The Emerging Threats of Deepfake Attacks and Countermeasures

# Shadrack Awah Buo

## 2020

The continuous evolution of cybercrime has culminated with deepfakes which severely magnify the threats of traditional frauds

# Scholarcy Highlights

* Developments in Artificial Intelligence (AI) have led to the emergence of deepfake technologies (DT), which pose a significant threat to global institutions
* The continuous evolution of cybercrime has culminated with deepfakes which severely magnify the threats of traditional frauds
* [17] Gosse, C; Burkell, J;, “Politics and porn: how news media characterizes problems presented by deepfakes,” Critical Studies in Media Communication, vol 37, no. 5, pp. 497-511, 2020
* It is critical to invest in awareness and training to help identify early signs of deepfake attacks
* In terms of future research, Facebook [38] has partnered with Microsoft and issued a public challenge worth $10 million, that will help produce a technology that can be us by everyone to detect deepfakes
* This process of crowdsourcing knowledge will help techs companies develop an effective solution against deepfake attacks

# Scholarcy Summary

## INTRODUCTION

Developments in Artificial Intelligence (AI) have led to the emergence of deepfake technologies (DT), which pose a significant threat to global institutions.

The faces of politicians being edited onto other individuals’ bodies who appear to say things that they never did are becoming commonplace

This growing phenomenon has been used in political scenarios to misinform the public on various debates [2].

Cybercriminals have taken advantage of the technology to misinform and defraud businesses and the individual [4]

Creating such deepfakes media requires expertise and specialist software and hardware.

This paper examines the threats posed by deepfakes that affect global institutions and the negative implications.

It provides an overview of deepfakes, how they are created and the benefits of its use.

The potential solutions to prevent deepfakes and what can be done to mitigate future risks are discussed

## AN OVERVIEW OF DEEPFAKES

The term deepfake is a combination of “deep learning” and “fake” [5]. The deepfake phenomenon started on the social media platform, “Reddit.” An anonymous user shared an altered pornographic video of a celebrity, their face had been swapped with a porn actor.

GANs employs two Artificial Neural Networks (ANNs) working together to create deepfakes

These ANNs are known as “detector” or discriminative network, and “synthesizer” or generative network [5].

They are trained on a large dataset of videos, images, and sounds to produce high-quality deepfakes [5].

It is expected that future GANs algorithms will be trained on smaller datasets and produce more convincing, higher quality deepfakes [11].

These developments would allow cybercriminals to create more authentic deepfakes which would have a devastating impact on their victims.

Aside from the risks that DT can pose, the section looks at some of the benefits and positive applications that DT can provide for society

## BENEFITS OF DEEPFAKE TECHNOLOGY

Voice assistant technologies such as Apple’s Siri and Window’s Cortana uses machine learning (ML)

These technologies apply similar AI-based algorithms to assist the end-user by answering queries and delivering content by voice-activated commands [7].

Ongoing research [13] is exploring ways to develop an AI system that will automate the process of producing educational content using DT.

It can be used to develop realistic virtual environments and natural-sounding in-game assistants which improve the user experience [4].

DT can have positive uses in health and social care

It can help individuals deal with the loss of loved ones by developing a digital version of their loved ones [4].

This will assist medical professionals in the early detection of illnesses and potentially save more lives

## THREATS OF DEEPFAKES TECHNOLOGY

In contrast to the benefits of DT pose major threats to global institutions for the following reasons; they used to defraud businesses and raise cybersecurity concerns to organisations.

They can be used as a source of misinformation in politics and the court of law.

This section explores the threats of deepfakes to Businesses, politics, and judicial systems

## Threats to Judicial Systems According to

Pfefferkorn [22], evidence tampering is one of the major threats posed by deepfakes in the judicial system.

Further issues may arise during cross-examinations when an offering party testifies affirmatively concerning details of a deepfake video while the opposing party denies the contents of the video [22].

This would negatively impact court cases because deepfakes might cause additional caseloads, and cost money and time to verify and authenticate the evidence before it can be admissible in court [22].

In a UK child custody case, a deepfake audio file was presented as evidence to the court by the mother [8].

There is an urgent need for new and effective countermeasures to prevent evidence tampering in future court cases

## Threats to politics

Another threat that can strengthen with the use of deepfakes is disinformation within politics [17].

One notable example was the circulation of an altered video of an American politician, Nancy Pelosi on social media.

In the video, she appeared intoxicated while mispronouncing her words [18].

Despite bipartisan calls for the video to be taken down, a Facebook spokesperson confirmed that the videos will not be removed because the platform does not have policies that dictate the removal of false information [20].

This has prompted world governments to look for ways to regulate the use of DT [7].

There is a growing need for policies to regulate the use of deepfakes on social media platforms for political gain

## Threats to Businesses

In addition to its impact on the legal system and politics, DT can have an adverse budgetary impact on businesses.

Symantec, a cybersecurity company, revealed that deepfakes and social engineering was used to defraud three CFOs (Chief Financial Officer) of undisclosed substantial funds [29].

In addition to these findings, Forrester Research [29] predicted a monetary loss of $250 million by the end of 2020 from deepfake frauds.

The installation of face scanners to grant access to restricted areas

If these areas are breached with the use of deepfakes, this could lead to unauthorised access to sensitive information and intellectual property.

Such an attack could lead to monetary loss due to the costs incurred from containing the breach, compensating customers, and heighten security costs [30]

## POTENTIAL SOLUTIONS TO DEEPFAKES

A multitude of solutions have been proposed and deployed against deepfakes.

Having access to a dedicated forensics team and expert tools needed for detecting deepfakes can be costly to manage and maintain [34]

To solve this issue, Lee and Un [34] proposed digital forensics as a service model, which leverages cloud computing technologies to provide robust forensics services at a cheaper price.

Using blockchain and smart contracts, Hasan, and Salah’s solution acts as a transparent digital signature on media content to prove their authenticity.

This solution relies on time-sequence logs to track the history of media contents, monitoring where it was used online to later determine their origins [28].

Marcel, “Vulnerability assessment and detection of Deepfake videos,” Crete, 2019

## Findings

Using CNNs, researchers [27] were able to successfully detect and identify 99.1% of deepfakes

## CONCLUSION

The continuous evolution of cybercrime has culminated with deepfakes which severely magnify the threats of traditional frauds.

Technology firms and governments should consider passing legislation that will criminalize the use of deepfakes with the intent to defame the character of individuals.

In terms of future research, Facebook [38] has partnered with Microsoft and issued a public challenge worth $10 million, that will help produce a technology that can be us by everyone to detect deepfakes.

This process of crowdsourcing knowledge will help techs companies develop an effective solution against deepfake attacks.