in fake news business models that generate web traffic to fake news pages to earn profit through advertising, which has been discussed in previous research (e.g., Figueira & Oliveira, 2017), did not come up in the present study. A likely reason is that the study analyzed news articles from journalists who wish to avoid being associated with actors in the field of journalism that rely on unethical methods such as clickbaiting (cf. Aldwairi & Alwahedi, 2018).

Zannettou et al. (2019) list a number of actors associated with deepfakes, ranging from governments, political activists, criminals, and malevolent individuals creating fake content to paid and unpaid trolls, conspiracy theorists, useful idiots, and automated bots disseminating it through social media platforms. According to Zannettou et al. (2019), the motivation behind these actors' actions may include malicious intent to hurt others in various ways, manipulate public opinion with respect to specific topics, sow confusion or discord to the public, monetary profit, passion about a specific idea or organization or, as MacKenzie and Bhatt (2018) note, plain fun and amusement. Our findings highlight that there are also individuals and organizations such as television companies that generate and support deepfakes in order to develop and apply deepfake technology for legit use such as paid work for music videos. These are considered as early examples of the benefits anticipated from applying GANs.

In regard to legitimate uses for deep learning technology, previous research has addressed movie studios, personalized advertisement companies, and media broadcasters as potential benefiters. For example, Netflix could enable watchers to pick on-screen actors before hitting play or even enable watchers themselves to star in the movie (Chawle, 2019). The present study identified a number of additional areas for legitimate uses of the technology, including educational media and digital communications, games and entertainment, social and healthcare, material science, and various business fields such as fashion and personalized e-commerce.

According to our study, deepfakes are a major threat to society, the political system and businesses because they put pressure on journalists struggling to filter real from fake news, threaten national security by disseminating propaganda that interferes in elections, hamper citizen trust toward information by authorities, and raise cybersecurity issues for people and organizations. In this vein, the study largely supports the findings from previous research (Aldwairi & Alwahedi, 2018; Bates, 2018; Chawla, 2019; Hamborg et al., 2018; Lin, 2019; Wagner & Blewer, 2019) and, at the same time, details these threats through examples of existing and potential uses of deepfakes.

On the other hand, there are at least four known ways to combat deepfakes, namely 1) legislation and regulation, 2) corporate policies and voluntary action, 3) education and training, and 4) anti-deepfake technology. While legislative action can be taken against some deepfake producers, it is not effective against foreign states. Rather, corporate policies and voluntary action such as deepfake-addressing content moderation policies, and quick removal of user-flagged content on social media platforms, as well as education and training that aims at improving digital media literacy, better online behavior and critical thinking, which create cognitive and concrete safeguards toward digital content consumption and misuse, are likely to be more efficient. Government authorities, companies, educators, and journalists need to increase citizens' awareness of the threats posed by AI to media trust, and prohibit fraudulent usage of such technologies for commercial, political, or anti-social purposes. In this vein, our results support and complement those presented by previous studies (Anderson, 2018; Atodiresei et al., 2018; Britt et al., 2019; Cybenko & Cybenko, 2018; Figueira & Oliveira, 2017; Floridi, 2018; Spivak, 2019).

Technological solutions, including automated tools for deepfake detection, content authentication, and deepfake prevention constitute a dynamic field of security methods. Consequently, there are numerous business opportunities for technology entrepreneurs, especially in the areas of cybersecurity and AI. The study highlights that deepfake technology is progressing at an increasing pace. It is quickly becoming impossible

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

for human beings to distinguish between real and fake videos. Hence, our results list numerous cues for detecting deepfakes, and suggest harnessing AI in order to detect AI-generated fakes as an efficient combat strategy. At the same time, the study reckons that there are growing needs for online source verification and content authentication, and that ubiquitous truth layers based on digital watermarks should be used. Further, another emerging technology, namely blockhain can be of help. Blockchain technology is not only highly resistant to forgeries and can store data in an accruable, safe, transparent, and traceable way, but it can also track and certify the origins and history of the data (Floridi, 2018). Again, these results are in line with previous research (Anwar et al., 2019; Atanasova et al., 2019; Bates, 2018; Chawla, 2019; Floridi, 2018; Hasan & Salah, 2019; Qayyum et al., 2018; Spivak, 2019). In the spirit of a review study, our results contribute to the emerging field of deepfakes by pulling together dispersed findings from the sparse academic research on fake news and deepfakes, and by fine-tuning those findings with examples and discussion on deepfakes taking place in public news articles.

All said, there are certainly limitations in the study. First, although the empirical research covered 84 online news articles on deepfakes, there are many more articles available and, given the speed of development of this technology, those articles could also provide additional information on deepfakes and suggest more methods to fight them. Second, our empirical material was collected from public sources, namely online news media sites. Using other types of data, such as deepfake-focused online community discussions and interviews with GAN developers and deepfake artists, some of whom are known to the public due to their work not only as deepfake technology developers but also as anti-deepfake technology developers, could give additional insight on strategies to combat deepfakes. Also, commentary sections in some of the analyzed news articles had extensive amount of opinions and ideas from readers; analysis of those comments might give additional insights on how deepfakes are perceived by a larger audience, and thus what education-oriented combat methods should emphasize. These limitations provide ample opportunities for future research on deepfakes.

**References**


Aldwairi, M., & Alwahedi, A. 2018. Detecting Fake News in Social Media Networks. *Procedia Computer Science*, 141: 215–222. https://doi.org/10.1016/j.procs.2018.10.171

Anderson, K. E. 2018. Getting acquainted with social networks and apps: combating fake news on social media. *Library HiTech News*, 35(3): 1–6.

Anwar, S., Milanova, M., Anwer, M., & Banihirwe, A. 2019. Perceptual Judgments to Detect Computer Generated Forged Faces in Social Media. In F. Schwenker, & S. Scherer (Eds.), *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*. MPRSS 2018. Lecture Notes in Computer Science, vol. 11377. Springer, Cham. https://doi.org/10.1007/978-3-030-20984-1_4

Atanasova, P., Nakov, P., Màrquez, L., Barrón-Cedeño, A., Karadzhov, G., Mihaylova, T., Mohtarami, M., & Glass, J. 2019. Automatic Fact-Checking Using Context and Discourse Information. *Journal of Data and Information Quality*, 11(3): Article 12. https://doi.org/10.1145/3297722

Atodiresei, C.-S., Tănăselea, A., & Iftene, A. 2018. Identifying Fake News and Fake Users on Twitter. *Procedia Computer Science*, 126: 451–461. https://doi.org/10.1016/j.procs.2018.07.279

Bates, M. E. 2018. Say What? 'Deepfakes' Are Deeply Concerning. *Online Searcher*, 42(4): 64.

Borges, L., Martins, B., & Calado, P. 2019. Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News. *Journal of Data and Information Quality*, 11(3): Article No. 14. https://doi.org/10.1145/3287763

Britt, M. A., Rouet, J.-F., Blaum, D., & Millis, K. 2019. A Reasoned Approach to Dealing With Fake News. *Policy Insights from the Behavioral and Brain Sciences*, 6(1): 94–101. https://doi.org/10.1177/2372732218814855

Chawla, R. 2019. Deepfakes: How a pervert shook the world. *International Journal of Advance Research and Development*, 4(6): 4–8.

Cybenko, A. K., & Cybenko, G. 2018. AI and Fake News. *IEEE Intelligent Systems*, 33(5): 3–7. https://doi.ieeecomputersociety.org/10.1109/MIS.2018.2877280

Day, C. 2019. The Future of Misinformation. *Computing in Science & Engineering*, 21(1): 108–108. https://doi.org/10.1109/MCSE.2018.2874117

De keersmaecker, J., & Roets, A. 2017. 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, 65: 107–110. https://doi.org/10.1016/j.intell.2017.10.005

Figueira, A., & Oliveira, L. 2017. The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121: 817–825. https://doi.org/10.1016/j.procs.2017.11.106

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

Fletcher, J. 2018. Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance. *Theatre Journal*, 70(4): 455–471. Project MUSE, https://doi.org/10.1353/tj.2018.0097

Floridi, L. 2018. Artificial Intelligence, Deepfakes and a Future of Ectypes. *Philosophy & Technology*, 31(3): 317–321.

Hamborg, F., Donnay, K., & Gipp, B. 2018. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*. https://doi.org/10.1007/s00799-018-0261-y

Hasan, H. R., & Salah, K. 2019. Combating Deepfake Videos Using Blockchain and Smart Contracts. *IEEE Access*, 7: 41596–41606. https://doi.org/10.1109/ACCESS.2019.2905689

Jang, S. M., & Kim, J. K. 2018. Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, 80: 295–302. https://doi.org/10.1016/j.chb.2017.11.034

Lin, H. 2019. The existential threat from cyber-enabled information warfare. *Bulletin of the Atomic Scientists*, 75(4): 187–196. https://doi.org/10.1080/00963402.2019.1629574

MacKenzie, A., & Bhatt, I. 2018. Lies, Bullshit and Fake News: Some Epistemological Concerns. *Postdigital Science and Education*. https://doi.org/10.1007/s42438-018-0025-4

Maras, M. H., & Alexandrou, A. 2019. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence & Proof*, 23(3): 255–262. https://doi.org/10.1177/1365712718807226

Qayyum, A., Qadir, J., Janjua, M. U. & Sher, F. 2019. Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News. *IT Professional*, 21(4): 16–24. https://doi.org/10.1109/MITP.2019.2910503

Spivak, R. 2019. "Deepfakes": The newest way to commit one of the oldest crimes. *The Georgetown Law Technology Review*, 3(2): 339–400.

Wagner, T.L., & Blewer, A. 2019. "The Word Real Is No Longer Real": Deepfakes, Gender, and the Challenges of AI-Altered Video. *Open Information Science*, 3(1): 32–46. https://doi.org/10.1515/opis-2019-0003

Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. 2019. The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. *Journal of Data and Information Quality*, 1(3): Article No. 10. https://doi.org/10.1145/3309699

**About the Author**

Mika Westerlund, DSc (Econ), is an Associate Professor at Carleton University in Ottawa, Canada. He previously held positions as a Postdoctoral Scholar in the Haas School of Business at the University of California Berkeley and in the School of Economics at Aalto University in Helsinki, Finland. Mika earned his doctoral degree in Marketing from the Helsinki School of Economics in Finland. His research interests include open and user innovation, the Internet of Things, business strategy, and management models in high-tech and service-intensive industries.

Citation: Westerlund, M. 2019. The Emergence of Deepfake Technology: A Review. Technology Innovation Management Review, 9(11): 39-52. http://doi.org/10.22215/timreview/1282

CC BY

Keywords: Deepfake, fake news, artificial intelligence, deep learning, cybersecurity.

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

**Appendix**: Studied news articles on deepfakes

| #ID | Author(s) | Title (shortened) | Date |
|---|---|---|---|
| *CBS News* | | | |
| CBS01 | Patterson, D. | From deepfake to "cheap fake"… | June 13, 2019 |
| CBS02 | N/A | I never said that! High-tech deception … | July 2, 2018 |
| CBS03 | Evans, C. | Spotting fake news in a world with … | April 17, 2018 |
| *CNET* | | | |
| CNET01 | Van Boom, D. | These deepfakes of Bill Hader are … | August 12, 2019 |
| CNET02 | Mack, E. | Researchers propose detecting deepfakes … | August 12, 2019 |
| CNET03 | Al-Heeti, A. | Sharing deepfake revenge porn is now … | July 1, 2019 |
| CNET04 | Singh, D. | Google, Facebook, Twitter put on notice … | July 17, 2019 |
| CNET05 | Solsman, J.E. | Deepfakes freak YouTubers out … | July 12, 2019 |
| CNET06 | Solsman, J.E. | Google, Facebook, Twitter aren't … | August 6, 2019 |
| CNET07 | Solsman, J.E. | Samsung deepfake AI could fabricate … | May 24, 2019 |
| CNET08 | Solsman, J.E. | Deepfakes may ruin the world … | April 4, 2019 |
| CNET09 | Solsman, J.E. | Facebook, YouTube, Twitter must team … | June 13, 2019 |
| CNET10 | Gonzalez, O. | Instagram chief Adam Mosseri: We don't … | June 25, 2019 |
| CNET11 | Keane, S. | Congress wrestles with 'deepfake' threat … | September 5, 2018 |
| CNET12 | Ng, A. | Deepfakes, disinformation among global … | January 29, 2019 |
| CNET13 | Ng, A. | Deepfakes are a threat to national … | September 13, 2018 |
| CNET14 | Kooser, A. | Deepfake your dance moves with an AI … | August 27, 2018 |
| CNET15 | Carson, E. | Reddit cracks down on 'deepfake' … | February 7, 2018 |
| CNET16 | Hautala, L. | If you like deepfakes, you might be … | February 11, 2018 |
| *CNN* | | | |
| CNN01 | O'Sullivan, D. | The Democratic Party deepfaked its own … | August 10, 2019 |
| CNN02 | O'Sullivan, D. | House Intel chair sounds alarm in … | June 13, 2019 |
| CNN03 | Metz, R. | The fight to stay ahead of deepfake … | June 12, 2019 |
| CNN04 | Metz, R., O'Sullivan, D. | A deepfake video of Mark Zuckerberg … | June 11, 2019 |
| CNN05 | Yurieff, K. | Instagram head says company is … | June 25, 2019 |
| CNN06 | O'Brien, S.A. | Deepfakes are coming. Is Big Tech ready? | August 8, 2018 |
| CNN07 | Metz, R. | Baby Elon Musk, rapping Kim Kardashian: … | June 23, 2019 |
| CNN08 | O'Sullivan, D. | Congress to investigate deepfakes as … | June 4, 2019 |
| CNN09 | Metz, R. | Researchers can now use AI and a photo … | May 24, 2019 |
| CNN10 | O'Sullivan, D. | Lawmakers warn of 'deepfake' videos… | January 28, 2019 |

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

| | | | |
|---|---|---|---|
| *Digiday* | | | |
| DD01 | Southern, L. | 'A perfect storm': The Wall Street Journal … | July 1, 2019 |
| DD02 | Southern, L. | How Reuters is training reporters to spot … | March 26, 2019 |
| *Forbes* | | | |
| FRB01 | Marr, B. | The Best (And Scariest) Examples Of … | July 22, 2019 |
| FRB02 | Baron, K. | Digital Doubles: The Deepfake Tech … | July 29, 2019 |
| FRB03 | Ballantine, M. | Are Deepfakes Invading The Office? | July 3, 2019 |
| FRB04 | Taulli, T. | Deepfake: What You Need To Know | June 15, 2019 |
| FRB05 | Sandler, R. | Instagram Won't Take Down Mark … | June 11, 2019 |
| FRB06 | Towers-Clark, C. | Mona Lisa And Nancy Pelosi: The … | May 31, 2019 |
| FRB07 | Dietmar, J. | GANs And Deepfakes Could Revolutionize … | May 21, 2019 |
| FRB08 | Leetaru, K. | DeepFakes: The Media Talks Politics While… | Mar 16, 2019 |
| FRB09 | Morris, I. | Revenge 'Porn' Gets Even More Horrifying … | Feb 5, 2018 |
| FRB10 | Morris, I. | Deepfake Porn Banned By Reddit And … | Feb 7, 2018 |
| *Fox News* | | | |
| FOX01 | Carbone, C. | Scary deepfake video shows Bill Hader … | August 13, 2019 |
| FOX02 | Carbone, C. | New deepfake technology makes Rasputin… | June 20, 2019 |
| FOX03 | Mikelionis, L. | 'Deepfake' clip of Mark Zuckerberg … | June 12, 2019 |
| FOX04 | Carbone, C. | Creepy deepfake AI lets you put words … | June 11, 2019 |
| FOX05 | Brandon, J. | Terrifying high-tech porn: Creepy … | February 16, 2018 |
| FOX06 | Farid, H. | Deepfakes give new meaning to the … | June 18, 2019 |
| FOX07 | Carbone, C. | New tool detects deepfakes with 96 … | June 24, 2019 |
| FOX08 | Brandon, J. | How AI-generated videos could be the … | March 12, 2018 |
| FOX09 | Wallace, D. | 'Deepfake' videos, other tech could … | June 14, 2019 |
| FOX10 | Carbone, C. | Creepy AI generates endless fake faces … | February 18, 2019 |
| *Financial Times* | | | |
| FT01 | Murphy, H. | Cyber security companies race to combat … | August 16, 2019 |
| FT02 | Waters, R. | Rising tide of online deepfakes bring … | June 13, 2019 |
| FT03 | Kuper, S. | The age of scepticism: from distrust to … | October 18, 2018 |
| FT04 | Khalaf, R. | If you thought fake news was a problem … | July 25, 2018 |
| FT05 | Murgia, M. | Why some AI research may be too … | June 19, 2019 |
| *The Guardian* | | | |
| GRD01 | Schwartz, O. | Deepfakes aren't a tech problem. They're … | June 24, 2019 |
| GRD02 | Hunt, E. | Deepfake danger: what a viral clip of Bill … | August 13, 2019 |

# The Emergence of Deepfake Technology: A Review

*Mika Westerlund*

| | | | |
|---|---|---|---|
| GRD03 | Chivers, T. | What do we do about deepfake video? | June 23, 2019 |
| GRD04 | Beard, M. | To fix the problem of deepfakes we must … | July 23, 2019 |
| GRD05 | Parkin, S. | The rise of the deepfake and the threat to … | June 22, 2019 |
| GRD06 | Hern, A. | New AI fake text generator may be too … | February 14, 2019 |
| GRD07 | Schwartz, O. | You thought fake news was bad? Deep … | November 12, 2018 |
| GRD08 | Hern, A. | My May-Thatcher deepfake won't fool … | March 12, 2018 |
| GRD09 | Chadwick, P. | The liar's dividend, and other challenges … | July 22, 2018 |
| GRD10 | Hern, A. | I thought online fakes would cause an … | January 2, 2019 |
| *PC Magazine* | | | |
| PCM01 | Eddy, M., Rubenking, N. | Detecting Deepfakes May Mean Reading … | August, 9, 2019 |
| PCM02 | Albanesius, C. | Deepfake Videos Are Here, and We're Not… | June 13, 2019 |
| PCM03 | Horowitz, B.T. | AI and Machine Learning Exploit… | May 13, 2019 |
| PCM04 | Rubenking, N. | How Lab Mice Are Helping Detect … | August 8, 2019 |
| PCM05 | Kan, M. | Latest Deepfake Tech Will Have You … | August 24, 2018 |
| PCM06 | Kan, M. | US Lawmakers: AI-Generated Fake Videos … | September 13, 2018 |
| PCM07 | Moscaritolo, A. | PornHub, Twitter Ban 'Deepfake' … | February 7, 2018 |
| PCM08 | Kan, M. | Facebook Declines to Delete Fake … | June 11, 2019 |
| PCM09 | Kan, M. | This AI Can Recreate Podcast Host Joe … | May 17, 2019 |
| PCM10 | Dickson, B. | When AI Blurs the Line Between Reality … | June 7, 2018 |
| *USA Today* | | | |
| USAT01 | Andrews, J. | Fake news is real – A.I. is going to make … | July 12, 2019 |
| USAT02 | Molina, B. | 'Deepfake' videos can be generated with … | May 24, 2019 |
| USAT03 | Brown, D. | Wait, is that video real? The race against … | May 13, 2019 |
| USAT04 | Walsh, C. | What is a deepfake? This video technology… | March 15, 2019 |
| *Washington Post* | | | |
| WP01 | Harwell, D. | Top AI researchers race to detect … | June 12, 2019 |
| WP02 | Zakrzewski, C. | The Technology 202: It's time for … | June 13, 2019 |
| WP03 | Sasse, B. | This new technology could send American … | October 19, 2018 |
| WP04 | Wang, A.B. | Biden says he won't use bots… | June 14, 2019 |