Title: AI in the Wrong Hands chapter 2

Author: Kharim Mchatta

Date: 13/6/2024

Disclaimer: This article is intended for educational purposes only. The author assumes no responsibility for any malicious use of the material or knowledge contained within this article.

In the first chapter of this article we talked about how sophisticated cyber threats have become and this is due to the rise of AI to generate digital forgery. When we talk about AI in the context of cyber threats we refer to how cybercriminals make use of Generative Artificial Intelligence (GenAI) which encompasses deep fakes and chatbots that make use of large language models (LLMs).

In this chapter we will be focusing on the deep fakes and how cybercriminals are making use of deep fakes for malicious purpose. Deep fakes made its appearance around the year 2017 where people used them for various purpose for example entertainment and pranks. Previously it was easy to tell that a certain deep fake was not real, but as technology advanced it has become ore complicated to differentiate the reality and what's fake.

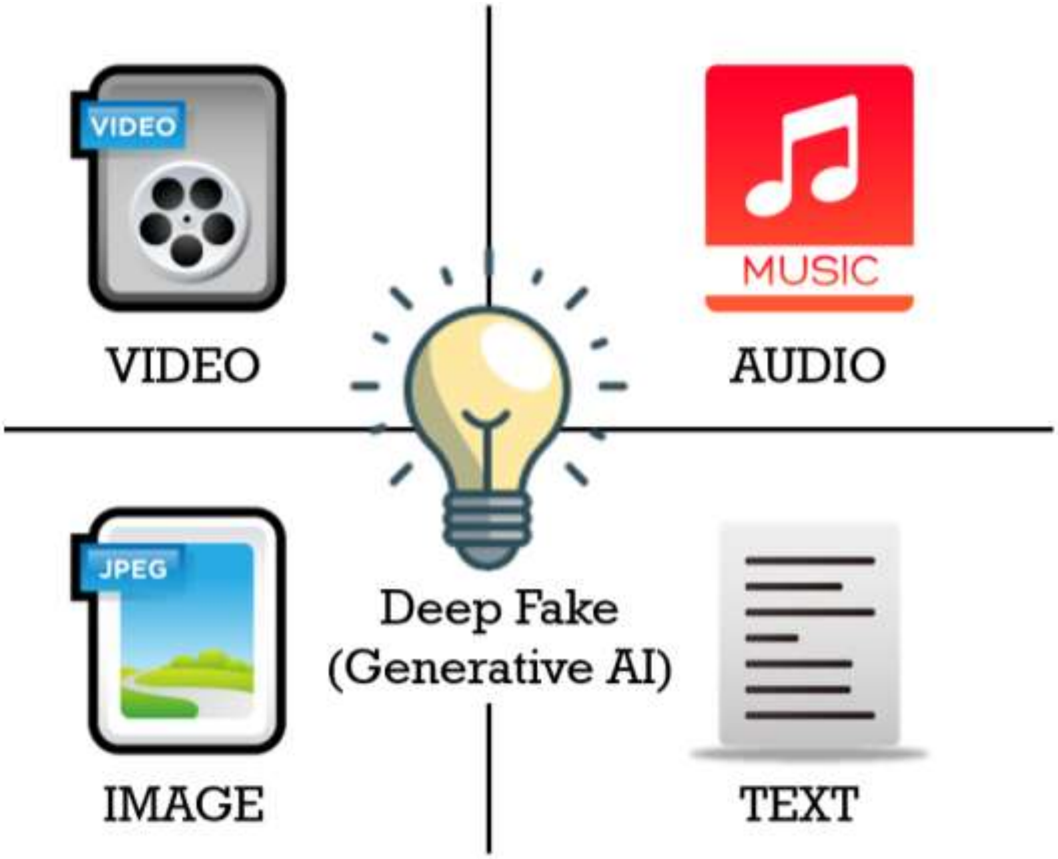
The image below is not a genuine representation of former President Barack Obama; rather, it is a deep fake video. This video features Jordan Peele simulating Obama, and without access to the behind-the-scenes footage, it would be difficult to discern that it is a deep fake. This example highlights the sophisticated advancements in deep fake technology in the current era.

Now let's dive in to more details on what these deep fakes are.

When we talk about deep fake, its all about the ability of an individual to manipulate digital media and text. Deep fakes give people the ability to swap faces in videos, images, change voices, and alter texts in documents and all this is through. Deep fake in 2024 have now reached to a level of maturity taking digital forgery to the next level.

Digital forgery is not a new concept since it was being done through graphic based and content changing alteration in the multimedia industry back in the days. Earlier generations of deep fakes did not have natural movements and reactions but also there was sudden shifts in audio quality, mismatches in color and lighting etc making it easy to identify that a certain digital media is a deep fake. The new generation deep fakes are more sophisticated that no one is able to recognize one by just looking at a media. Special tools are required in order to tell if a media is original or fake.

There are several types of deep fakes which you should be aware of and they are as shown on the image below



As illustrated in the image, deep fakes can manifest in various forms, reinforcing the notion that "seeing is no longer believing." This technology's ability to distort reality has advanced significantly, fostering a climate of distrust. The capabilities of deep fake technology continue to improve, raising concerns about its impact on societal trust.

The term "deep fake" encompasses a range of meanings. Unfortunately, it is often associated with malicious activities, similar to the word "hacker," which tends to evoke images of black hat hackers despite also referring to ethical, white hat hackers. It is important to recognize that not all deep fakes are inherently harmful; they can be used for a variety of purposes. Here are some examples of how deep fake technology is being utilized:

- **Entertainment and Media:** Deep fakes are used to create realistic visual effects in movies and TV shows, allowing actors to appear younger, portray multiple roles, or perform dangerous stunts without risk.
- **Education and Training:** In educational settings, deep fakes can create realistic simulations for training purposes, such as medical students practicing surgeries or pilots undergoing flight simulations.
- **Marketing and Advertising:** Companies use deep fake technology to create engaging and personalized advertisements, featuring realistic avatars that interact with potential customers.
- **Art and Creativity:** Artists and content creators use deep fakes to push the boundaries of their work, creating innovative and thought-provoking pieces that challenge traditional notions of art and media.
- **Accessibility:** Deep fakes can be employed to generate more accurate lip-syncing for dubbed foreign language films, making them more accessible to global audiences. Additionally, they can help create virtual assistants that provide more natural and engaging interactions for users.
- **Historical Preservation:** Deep fake technology is used to restore and animate old footage, bringing historical figures and events to life for educational and documentary purposes.

- **News Reporting:** Media companies can use deep fake technology to create virtual news anchors. This allows for consistent and round-the-clock news presentation without the need for human reporters to be physically present. These virtual reporters can deliver news with high accuracy and even adapt their style to suit different audiences.

On the other hand, deep fake technology is increasingly being exploited for illicit purposes, primarily by cybercriminals. Some of the illegal applications include:

- **Scams and hoaxes:** Cybercriminals can leverage deep fake technology to create false claims, scams, and hoaxes that may undermine the legitimacy of individuals or organizations. For example, a cybercriminal could produce a fabricated video of a manager admitting to illegal activities such as financial fraud or disseminating false information about the organization's operations. This could significantly impact the organization's reputation and brand image.
- **Election Manipulations:** Politicians are often the prime targets during elections. Cybercriminals may use deep fakes to stain the reputation of politicians, potentially affecting their vote counts and influencing election outcomes.
- **Social Engineering:** Social engineering has long been a favored tactic for cyberattacks due to its high success rate. Deep fake technology has enhanced the effectiveness of social engineering, making it easier for cybercriminals to commit fraud. For instance, in the UK, cybercriminals employed a social engineering technique known as vishing. They impersonated the CEO of a parent company in Germany, convincing the CEO of an energy company in the UK to transfer €220,000 to a fraudulent supplier’s bank account.
- **Identity theft and financial fraud:** Cybercriminals have found new methods of committing identity theft and financial fraud through the use of deep fakes. By creating falsified documents or cloning individuals' voices, they can establish accounts, purchase products, and commit fraud by impersonating their victims.

Out of all these cybercrime categories mentioned above the leading AI cybercrime fraud is leading the way.

Even though deep fake technology has advanced significantly, individuals can still detect its presence through various indicators. Signs of a deep fake often include unnatural movements or anomalies such as;

1. Unnatural eye movements, as deep fake technology struggles to replicate natural eye behavior.
2. Lack of blinking, which deep fakes often fail to simulate realistically. When there is no blinking then it is an indicator of a deep fake.
3. Irregular body shape, as deep fakes typically focus on facial features rather than entire body proportions.
4. When the video has unnatural facial expressions and facial morphing
5. Difficulty generating realistic hair characteristics, such as frizzy or messed up hair. Deep fakes have a hard time generating realistic individual characteristics such as hair.
6. Abnormal skin coloration, with deep fakes often unable to accurately reproduce natural skin tones.
7. Awkward facial expressions that diverge from natural human behavior. Most deep fake will have weird facial expressions which are far from the natural.
8. Odd body features or anomalies, such as incorrect numbers of fingers or glitches in the video. example you may find characters having more fingers in one hand or glitches in the video, or having odd standing positions.
9. Poor lip syncing, where spoken words do not match the movements of the lips. When there is bad lip syncing where the words being said does not align with the lip movement when words are being spoken then it is a deep fake.
10. Unnatural movement of people or animals, including body parts disappearing as they interact with objects.

While deep fake technology has advanced significantly, there are still elements it struggles to replicate realistically. Despite this, to an untrained observer, detecting a deep fake remains challenging due to its current sophistication. As deep fakes continue to improve, aided by ongoing technological advancements, their detection may become even more difficult without the use of specialized tools.

In addition to individuals falling victim to sophisticated attacks, organizations are also susceptible to AI-driven cyber threats. Safeguarding organizational assets is not merely a technological challenge; it also represents a human issue that necessitates adjustments spanning people, processes, and technology. Protecting financial transactions and data from evolving cyber threats requires a comprehensive approach. It is important for organizations to strengthen their verification processes, particularly for fund transfers. Clear and well-defined procedures should be established and adhered to when handling requests for fund transfers within the organization.

One of the most common attack that is faced by organization is deep fake phishing. While phishing has historically been a persistent threat, it continues to be highly effective in compromising organizational security. In recent years, attackers have advanced their tactics by leveraging artificial intelligence (AI), thereby enhancing the sophistication of phishing attacks.

Examining deep fake phishing reveals a relatively new technique employed by cybercriminals to manipulate victims and commit fraud. This approach combines sophisticated social engineering tactics with deep fake technology to deceive individuals effectively. Deep fake phishing operates on the foundational principles of social engineering, manipulating users into divulging sensitive information they would not typically share under normal circumstances. This form of attack capitalizes on exploiting trust and often circumvents traditional security measures. Attackers utilize deep fake technology to weaponize phishing attacks in various ways, including:

- Emails or messages: Organizations are increasingly susceptible to financial losses due to email compromise attacks, amounting to billions annually. The advent of deep fakes has heightened the threat posed by this attack vector, as cybercriminals personalize their emails to appear more authentic and credible.
- Video calls: Cybercriminals are leveraging deep fake technology to conduct Zoom calls, aiming to persuade victims into disclosing confidential information or manipulating them into making unauthorized financial transactions.
- Voice messages: Cybercriminals now clone voices to deceive individuals within the victim’s network, attempting fraud. Voice cloning has become increasingly accessible, requiring only a 3-minute audio clip of your voice. These deep fakes are used for leaving voicemails, sending voice notes, or making live calls. In 2022, 37% of organizations reported incidents of deep fake voice fraud.

The devastating impact of deep fakes is evident, and we have only begun to understand its potential implications. This underscores the critical importance for organizations to prioritize their vigilance against deep fake phishing attacks.

- The rapid advancement and accessibility of deep fake technology are alarming, posing significant concerns. In 2023, incidents of deep fake phishing and fraud saw a drastic increase, highlighting the escalating threat landscape.
- Deep fakes are highly targeted that enable cybercriminals to craft sophisticated, personalized attacks tailored to specific organizations and individuals they aim to exploit.
- Deep fakes are challenging to detect due to their ability to replicate writing styles, clone voices with high accuracy, and generate AI-generated avatars that closely resemble the victim. These capabilities make deep fake phishing attacks particularly difficult to identify and mitigate.

Having examined how cybercriminals leverage deep fake technology in social engineering attacks, specifically targeting phishing, and recognizing the imperative for organizations to address deep fake phishing seriously, we can now explore strategies for organizations to safeguard themselves against these threats.

- Organizations should enhance their awareness programs by incorporating education about synthetic content. Emphasis should be placed on caution against trusting online personas, individuals, or identities solely based on videos, photos, or audio clips present in their online profiles.
- Training programs should incorporate deep fake recognition and reporting procedures. Employees need to be equipped with the skills to identify potential deep fakes by recognizing visual anomalies such as inconsistencies in lip syncing, unnatural movements, and suspicious requests. This proactive approach is crucial in mitigating the risks posed by deep fake technology.
- Organizations should implement AI-based protection mechanisms along with robust authentication methods to mitigate fraud risks effectively. Techniques such as multi-factor authentication and zero-trust frameworks can significantly enhance security by reducing the likelihood of identity theft and preventing lateral movement within networks.

Several documented cases reported by trusted sources illustrate the real-world impact of deep fake technology. Let's explore a few scenarios to provide context and understanding of these occurrences.

- Our first case was reported by CNN that an employee that was working for a multinational organization was tricked into paying out $25 million. The cybercriminals made use of deep fake technology to pose as the company's CFO in a video conference call and this was according to Hong Kong police. The attacker convinced the employee into jumping into a video conference where he saw other several people who looked like fellow staff members but were actually deep fake characters.
  Link to the full article: https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html

- In our second case, the guardian reported that the CEO of wpp was targeted by deep fake where cybercriminals impersonated him using a fake whats app account, voice clone and YouTube video use in an online meeting.

  Link to full article: https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam

- In our third case we look at cbs new on their segment of age of AI they reported that cybercriminals are now using AI to clone voices to scam people. A woman called Jennifer got a call from her 15-year-old daughter called Bryana called her and said **“mummy this bad man got me help me help me help me”** and then a man gets on the call and told Jennifer **“listen here I have your daughter”** and he went on to ask for a million dollars.

AGE OF AI
CBS EVENING NEWS WITH NORAH O'DONNELL
HOW SCAMMERS ARE USING A.I. TO CLONE VOICES

Link to the full video: https://youtu.be/pJZYd_65xs4?si=1AZrSNdJOL0RpIDq

- In our last case NBC new reported that a man called john got a call from his daughter who said **“I have had an accident and I need your help”** and another voice joined the call and said **“im going to give your daughter back to you but am going to need some cash”**

Link to full video: https://youtu.be/V6_jCGzR020?si=0oxKtykMpvElt6Ts

CONCLUSION

As deep fake technology continues to advance, it is imperative that we remain vigilant to avoid falling victim to such attacks. The potential consequences are severe, and ensuring our safety from these threats is crucial.

Organizations must take this threat seriously and strive to understand these sophisticated attacks to effectively protect themselves. Deploying AI security mechanisms and implementing a zero-trust architecture are essential steps in safeguarding against deep fake threats. Current deep fakes represent only the beginning of this technology's capabilities, and as it evolves, their legitimacy will increase. Therefore, we must all stay alert and proactive in combating these attacks.

REFERENCE

https://arxiv.org/html/2402.04373v1

https://www.fortinet.com/resources/cyberglossary/deepfake

https://www.trendmicro.com/en_us/research/24/b/deepfake-video-calls.html

https://www.crowdstrike.com/cybersecurity-101/social-engineering/deepfake-attack/

https://www.theguardian.com/technology/article/2024/may/10/ceo-wpp-deepfake-scam

https://www.forbes.com/sites/forbestechcouncil/2024/01/23/deepfake-phishing-the-dangerous-new-face-of-cybercrime/

.