Advancement's Achilles' Heel: The Disservice of Deepfakes

Caroline Frisiras

Department of Journalism and Media Management, University of Miami

Abstract

Advancements in artificial intelligence have led to the rise of deepfake technologies. Using such neural network models as Generative Adversarial Network and the Video Autoencoder, computer users are able to manipulate existing photos and videos of people, allowing them to create the appearance of actions or words the subject did not say or do. With software for video detection and methods for harm mitigation still in the works, the rise of deepfake technology brings about questions of computing ethics: how society is working to protect vulnerable populations, and how industry leaders and authority figures ought to regulate the open-source niche of deepfake technology to best combat the spread of misinformation and exploitation that goes along with it? The reality is that *anyone* can have their photos used in deepfake videos without their knowledge, and an analysis of the implications of open-source development is long overdue.

Keywords: artificial intelligence, deepfake, neural network, computer ethics

Advancement's Achilles' Heel: The Disservice of Deepfakes

Introduction

A Taylor Swift "porno." Joe Biden discouraging Democrats from taking part in the upcoming election. Richard Nixon admitting that the moon landing was fake. All these occurrences – according to video evidence on the internet – really happened. But in today's day and age, can people believe what they see on the internet? Even if, to the human eye, the videos look indistinguishable from a genuine person saying or doing these things? The answer is clear: *nothing* seen on the internet can be taken at face value, no matter how authentic any one video may appear to be. While artificial intelligence, a field striving to achieve human-like thinking abilities from machines, has existed as a concept since the 1950s (Anyoha, 2017), in recent years we have seen rapid advancements in both the abilities and applications of the technology. The field as a whole boasts an impressive list of day-to-day uses, spanning across nearly every aspect of our lives imaginable, with utilizations in the fields of medicine, entertainment, security, finance, transportation, marketing, gaming, and many more. In this paper, I will provide a thorough analysis of one specific subsect of artificial intelligence: synthetic media, more commonly referred to as "deepfakes". Because deepfake technology is rapidly developing, and artificial intelligence is becoming increasingly integrated into society, it is important to understand the emerging technologies, alongside the inherent risks of their widespread use.

In layman's terms, deepfakes are a type of artificial intelligence that allows for someone's face to be edited, depicting them saying or doing things they never actually said or did. The United States Government Accountability Office explains that, "[a] deepfake is a video, photo, or audio recording that seems real but has been manipulated with AI. The underlying technology

can replace faces, manipulate facial expressions, synthesize faces, and synthesize speech" (United States Government Accountability Office, 2020a).

Deepfakes are an ever-evolving subsect of artificial intelligence, and the internet has seen a massive spike in the number of deepfake videos in the last few years. The "2023 State of Deepfakes" report (Home Security Heros, 2023) found a total of 95,820 deepfake videos online – a staggering 550% increase from the total number of videos in 2019. The ability to edit *anyone* to make them say or do *anything* is certainly a terrifying prospect. Seeing as the dissemination of deepfake videos is happening at an alarmingly high rate, and anyone with photos of them online (which, in a digital age in most people) can be used in a deepfake videos without their knowledge, a comprehensive review of this topic is both warranted and timely.

I will first review the history of deepfake development, followed by an explanation of the key technical features of them. I will then examine the various applications of deepfakes, both positive and negative, in social, political, and entertainment contexts. Afterwards, I will consider the ethical and legal issues regarding deepfakes, alongside ways in which we can detect and mitigate negative side effects of the technology. Finally, I will analyze and argue how I personally best see fit for the applications and regulations of deepfake technology moving forward.

Deepfake Technology

To understand how deepfakes operate in a modern society, it is crucial to examine both the origins of the technology, alongside its key components and advancements. The following section will give overview of the history and development of deepfakes, how deepfake software works to create their ending results, and applications for deepfakes in the modern world.

History and Development

The term “deepfake” was coined in 2017 via a reddit user who used the technology to face-swapping women into pornographic videos (Somers, 2020). This is essentially what catapulted the widespread popularity of deepfake technology. That being said, the processes and technology used to create deepfake videos have been established long before 2017.

Video Rewrite

The first stepping stone in the development of deepfake technology was the 1997 “Video Rewrite” technique, created by Christoph Bregler, Michele Covell, and Malcolm Slaney. The aim of the project was to “[use] existing footage to create automatically new video of a person mouthing words that she did not speak in the original footage” by essentially reordering images of the subject’s mouth to align with the phoneme sequence of the desired speech, stitching mouth images with background footage to be able to edit the speaker’s utterances (Bregler et al., 1997, p. 1). Video Rewrite was initially created with the intention of being a useful tool in endeavors, such as movie editing, and for a while the technology was not used in malicious or misrepresentative ways.

Deep Neural Networks

It was not until years later, in the mid-2010s that Deep Neural Networks (henceforth referred to as DNNs) used for facial recognition emerged in academia (Bernaciak & Ross, 2022). A neural network essentially mimics the way the human brain works; it takes inputs, categorizes them via “hidden layers,” and produces outputs. A neural network consisting of two or more hidden layers is considered to be a DNN. DNNs are known for being able to classify more complex parameters of information, as well as more data. Essentially, machine learning utilizes neural networks to allow for computers to intake, sort through, and “learn” information in the way human brains would.

Current Software

DeepFace is the most notable of these early, machine-learning facial recognition projects. Initially published in 2014, DeepFace was a major breakthrough in the possibilities of facial recognition machine learning. The system had nine neural networks with over 120 million parameters and was trained with a labeled dataset of four million facial images in total (Taigman et al., 2014).

The discovery of additional DNN models, most notably the Generative Adversarial Network (GAN) and Variational Autoencoder (VAE), both released in 2014, transitioned machine learning in facial recognition from simply identifying images to swapping and reenacting them (Bernaciak & Ross, 2022).

How Deepfakes Work

There are two main ways in which deepfake videos are created: using Generative Adversarial Networks, or (more popularly) with Variational Autoencoders. Even though GANs are more difficult to train, they produce a higher caliber image than their easier-to-use counterpart, VAEs (Nanos, 2024). While both machines ultimately accomplish the same result, they go about them with different methods.

Generative Adversarial Network

GANs essentially put two neural networks against each other; one to *create* the deepfake image (referred to as a Generative Neural Network (GNN) – the *generative* component of GAN), and the second (known as the discriminant classifier – the *adversarial* component of GAN) to give feedback to the GNN and detect if the image created by the generative network is real or fake (Finger, 2022). This feedback cycle continues, allowing for the models to produce visuals that are incredibly realistic, and often undetectable as deepfakes to the average human eye.

Variational Audoencoders

The more common method for deepfake creation employs the use of VAEs, which utilizes encoders and decoders to create deepfake content. Essentially, encoders compress and abstract images in order to be represented in a lower-dimensional space, referred to as "latent vectors" (Bernaciak & Ross, 2022). Once these images are encoded, they reside in the "latent space," a multi-dimensional space with each dimension representing a different feature or attribute of the data (e.g., one dimension may represent factors like age or gender) (Bernaciak & Ross, 2022). Decoders turn the abstractions back into images, able to interpolate (or seamlessly combine) the images (Somers, 2020).

That being said, the average creator of deepfake videos likely is not creating or training their own neural networks to generate deepfake content. Many users utilize software such as "Zao," "DeepFaceLab," "FaceApp," and, the since removed, "DeepNude," to aid in the creation of deepfake content (Johnson & Johnson, 2023). GitHub, an open-source development community for coders, houses countless deepfake software available for users, making video creation simple and accessible for those without extensive experience and training in artificial intelligence. Deepfake software and applications are available to every operating system, having downloads for Mac, Android, Windows, and PC systems alongside availability in Google Play and Apple App Store. The simplicity of creating deepfake videos, alongside the freely available software to do so, allows for endless possibilities and applications of deepfake content.

Applications of Deepfake Technology

While the name "deepfake" is most commonly associated with computer-generated pornographic content, deepfakes serve an array of practical uses in a modern society. Deepfakes are used frequently in the entertainment and media production industries. In post-production of

television shows or movies, deepfake technology can assist with a variety of editing processes, such as dubbing and voiceover work, bringing deceased actors back to life in order to finish a project or series of projects, or allowing for the editing of actors into scenes they were not initially in.

Reuters, an international news agency based out of London, United Kingdom, created an artificial intelligence deepfake news reporter to present their sports news summary (Mayhew, 2022). Snoop Dogg utilized deepfake technology in his music video, allowing deceased rapper Tupac Shakur to appear rapping along to Snoop Dogg's "I C Your Bullsh*t" — a song that was released 24 years *after* Shakur's death. The Mona Lisa has been brought to life through Samsung's AI lab in Moscow, Russia. Deepfakes have been used in the fashion market in the form of technology that allows for customers to deepfake their face onto clothing models to see how clothing items would look on them (Chow, 2022). David Beckham's "malaria no more" campaign uses deepfake technology to allow him to speak nine different languages, reaching audiences of all cultures and backgrounds. Safe to say, deepfake technology is not inherently negative; it can serve a valuable purpose in art and entertainment, bring historical figures back to life, and upgrade the quality of technology experienced in day-to-day activities. But at what cost do these technologies come at?

Ethical and Legal Considerations

Although the caliber and precision deepfake technology has developed into is impressive, there are many ethical and legal problems that arise from the creation and dissemination of deepfake content. In the following sections, key ethical issues in deepfake content will be explained, followed by legal exposition, current regulations, and why the nature of artificial intelligence makes it nearly impossible to adequately regulate the use of deepfake technologies.

Ethical Considerations

Despite varying potential applications of the technology, the vast majority of deepfake technology is used for pornography, with about 96% of deepfake videos online hosting pornographic content (Ajder et al., 2019), and pornography is considered the most well-known application for deepfake technology (Lyon & Tora, 2023, ch.2). Pornographic content, misinformation in political contexts, and the potential for *anyone* to fall victim to deepfake creations are creating a multitude of gray areas in computing ethics. *Exploring Deepfakes* gives a three-part ethical breakdown of the technology, suggesting such evaluations should look at the independent video and examine for consent, respect, and deception in the content. (Lyon & Tora, 2023).

Pornographic Deepfakes

The most controversial and concerning aspect of deepfake technology is that it has become a means of sexual exploitation for celebrity name, image, and likeness (NIL). Pornographic deepfake videos feature almost exclusively female celebrities and public figures, ranging from actresses to corporate figures. On deepfake pornography websites, 99% of women featured in deepfake videos work in the entertainment industry, with the other 1% in news and media; however, other platforms, such as YouTube, feature politicians and business owners as well (Ajder et al., 2019).

Even more alarming is the notion that due to the unregulated nature of deepfake technology, users are able to essentially create their own pornographic material. Because deepfake production is unregulated, the creation and possession of otherwise illegal content, such as child pornography, can go unrestricted. While it is illegal to create, possess, and *view* child pornography (Seto, 2013), this prohibition does not stop or regulate *creation* of such content that

is created and saved for personal use — it merely makes it so that *posting* or *being caught* with such content puts one at risk for legal repercussions.

Political Deepfakes

Another major ethical concern with deepfake videos is their ability to be used in political contexts, altering the words of public figures and spreading false information about current states of affairs. With an upcoming election in 2024, the ability for deepfakes to influence public opinion brings concerns about the integrity of our democratic processes, especially with the ability for dissemination of malicious and misleading information to cross language barriers seamlessly. For example, CNN reports that Taiwan's election faced misinformative deepfake videos this past December, of U.S. congressmen soliciting votes in favor of their presidential candidate (Wolf, 2024). In the U.S, voters received calls from someone who appeared to be Joe Biden – but was rather a deepfake of his voice, urging voters not to vote in the democratic primaries (Steck & Kaczynski, 2024). In recent years, political deepfakes have been trending globally, and with the advancements seen in artificial intelligence's ability to mimic real human behavior, bad actors are able to manipulate footage and audio of politicians flawlessly. Similar to those in the entertainment industry, politicians have very public platforms, with countless high quality photos and videos of them online, making them a prime target for deepfake manipulation. Popularly circulated deepfakes of male politicians tend to be centered around altering their utterances. Not shockingly, female politicians are more popularly victims of pornographic deepfake content.

Sexism and Technology

Not only is the common use of deepfake videos deeply unsettling for the individuals who experience such violating content of them circulating the internet, but it speaks to a greater issue

about the dehumanization, sexualization, and violation of women happening at the hands of a male-dominated technology industry. Technology as a whole, but specifically artificial intelligence, has been an area in which implicit bias runs rampant. Seeing as the vast majority of artificial intelligence researchers at top companies are men — with only 15% and 10% of researchers being female at entities like Facebook and Google, respectively — artificial intelligence has been known to include a narrow, white-male dominated version of what a "normal" person looks like (Pichhi, 2019). Because these technologies are developed primarily by men, in a society that largely operates at the hands of men, the dangers and vulnerabilities they bring to girls and women around the world have gone rather overlooked since the widespread deployment of deepfake technology began.

When software with capabilities are circulated on open-source platforms allowing for the constant sharing, creation, and saving of demeaning content, it becomes clear that *something* is amiss. Unfortunately, the issue of sexual exploitation of women via the internet is a conversation that the scope of this paper could not even begin to adequately address. Nevertheless, deepfakes have certainly created a gray area in exactly *how much* freedom should be given in the advancement of artificial intelligence.

Legal Considerations

On the one hand, the technology community has long thrived off of open-source platforms, such as GitHub, that allow for collaboration and constant improvement of certain technologies or software. To limit this access goes against the First Amendment and its provisions regarding freedom of speech. The exception to this rule includes obscenity and child pornography. However, the issue is not the platforms themselves, per se — it is what users have the potential to utilize them for.

On the other hand, deepfake videos are disproportionately used for harmful, demeaning, and non-consensual pornographic content, leading one to believe that some degree of platform access limitation or production-stage monitoring may be necessary. While GitHub's Acceptable Use Policies indicate that users must follow all applicable laws and regulations, and explicitly bans content that "is sexually obscene or relates to sexual exploitation or abuse, including of minors," they do allow for sexual content that is in "artistic, educational, historical or journalistic contexts" (GitHub, 2024). Content that violates GitHub's terms of use may be removed from the platform, although both the video and the software that enabled video's production are still available for the creator of the video. Ultimately, the government is the entity at play that ought to do more to protect the exploitation of victims via deepfake technology.

Legal Exposition

Because the field of artificial intelligence is a relatively new and ever-evolving one, and experiences many of its advancements and widespread applications via open-source communities, it is a particularly difficult field to regulate. The issue is that the platforms and technologies that allow for deepfake creation are not inherently malicious -– many of them being used solely for entertainment and educational purposes, making the outlawing of deepfakes as a whole, improbable.

Nevertheless, the lack of ability to regulate this type of technology proves dangerous for victims of sexually explicit deepfakes made and released without their consent. There are currently no uniform regulations for all 50 states for adult victims of deepfake pornography, and the few states that have adopted laws surrounding sexually explicit deepfake content have done so in different ways. The House of Representatives Staff Analysis of Florida Bill CS/CS/HB 1453: Sexually Explicit Material explains the rationale:

> [N]o state completely bans the creation or distribution of all deepfake content. A complete ban of such images would likely violate constitutional protections under the First Amendment. However, certain categories of speech, including defamation, fraud, true threats, and the imminent-and likely incitement of violence, are not entitled to protections under the First Amendment, 59 and some deepfake content is likely to fall into one of these categories and therefore may be regulated (Justice Appropriations Subcommittee et al., 2022, p.7).

Current and Proposed Regulations

While states like Georgia, Hawaii, Texas, and Virginia have taken an approach that criminalizes (non-consensual) sexually explicit deepfakes, other states like California and Illinois have merely given victims the rights to sue creators who infringe upon their likenesses with this type of content (Mulvihill, 2024).

Currently, deepfake videos involving pornographic content of children (minors under 18 years of age) are outlawed federally, as the language in 18 U.S. Code § 2252 specifies that any *visual depiction* of minors engaging in sexually explicit activity are punishable by both fines and imprisonment. The law specifies that "'child pornography' means any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture, whether made or produced by electronic, mechanical, or other means, of sexually explicit conduct" (18 U.S.C. § 2256).

A recently proposed bipartisan and bicameral *Disrupt Explicit Forged Images and Non-Consensual Edits Act of 2024 (DEFIANCE Act)* aims to classify deepfakes of this nature as "digital forgery," giving a federal civil right of action to victims of this technology. The Bill proposes that "an identifiable individual whose intimate visual depiction is disclosed [...] without

the consent of the identifiable individual, where such disclosure was made by a person who knows or recklessly disregards that the identifiable individual has not consented to such disclosure, may bring civil action against that person in an appropriate district court of the United States for relief" (Ocasio-Cortez, 2024). The bill would compensate victims in the amount of actual damages or liquidated damages of $150,000, alongside the cost of litigation and attorney's fees.

Legal Issues

While the proposed DEFIANCE Act certainly serves as a step in the right direction, the reality is that the dissemination of deepfake videos happens on such a massive scale that by the time a victim is made aware of it, so is the rest of the world. Civil litigation serves to award monetary or injunctive relief to victims but cannot undo the damage already done by widespread viewership of deepfake videos of this nature. Once a video is posted to a platform like X and gains viral recognition, users can download or screen-record videos, keeping them in possession even after being removed from the platform. While celebrities may have the resources to target the initial creator of the video who infringed upon their NIL, many victims do not. And to eradicate every instance in which *all* deepfake videos have been — and may continue to be — shared, saved, or viewed is impossible.

Detection and Mitigation

Because deepfake technology (and artificial intelligence as a whole) has advanced at such an unprecedented rate, it is becoming increasingly difficult to detect what is, and is not, real. While traditional recommendations for deepfake detection include phenomena that can be witnessed by the human eye, such as checking for unnatural blinking patterns, mis-matched earrings, or facial features lacking typical definition (United States Government Accountability

Office, 2020b), deepfakes are becoming more and more advanced, resulting in high-definition videos indistinguishable from real people to the naked eye.

Detection

Efforts have been made towards developing software that can accurately detect deepfake videos, but there is certainly a long way to go in the development of a consistently accurate detection method for deepfake technology. Because of the ability deepfakes have to violate both personal boundaries and sway public opinions, being able to determine the authenticity of videos on the internet remains a pressing issue and key goal for software development in 2024. Artificial intelligence is developed chiefly via open source software, and creating precise video detection software is proving to be an all-hands-on-deck challenge.

Deepfake Detection Challenges

In 2020, a massive "DeepFake Detection Challenge" (DFDC) was hosted by Kaggle, a data science and machine learning competition platform operated by Google. Provided with a full dataset of 124,000 videos utilizing eight different facial modification algorithms (Meta, 2020), the DFDC aimed to accelerate progress in detection methods for deepfake videos. The highest performing model in the challenge only reached about 65% precision in its detection (Meta, 2020).

Deepfake Video Detection (DVD) still has a way to go in its development and implementation, and currently there does not appear to be one foolproof system created. DFDCs are a common method for encouraging video detection and have been hosted and encouraged by entities like the U.S. GAO, Meta, Kaggle, and Massachusetts Institute of Technology in hopes of creating an accurate and effective method for detecting deepfake videos.

Current Detection Software

Today, deepfake audio detection methods are more developed and accurate, although they still are not perfect. A recently released NPR article examines the three primary deepfake audio detection providers: Pindrop Security, AI or Not, and AI Voice Detector. The latter two software are available for personal use, but Pindrop Security solely exists for businesses in its current state. All three companies provide mere *probabilities* of the likelihood that the audio clips are — or are not — deepfakes. NPR reported that all three platforms incorrectly identified at least some of their sample clips, and while Pindrop Security achieved approximately 96% accuracy, AI Voice Detector wrongly identified approximately 6% of NPR's samples, returning an additional 38% of samples as "inconclusive." The AI or Not program was only capable of correctly identifying about half the clips submitted (Jingnan, 2024). The article featured input from Sarah Barrington, an AI and Forensics researcher at the University of California, Berkeley, who explained that accuracy is of the essence: "If we label a real audio as fake, let's say, in a political context, what does that mean for the world? We lose trust in everything, [...] And if we label fake audios as real, then the same thing applies. We can get anyone to do or say anything and completely distort the discourse of what the truth is" (Jingnan, 2024).

Mitigation

Mitigation aims to reduce or alleviate the negative side effects of improper deepfake technology use. Typically, platforms are responsible for governing the content posted on their site. Once a video on a platform is flagged for obscene, illegal, or misinformative content, it will be deleted. Platforms like X have adopted policies for synthetic and manipulated media, stating that violating posts will be deleted, with the potential for the poster's account to be locked (Twitter, 2023). In terms of mitigating one's risk of *becoming* a victim of deepfakes, the best way

to protect oneself is to not have any photos or videos of themselves available on the internet (Lyon & Tora, 2023, ch.2).

Perhaps the only way to prevent the risk of deepfake technology being used for the wrong reasons would be to regulate and monitor the actual *production* of deepfake content — rather than enforcing regulations once it has already been posted. Had this technology been maintained as proprietary knowledge within entertainment and technology professionals, it may have been a different story. However, that is simply not the nature of artificial intelligence. As explained in *Evolving Deepfakes*, "[t]he technology is 'out of the bag' and cannot be erased. Even if the technology were to be banned immediately, state-level actors would still be able to wield it in potentially dangerous ways and could ignore any laws or restrictions placed on its use" (Lyon & Tora, 2023). The anonymity of the internet – specifically anonymity used in advanced technology communities for the posting and creation of otherwise illegal content – aims to ensure that there cannot be a way to *proactively* enforce regulations for manipulated media.

Conclusion

As someone who has grown up in an era of unprecedented technological advancements, I tend to welcome the idea of new technology. I have also always known just how dangerous of a place the internet can be. In doing research for this paper, I was admittedly impressed with the positive applications deepfake technologies have, and the precision in which these softwares can operate. Machine learning has surpassed our wildest expectations and only continues to improve, with neural network models like GAN and DNN constantly raising the quality and scope of their uses. Deepfake technology is versatile and can be utilized in all types of personal and professional endeavors, with almost no bounds on what the technology can be used *for*.

The versatility of deepfake technology is, perhaps, also one of its biggest downfalls. This technology should have never fallen into common use – although due to the open-source nature of artificial intelligence, its common usage was rather unavoidable. The consequences of giving unlimited and (essentially) unrestricted access to softwares used for such exploitative content are daunting. Fittingly, the vast majority of deepfake sexual exploitation is happening in spaces where women are traditionally not represented. The technology industry as a whole has always lacked gender-diversity (alongside other kinds), and women who *are* represented frequently struggle with the treatment they receive (O'Mara, 2022).

As a knee-jerk reaction, it almost feels negligent for all governing parties involved to have allowed for the open-source development and dissemination of deepfake softwares given their overwhelming use for obscene, offensive, and violating content. This issue, however, does not stem from deepfake technology specifically; it was merely exacerbated by it. Deepfakes are merely a small scale representation of misogynistic behavior reinforced by technology.

Since its inception, the internet has *always* had a market for pornographic content, and a subsect of that market has *always* been users creating, viewing, and sharing content that contains exploitative and violating themes against women. While deepfakes did not *invent* non-consensual, obscene, and offensive pornography of women on the internet, the development of deepfake technology *thrives* off of it. It is precisely what gave it the popularity to be continuously developed and improved, raising the bar with every video.

Perhaps these development systems should have been regulated from the beginning. Maybe they should have never been made. In either case, nothing deleted from the internet is ever *really* gone in today's day and age — deepfakes are here to stay. How we move forward as a

society to safely and respectfully allow for open source communication development, and ensure the protection of victims of exploitative content, is still very much uncharted territory.

Implementing ways to proactively and effectively regulate artificial intelligence must be at the forefront of our future technological developments. We need more representation of, and consideration for, marginalized and vulnerable populations in the technology industry and its governing bodies if we want to continue seeing conducive and positive developments in artificial intelligence. Deepfakes are a cutting edge technology, with limitless bounds on what they *can* do. However, to preserve the privacy, reputation, safety, and autonomy of deepfake victims around the world, there ought to be major advancements in regulatory groundwork surrounding what they *cannot* do. Deepfakes — at their core, and despite their origins — are used to graphically degrade and exploit women on the internet. To chalk it up to "human nature" disregards the immediate issue at hand. As antithetical as it may sound, the integration and positive applications of artificial intelligence are being harmed by the very software it strives to advance. Deepfakes are arguably one of the greatest revolutions in machine learning technology. But until we can learn to properly harness new technologies for good, the industries they reside within and foundations they are built upon will forever be tainted. Deepfakes present a disservice to the technology industry, and an advancement's Achilles' heel.

References

Ajder, H., Giorgio Patrini, Cavalli, F., & Cullen, L. (2019, September). *The State of Deepfakes: Landscape, Threats, and Impact.* https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

Anyoha, R. (2017, August 28). The History of Artificial Intelligence. Science in the News. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/

Bernaciak, C., & Ross, D. A. (2022, March 14). How Easy Is It to Make and Detect a Deepfake?. SEI Blog. https://insights.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake/

Bregler, C., Covell, M., & Slaney, M. (1997, August 1). *Video Rewrite: Driving Visual Speech With Audio* ACM Conferences. https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/human/bregler-sig97.pdf

Chow, M. (2022, June 10). What Are the Positive Applications of Deepfakes?. *Jumpstart Magazine*. https://www.jumpstartmag.com/what-are-the-positive-applications-of-deepfakes/

Finger, L. (2022, September 8). Overview Of How To Create Deepfakes - It's Scarily Simple. *Forbes*. https://www.forbes.com/sites/lutzfinger/2022/09/08/overview-of-how-to-create-deepfakesits-scarily-simple/?sh=1b7a02b52bf1

GitHub. (2024). *Github Acceptable Use Policies*. Github. https://docs.github.com/en/site-policy/acceptable-use-policies/github-acceptable-use-poli

cies

Home Security Heros. (2023). *2023 State of Deepfakes*. 2023 State Of Deepfakes: Realities, Threats, And Impact. https://www.homesecurityheroes.com/state-of-deepfakes/#key-findings

Hsu, J. (2024, February 21). *Deepfakes are out of control – is it too late to stop them?*. New Scientist. https://www.newscientist.com/article/2418188-deepfakes-are-out-of-control-is-it-too-late-to-stop-them/#:~:text=A%20study%20last%20year%20by,cent%20using%20a%20woman%27s%20likeness

Jingnan, H. (2024, April 5). *Using AI to detect AI-generated deepfakes can work for audio - but not always*. NPR. https://www.npr.org/2024/04/05/1241446778/deepfake-audio-detection#:~:text=NPR%20identified%20three%20deepfake%20audio,available%20for%20individuals%20to%20use

Johnson, D., & Johnson, A. (2023, June 15). *What are deepfakes? how fake AI-powered audio and video warps our perception of reality*. Business Insider. https://www.businessinsider.com/guides/tech/what-is-deepfake

Justice Appropriations Subcommittee, Public Safety Subcommittee, Criminal Justice & Public Safety Subcommittee, & Harding et al. (2022, February 28). CS/HB 1453 Sexually Explicit Material. https://www.flsenate.gov/Session/Bill/2022/1453/Analyses/h1453b.JUA.PDF

Lyon, B., & Tora, M. (2023). *Exploring deepfakes*. O'Reilly Online Learning. https://learning.oreilly.com/library/view/exploring-deepfakes/9781801810692/B17535_TOC_ePub.xhtml

Mayhew, F. (2020, February 11). *Reuters creates prototype automated video match report led by AI sports presenter*. Press Gazette. https://pressgazette.co.uk/news/reuters-creates-prototype-match-report-led-by-ai-artificial-sports-presenter/

Meta. (2020, June 25). *Deepfake Detection Challenge Dataset*. AI at Meta. https://ai.meta.com/datasets/dfdc/

Mulvihill, G. (2024, January 31). *What to know about how lawmakers are addressing deepfakes like the ones that victimized Taylor Swift*. AP News. https://apnews.com/article/deepfake-images-taylor-swift-state-legislation-bffbc274dd178ab054426ee7d691df7e

Nanos, Georgios (2024, March 18). *VAE Vs. GAN For Image Generation.* Baeldung. https://www.baeldung.com/cs/vae-vs-gan-image-generation#:~:text=Moreover%2C%20VAEs%20are%20frequently%20simpler,detailed%20plausible%20data%20than%20VAEs.

Office, United States Government Accountability Office. (2020a, February). *Science and Tech Spotlight: Deepfakes.* U.S. GAO. https://www.gao.gov/assets/gao-20-379sp.pdf

Office, United States Government Accountability Office. (2020b, October 20). *Deconstructing Deepfakes – How do they work and what are the risks?*. U.S. GAO.

https://www.gao.gov/blog/deconstructing-deepfakes-how-do-they-work-and-what-are-risks

Picchi, A. (2019, April 17). *How tech's white male workforce feeds bias into AI*. CBS News. https://www.cbsnews.com/news/ai-bias-problem-techs-white-male-workforce/

Rep. Ocasio-Cortez leads bipartisan, bicameral introduction of Defiance Act to combat use of non-consensual, sexually-explicit "Deepfake" media. Representative Ocasio-Cortez. (2024, March 7). https://ocasio-cortez.house.gov/media/press-releases/rep-ocasio-cortez-leads-bipartisan-bicameral-introduction-defiance-act-combat

Science, Technology, Assessment, and Analytics. (2020, February). *Science & Tech Spotlight: Deepfakes*. The United States Government Accountability Organization. https://www.gao.gov/assets/gao-20-379sp.pdf

Seto, M. C. (2013). *Internet Sex Offenders*. American Psychological Association. https://psycnet.apa.org/doiLanding?doi=10.1037%2F14191-000

Somers, M. (2020, July 21). *Deepfakes, explained*. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Steck, E., & Kaczynski, A. (2024, January 22). *Fake Joe Biden robocall urges New Hampshire voters not to vote in Tuesday's Democratic primary.* CNN Politics. https://www.cnn.com/2024/01/22/politics/fake-joe-biden-robocall/index.html

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the Gap to

Human-Level Performance in Face Verification. https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf

Twitter. (2023, April). *Our synthetic and manipulated media policy | X help*. Twitter. https://help.twitter.com/en/rules-and-policies/manipulated-media

Wolf, Z. (2024, January 24). *The deepfake era of US politics is upon us*. CNN Politics. https://www.cnn.com/2024/01/24/politics/deepfake-politician-biden-what-matters/index.html