Combating Misinformation and Deep Fakes

Roman Kazorin
School of Governance and Politics, MGIMO University
Andrey Navolokin
School of Governance and Politics, MGIMO University

Abstract

With the advent of the 21st century, mankind has entered a completely new, previously unknown world of technology. The development of knowledge-intensive industries does not stand still, and today both public and private corporations around the world are engaged in the development of advanced technologies. Over the last 3 years, artificial intelligence (AI) has taken a huge step in its development.

By now, applications and devices with artificial intelligence are all around us, in our homes, in our offices and even in urban infrastructure. AI already has extensive learning capabilities and has started to replace humans in many areas of life, but it also makes our jobs easier and allows us to spend more time doing the things we love. In the context of the capabilities and associated risks of artificial intelligence, particularly in relation to its autonomous training, a diverse range of perspectives is pertinent.

The possible negative consequences of the development of artificial intelligence, among other respected personalities, were discussed by Stephen Hawking, who argued that the emergence of artificial intelligence could be the "worst event in the history of our civilization" unless society finds a way to control its development[1].

There is no doubt that artificial intelligence has been around since the middle of the last century. Thus, in 1969, the Shakey robot was built. He could

[1] https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html

reason about his actions, analyse commands, breaking down a task into simple partsa—It was the first robot in the history of mankind, which combined logic with physical actions. However, artificial intelligence as we know it today is a relatively recent phenomenon. In 2015, a research organization, OpenAI, was founded to develop artificial intelligence. And with the emergence of platforms such as ChatGPT, Midjorney and others, the world, people, skills and training requirements have changed completely.

Introduction

In recent years, fake photo and video production technologies have become widespread. They are created by means of computer image synthesis technologies based on artificial intelligence, which transfer facial features from a human image to the target photo (video) with a high degree of credibility. These technologies are referred to as deepfakes.

The synthesis of images, video or audio when applying deepfake technologies may not have socially dangerous purposes and do not necessarily affect the rights of other citizens. For instance, their usefulness is obvious in cinema, art and advertising. However, in the digital age, the emergence of deepfakes presents a serious challenge in the fight against misinformation. After all, these technologies allow to create audio and video content that is strikingly realistic, where a person's face or voice can be replaced or synthesized with amazing accuracy. This, in turn, can play a decisive role in influencing many social and political processes[2].

Deepfakes can cause widespread misinformation by presenting audio-visual content that is convincing but completely fabricated. As a result, the reputation of individuals and celebrities suffers, and fake news becomes more

[2] https://cyberleninka.ru/article/n/dipfeyki-kak-ugroza-pravam-cheloveka

widespread. Ultimately, the main threat of deepfakes appears to be their influence on public opinion, and even on elections.

Accordingly, the emergence of deepfakes in the digital landscape requires an active response from both the state and large corporations, which in turn will span technological, legal and societal norms.

Real World Impact

In recent years, deepfake technology has emerged as a tool capable of significant real-world impact. Situations that have recently occurred in such countries as Gabon and the United States prove this opinion.

In 2019, Gabon witnessed a politically charged situation fuelled by a deepfake video of President Ali Bongo[3], as reported by the French Institute for Environmental and Individual Risks ("IFREI"). Released amidst rumours about Bongo's health, the video depicted him giving a speech in an unnatural manner, contributing to political unrest and a subsequent failed military coup. This incident highlights the potential of deepfakes to destabilise governments and manipulate political landscapes.

As Politico Magazine reports, there have been several glaring cases of deepfakes used in the U.S. over the past few years. In May 2019, for instance, a video of a "drunk" Nancy Pelosi, speaker of the U.S. House of Representatives, went viral. The footage was edited to slow down the politician's speech, giving the impression that she was inebriated. According to another Politico article, in November 2018, a former Trump spokesperson posted an edited video online, which misleadingly suggested that a CNN reporter had aggressively interacted with a White House staff member[4]. Both instances gained considerable attention on social media.

[3] https://www.ifrei.org/article50-2019-Failed-Coup-in-Gabon-The-Deepfake-Theory
[4] https://www.politico.eu/article/deepfake-videos-the-future-uncertainty/

Referred to as “cheap fakes”, these videos were made using simple video-editing techniques. Despite being rapidly exposed as false, they significantly influenced media coverage worldwide for a number of days.

These examples illustrate the wide range of consequences that can arise from deepfakes. As this technology continues to advance and become more accessible, the potential for exploitation in different fields grows significantly. This raises major ethical, societal, and legal considerations.

The Dual Role of AI

Whether AI solely causes harm to modern society is a pivotal question, and reaching a consensus on it is impossible. Artificial intelligence is at a tipping point where its potential to bring benefit or harm is in the hands of not only its creators, but also its users[5]. The dual role of artificial intelligence in both creating and detecting deepfakes is particularly tangible in the context of fighting fake news. AI detection methods are becoming increasingly sophisticated in detecting anomalies that distinguish fake from genuine content and that cannot be detected at a glance. Such detection methods are often built on identifying frames or audio cues that are difficult or simply impossible for deepfake algorithms to reproduce accurately. These aspects include the natural blink pattern of the eyes, facial expressions, skin texture, and even the way light reflects off a person’s skin[6].

AI doesn’t stand still, it is becoming increasingly adept at creating deepfakes. Therefore, the simultaneous development of AI-based detection technologies is vital to detect and mitigate the spread of various forms of misinformation. To counter the rise of deepfakes, Intel introduced its software called FakeCatcher[7]. Using advanced artificial intelligence, FakeCatcher operates in real-time and accurately identifies fake videos. This is a critical advantage in a

[5] https://www.forbes.com/sites/cognitiveworld/2019/01/07/the-dual-use-dilemma-of-artificial-intelligence/?sh=701382cf6cf0

[6] https://www.cryptopolitan.com/the-role-of-artificial-intelligence-on-fraud/

[7] https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html

world where misinformation can go viral in seconds. Using a photoplethysmography technique called PPG, which examines the change in color of blood as it circulates through the body, infrared light measures the volumetric fluctuations in blood circulation. PPG signals are captured from the subject's face and converted into PPG maps, and a deep learning approach is used to classify whether a video is fake or not. Eye gaze recognition is also used in FakeCatcher to improve the accuracy of the tool.

Strategies to Counteract

Consequently, there are currently 2 prevailing ideas on how deepfakes should be addressed.

The first approach is to develop effective detection tools, which are essential in swiftly identifying and curtailing the spread of fake content, thus maintaining the integrity of information shared online. However, the Brookings Institution highlights that deepfakes may become indistinguishable from real content in the near future, making automated deepfake detection increasingly difficult.

The second approach comprises the main idea—it is necessary to make people as much as possible aware of the deepfake technology. Virtual literacy courses are a prime example of this type of initiative. For example, the Massachusetts Institute of Technology (MIT) has recently launched a course aimed at addressing misinformation, particularly focusing on deepfakes[8].

Findings

The development of artificial intelligence (AI) and deepfake technology in the context of modern society is having a more and more tangible impact on humanity in the context of the information field every day. The development of

[8] https://news.mit.edu/2022/fostering-media-literacy-age-deepfakes-0217

AI, especially in the last three years, has significantly changed our lives, offering society both achievements and challenges caused by these technologies.

Deepfake technology, capable of creating very realistic audiovisual content, is a serious threat in the digital age. It is capable of manipulating public opinion, influencing the political landscape and spreading misinformation. Real incidents demonstrate the destabilising power of this technology, highlighting the need for effective countermeasures. Therefore, the impact of AI on society will largely depend on how humanity uses and regulates these technologies.

1. Stephen Hawking: AI Could Be 'Worst Event in Civilization' // CNBC URL: https://www.cnbc.com/2017/11/06/stephen-hawking-ai-could-be-worst-event-in-civilization.html(accessed: 21/12/2023).
2. Добробаба М. Б. Дипфейки как угроза правам человека // Lex russica. — 2022. — Т. 75. — № 11. — С. 112–119. — DOI: 10.17803/1729-5920.2022.192.11.112-119.
3. Дипфейки как угроза правам человека // Cyberleninka URL:https://cyberleninka.ru/article/n/dipfeyki-kak-ugroza-pravam-cheloveka (accessed: 21/12/2023).
4. Failed Coup in Gabon: The Deepfake Theory // IFREI URL: https://www.ifrei.org/article50-2019-Failed-Coup-in-Gabon-The-Deepfake-Theory (accessed: 22/12/2023).
5. Deepfake Videos: The Future Uncertainty // Politico URL: https://www.politico.eu/article/deepfake-videos-the-future-uncertainty/ (accessed: 22/12/2023).
6. The Dual-Use Dilemma of Artificial Intelligence // Forbes URL: https://www.forbes.com/sites/cognitiveworld/2019/01/07/the-dual-use-dilemma-of-artificial-intelligence/?sh=701382cf6cf0 (accessed: 22/12/2023).
7. The Role of Artificial Intelligence on Fraud // Cryptopolitan URL: https://www.cryptopolitan.com/the-role-of-artificial-intelligence-on-fraud/ (accessed: 23/12/2023).
8. Intel Introduces Real-Time Deepfake Detector // Intel Newsroom URL: https://www.intel.com/content/www/us/en/newsroom/news/intel-introduces-real-time-deepfake-detector.html (accessed: 23/12/2023).

9. Fostering Media Literacy in the Age of Deepfakes // MIT News URL: https://news.mit.edu/2022/fostering-media-literacy-age-deepfakes-0217 (accessed: 23/12/2023).