Influencing, disinformation, and fakes problems caused by Deep Fakes

FIT1055 IT PROFESSIONAL PRACTICE AND ETHICS

Kang Hong Bo | 32684673 | FIT1055 | 31/3/2023

Introduction

From a blog at Q5ID written in (2022). The first concept of the deep fake was first a program to rewrite the video which is created in 1997 by Christoph Bregler, Michele Covell, and Malcolm Slaney. The program altered existing video footage to create new content of someone mouthing words they didn't speak in the original version. This program was the first system to automate facial reanimation completely. Following the video rewrite program, Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor developed the active appearance model (AAM) in 2001. AAM is a computer vision algorithm that compares a new image to a statistical object shape and appearance model. This study improved the efficiency of face matching and tracking significantly.

In recent years, the rise of deep fake technology has sparked growing concern over its potential to spread disinformation, manipulate public opinion, and undermine trust in media and institutions. Deepfakes are computer-generated audio, video, or images that are manipulated using machine learning algorithms, enabling the creation of highly realistic content that can be used to deceive or mislead viewers. With the increasing ease of creating and using deep fakes, there is a possibility that they could be used to spread false information and intercept political systems, cause a social problem and political issue, and damage reputations and trust in institutions.

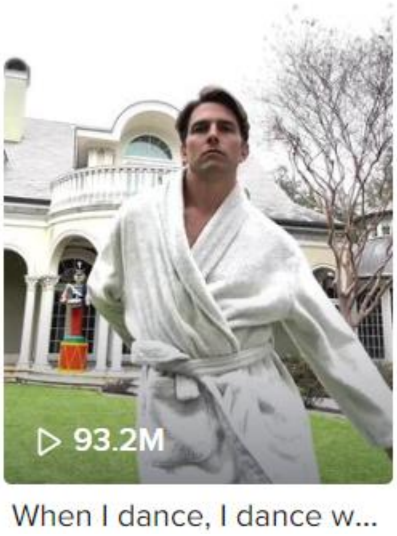
Moreover, the potential uses of deep fakes are not limited to political or social manipulation. They can also be used for malicious purposes such as fraud, cybercrime, and cyberbullying. For example, deep fakes can be used to create convincing phishing emails or to impersonate someone online, leading to identity theft and financial fraud. Additionally, they can be used to harass or threaten individuals by creating fake videos or images that appear to show them engaged in illegal or inappropriate behavior.

As deep fake technology becomes more advanced and accessible, the risks associated with its use are likely to increase. This highlights the importance of ongoing research and the development of strategies to detect and prevent deep fakes. It also emphasizes the need for individuals to become more media-savvy and critically evaluate the authenticity of digital media.

Overall, the emergence of deep fakes poses a significant threat to digital media's credibility and our ability to trust the information we encounter online. It is crucial that we take proactive steps to address this threat and safeguard the integrity of digital media and democratic processes.

Literature review

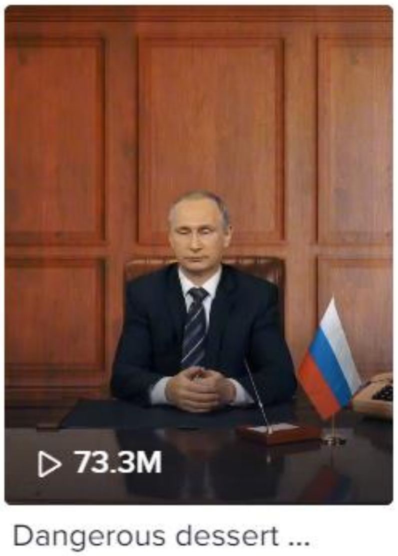
In this decade deep learning and artificial intelligence (AI) have led to lots of new technology and software. Deepfake is one of the software which is involved in both technics. One of the popular deep fake technologies is face swapping, which replaces the face of one person in a video with another person. The DeepFaceLab project on GitHub (iperov,2020) is an example of a tool that creates face swaps using deep learning algorithms. This project said that 95% of deep fake videos are created with DeepFaceLab and there are some channels that use deep faking celebrity's or politician's face to create content for example, deeptomcruise (TikTok content creator 5.1Million follower) making content which change the face to Tom Cruise and 1facerussia (TikTok content creator 10.1Million follower) making content with using Putin's face.



When I dance, I dance w...

(deeptomcruise 2022-12-27 video

with 93,2 Million views)



Dangerous dessert ...

(1facerussia 2022-2-17 video

with 73.3 Million views)

As we can see every people can access and use this software and impersonate others since it is open-source software. This will lead to the misuse of software such as political propaganda, revenge porn, and other types of fraud. In addition to disinformation by using Deepfake, an article by Satariano and Mozur (2023) published in The New York Times has shown two videos of broadcasters who appeared intended to promote the interests of the Chinese Communist Party and undercut the United States for English-speaking viewers by having the video generated by AI from pro-China bot with the news of the United States lacking action against gun violence and the other news which attempt to promote China's role in geopolitical relations at an international summit meeting. Besides this disinformation, deep fakes also involve in the disinformation of the war problem. An article published by Allyn (2022) at NPR organization has stated that a video of the Ukrainian President Volodymyr Zelenskyy appearing to tell his soldiers to lay down their arms and surrender the fight against Russia that ran about a minute long and it is circulated on social media and was placed on a Ukrainian news website by hackers Wednesday before it was debunked and removed.

The dangers of deep fake technology were highlighted by a recent report by Hao (2021) published at MIT Technology Review, which revealed the existence of a deep fake app called "Face-swaps Women into Porn." This app allows users to replace the faces of women in pornographic videos with those of their own choosing, raising concerns about the potential harm to the individuals whose faces are being swapped. While an article in Vice written by Cole (2018) highlighted that Reddit users created an app called FakeApp, which allows users to replace the faces of actresses in pornographic videos with those of their own choosing raising concerns about the potential harm to the individuals whose faces are being swapped. Besides a Twitter post by Luccioni (2021) pointed out that having this machine learning algorithm that can undress women is an ethically dubious task. For extension, another article which is published in Vice by Cole (2019) also wrote out that an app called Deep Nude programmed to undress the women in the photo, and she emphasis that deep fakes aren't only constructed disinformation but also aimed at women. This is an "invasion of sexual privacy," Danielle Citron, professor of law at the University of Maryland Carey School of Law, who recently testified to Congress about the deep fake threat, told Motherboard.

Moreover, Deepfake technology also has the potential to create realistic audio and video simulations that can be used to manipulate people and events, as highlighted in a new at GIZMODO by Brown (2019). In this article, he states that news about a CEO who is being lied to by a call from a person who modified the voice to the boss of the CEO and make him transfer money to a Hungarian supplier. And from this news, we see how an AI-assisted voice can deep fake as someone's voice which is normally heard and still not able to recognize. In addition, an article in Forbes written by Brewster (2021) highlighted a bank manager in Hong Kong who received a call from a man whose voice he recognized—a director at a company with whom he'd spoken before. The director's company was about to make an acquisition, so he needed the bank to authorize some transfers to the tune of $35 million. But he didn't know that he is in an elaborate swindle which is conducted by fraudsters by using deep voice technology to clone the director's speech. In these two cases, we can see that scamming using the voices of people we know or recognize and by having this cloning algorithm is quite hard to differentiate whether it is a fake synthesis of voice or is the true voice.

In short, deep fake technology has introduced a new level of risk to the world. While it has enabled creative and entertaining content, its potential misuse can lead to devastating consequences. From political propaganda to revenge porn, and disinformation to identity theft, the dangers of deep fakes are significant. The ability to manipulate audio and video simulations to clone someone's voice or face raises serious ethical concerns, particularly for women who are disproportionately affected. It is essential to develop robust methods to detect and prevent the misuse of this technology to protect individuals' privacy, safety, and security. As deep fake technology continues to evolve, it is crucial to consider its potential negative impacts and take appropriate measures to minimize them.

Problem statement – What, why, and How

Goodfellow et al. (2014) have published a new framework for estimating generative models via an adversarial process to train two models. These two models can train by continuously sending back and forth the data, and the entire system can be trained with backpropagation. And this framework is called a Generative adversarial network. Having this GAN mindset Schreiner (2022) has summarized how GAN is used in this short time and deep fake uses this algorithm as the basis. After that more researchers combined GANs and trained GANs with other GANs which started how credible the image is produced. And since Nvidia spot out that to train the network stage, which let the forger AI learns to create low-resolution image then gradually increase them. All these improvements of the GANs lead to the research above where deep fake porn as it is one of the major problems in this area.

Research studies have highlighted the potential impact of deep fakes on various aspects of society. Similarly, a research project by Suwajanakorn, Seitz, and Kemelmacher-Shilizermnan (2017) has researched lip sync using an Obama video. The study used a deep fake video to create 4 different videos with the same speech by Obama and by this research, we can see that it is easy to create false news just by having audio then sync to the news which is published before. By having this algorithm, we can see how we can reproduce a video that makes it seems real and hard to recognize by our raw eyes. Furthermore, scamming through calls with a regenerated voice could easily scam people, what if making a lip sync video with a regenerated voice of others it will make us easier to trust in this swindle.

Moreover, a report published by Helmus at the RAND Corporation (2022), highlighted the potential of deep fakes to exacerbate existing social and political tensions. The report noted that deep fakes can be used to create fake news stories that play on people's fears and biases, leading to increased polarization and conflict. The widespread availability of deep fake technology also poses a threat to the integrity of digital media and raises concerns over the potential for malicious actors to use deep fakes to spread false information for personal or political gain. In addition, the use of deep fakes has the potential to reduce trust in institutions, including the media, government, and law enforcement. As deep fakes become more convincing, people may question if it is a piece of real news, leading to a breakdown in trust and a rise in conspiracy theories and misinformation. This could have serious consequences for the functioning of democracies and societies worldwide. As technology continues to evolve, it is serious that individuals and organizations remain vigilant and take steps to protect themselves against the potential harms of deep fakes. Ultimately, the widespread use of deep fakes raises important questions about the future of digital media and the role of technology in shaping our societies and cultures.

So how we can know that the video we see is not a fake video? In a short article was released on CNN Business by O’Sullivan (2019), United States started a program call The Pentagon which is in Defense Advanced Research Projects Agency (DARPA), is working with several country researchers to get ahead deep fake. They give the computer fake and real videos to train the computer to detect deep fake videos. They are also focusing on fake audio.

Conclusion and Discussion

In conclusion, Deep fake technology has advanced rapidly in recent years and is becoming increasingly sophisticated, making it easier to create convincing fake videos and images. It poses a significant risk to individuals, organizations, and society. This includes threats to privacy, security, democracy, and human rights. As we know the detection of deep fakes is not yet done. Therefore, we need to be aware to avoid being one of the victims. Additionally, the responsible use of deep fake technology must be emphasized, particularly in industries such as entertainment and advertising, where the line between reality and fiction can be blurred. Overall, deep fakes are a rapidly evolving technology that requires careful consideration and management to maximize their benefits while minimizing their negative consequence.

References

Anonymous, A Quick History of Deepfakes: How It All Began, Q5ID, November 16, 2022: https://q5id.com/blog/a-quick-history-of-deepfakes-how-it-all-began#:~:text=Deepfakes%20started%20with%20the%20Video,speak%20in%20the%20original%20version.

iperov, DeepFaceLab, Github, 2020: https://github.com/iperov/DeepFaceLab

Satariano A., Mozur P., The People Onscreen Are Fake. The Disinformation Is Real., New York Time, Feb 7,2023: https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html

Allyn B., Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn, NPR, March 16, 2022: https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

Hao K., A horrifying new AI app swaps women into porn videos with a click, MIT Technology Review, September 12, 2021: https://www.technologyreview.com/2021/09/13/1035449/ai-deepfake-app-face-swaps-women-into-porn/

Cole S., We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now, Vice, 25 January 2018: https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley

Luccioni S., Twitter post, October 28, 2021: https://twitter.com/sashamtl/status/1453491661720391685

Cole S., This Horrifying App Undresses a Photo of Any Woman with a Single Click, Vice June 27, 2019: https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman

Brown J., Scammer Successfully Deepfake CEO's Voice to Fool Underling into Transferring $243,000, GIZMODO, September 3, 2019: https://gizmodo.com/scammer-successfully-deepfaked-ceos-voice-to-fool-under-1837835066

Brewster T., Fraudsters Cloned Company Director’s Voice In $35 Million Bank Heist, Police Find, Forbes, Oct 14, 2021: https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=230ec98d7559

Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., Generative Adversarial Networks, arxiv, 10 Jun 2014 : https://arxiv.org/abs/1406.2661#

Schreiner M., Deepfakes: How it all began – and where it could lead us, The decoder, April 28, 2022: https://the-decoder.com/history-of-deepfakes/

Suwajanakorn S., Seitz S. M., and Kemelmacher-Shlizerman I., ACM Transactions on Graphics 36(4), Grail, July 2017. https://grail.cs.washington.edu/wp-content/uploads/2017/08/suwajanakorn2017sol.pdf

Helmus, Todd C., *Artificial Intelligence, Deepfakes, and Disinformation: A Primer,* RAND Corporation, PE-A1043-1, July 2022. As of March 22, 2023: https://www.rand.org/pubs/perspectives/PEA1043-1.html

O'Sullivan D., When seeing is no longer believing, CNN Business, 2019: https://edition.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/

In our unit, we will apply the APA citation technique. APA style uses the author/date method of citation in which the author's last name and the year of the publication are inserted in the actual text of the paper.

You may refer to more details on the APA reference style here