IMPACT OF ARTIFICIAL INTELLIGENCE IN DISINFORMATION: THE CASE OF DEEPFAKE [STIC-B415]

Author : *Lamblin Fabrice Ngueyap Youbissi*
student ID : *000555424*

Université Libre de Bruxelles

Supervised by: Sebastien De Valeriola

January 6, 2024

Abstract

my abstract will be here

1 Introduction

Artificial intelligence (AI) is growing very fast, and has had a significant impact on various aspects of our society, offering unprecedented opportunities while raising significant issues. One of this issue is misinformation. It has become one of the major problems of the modern world, given that to understand the world in which we find ourselves, and to make decisions in most cases, we need to be informed. We need data on our environment. The spread of misinformation has been revolutionized by AI's ability to create synthetic content, thanks in particular to deepfake technology. What's even more worrying is that *Social network algorithms, far from being neutral, are not designed to sort the true from the false, but to select, classify, prioritize and target information likely to capture the attention of a maximum number of users[1].* This means that social networks frequently focus on user retention and advertising revenue generation. And as a result, algorithms are designed to favor content that provokes reactions, whether positive or negative. This dynamic creates an information bubble where users are often exposed to content that reinforces their current preferences, without necessarily evaluating the veracity of the information presented. As AI becomes more and more established in various sectors, disinformation finds new ways to spread. It is therefore necessary to pay particular attention, from understanding the mechanisms of AI to the very nature of misinformation and the theoretical underpinnings of this complex issue. The first objective of this article will be to establish the theoretical foundations of our analysis, define artificial intelligence clearly and examine its applications. In addition, we will explore the complex mechanisms of misinformation, which will help us to better understand the role of AI in this context, particularly through deepfake. The second section will examine the real effects of AI, highlighting the speed with which misinformation spreads, its destabilizing effects on public trust and its far-reaching consequences for institutions and society as a whole. The fourth section will identify the challenges and issues inherent in this complex reality, focusing on detecting and combating deepfakes, as well as the crucial issues of responsibility, ethics and regulation. The fifth section will present solutions and perspectives for reducing the negative effects of AI on misinformation. Overcoming this global challenge will require the development of innovative detection technologies, public awareness initiatives and working together on an international scale. Finaly In part six, we will illustrate our analysis with concrete case studies, examining an example of deepfakes used for disinformation and analyzing their effects on society and individuals.

2 Some theoretical basis

2.1 What's deepfake and use?

The image or video synthesis technique known as "deepfake" is based on artificial intelligence, more specifically on deep neural networks. *The term "deepfake" emerged in late 2017 when a redditor (someone who uses reddit platform which is an American social media platform for web content rating by votes along with discussion of websites) posted realistic pornographic videos featuring Hollywood actresses who weren't really part of it [2]* .Deep fake requires huge data to train models or neural networks to create fake image, video,audios. According to IBM [3] *Neural networks attempt to mimic the human brain, combining computer science and statistics to solve common problems in artificial intelligence [4].* A neural network is made up of several layers of neurons, organized into an input layer (which receives the initial data to be processed), one or more hidden layers (which perform complex transformations of the data). Each neuron in these layers applies an activation function to the data received, and an output layer (the final layer produces the results of the network's analysis or prediction). In these layers, each neuron is connected to others, and each connection has an associated weight and threshold.

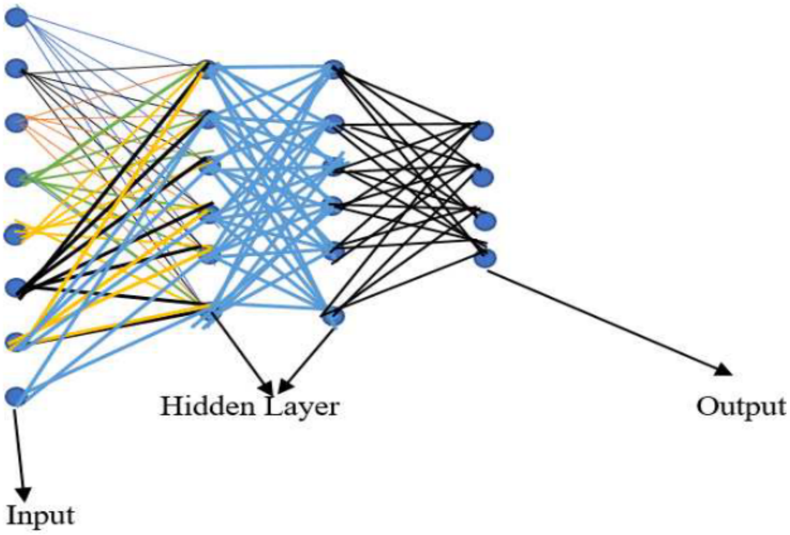

Figure 1: Multi layered neuron network [2]

Each artificial neuron receives data, performs a mathematical operation on it using specific weights and thresholds, and then, if the data exceeds a certain threshold, the result is passed on to the next layer. Alternatively, no information is sent at all. Training data helps neural networks adjust their weights and thresholds, improving their ability to perform specific tasks over time. It's an iterative learning process that aims to reduce the number of errors between the network's predictions and expected results. It is therefore important to say that,*If problem for processing is face generation its more complex because the network has to reads in input and then extract features like eyes, nose, mouth, texture, facial features, determine contort of* *such features and much more, neural network to generate face should have high processing and a large data and a lot of time as well as complex neural network to understand and train in face recognition [2].*

2.1.1 How does deepFake works ?

Deepfakes work through the use of deep neural networks, in particular adversarial generative networks (AGNs). *A generative adversarial network (GAN) is a machine learning (ML) model in which two neural networks compete with each other by using deep learning methods to become more accurate in their predictions. GANs typically run unsupervised and use a cooperative zero-sum game framework to learn, where one person's gain equals another person's loss. The two neural networks that make up a GAN are referred to as the generator and the discriminator. The generator is a convolutional neural network and the discriminator is a deconvolutional neural network. The goal of the generator is to artificially manufacture outputs that could easily be mistaken for real data. The goal of the discriminator is to identify which of the outputs it receives have been artificially created [5].* The process begins with the collection of massive datasets, typically videos and images of the subject to be reproduced. Two main elements make up the GAN architecture: the generator, which transforms the original subject data to make it similar to the target subject, and the discriminator, which distinguishes between real and generated data. Both elements are trained iteratively at the same time, resulting in a dynamic competition in which the generator constantly improves to deceive the discriminator, and the latter modifies its discernment capabilities. Once the model has been sufficiently trained, the generator is used to create synthetic data, such as an animated face, which can be integrated into an existing video. Post-processing procedures can then be used to enhance the visual quality of the deepfake.

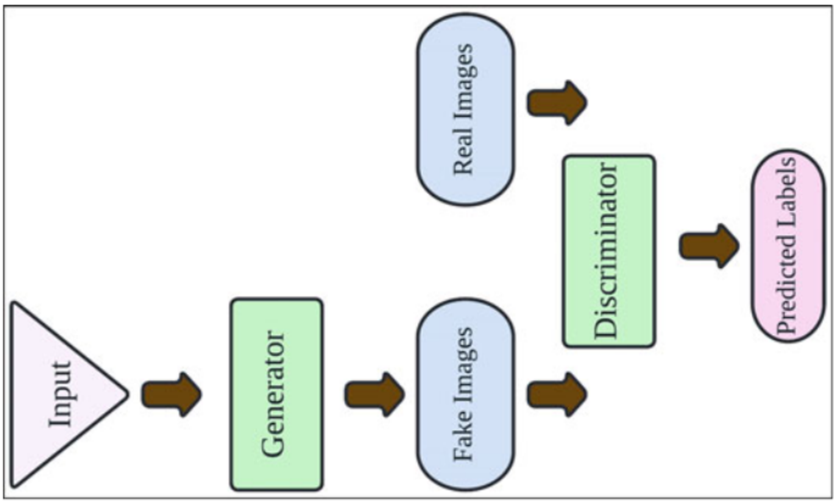

Figure 2: Architecture of GANs [6]

Deepfakes have given rise to concern because of their potential for misuse, particularly in disinformation and the manipulation of public opinion. *Deep fake can come in use for helping people who have lost their speech to give them new improved voice, commercially deepfake can be used in improving animation or movie quality putting in creative imagination to work as well is therapeutic to people who have lost their dear once. Negative aspects of deep fake include creating fake images, videos, audios that look very real can cause threats to an individual's privacy, organizations, democracy, and even national security.[2]*

2.2 Misinformation

Misinformation has become one of the most dangerous threats of our time. The rapid spread of false information can have serious consequences for society, because information is fast-moving and connectivity is global. It's not a recent phenomenon, but with the Internet and social media, its scope and scale have increased dramatically. So then, malicious people use these platforms to spread false information, frequently with the aim of manipulating public opinion, creating confusion or even social conflict. *Although people use many cognitive heuristics to make judgments about the veracity of a claim (for example, perceived source credibility), one particularly prominent finding that helps explain why people are susceptible to misinformation is known as the 'illusory truth' effect: repeated claims are more likely be judged as true than non-repeated (or novel) claims. Given the fact that many falsehoods are often repeated by popular media,politicians, and social-media influencers, the relevance of illusory truth has increased substantially [7].* The mechanisms of disinformation are a complex combination of influences, distortions and manipulations designed to distort the perception of reality. And like we are going to see in the next section, Artificial intelligence through deep Fakes technologies is a weapon of mass destruction in the disinformation process, because the faking models generated by these technologies are becoming more and more realistic and more and more targeted. *as more and more nation-state actors turn to social media manipulation and information warfare to assist in their geopolitical ambitions, deep fakes will become a more serious threat. It is impossible to predict when the deep fake problem will come to a head on the international stage, when a malicious actor will release a deep fake in the hopes of altering an election or spurring a war, but that future is not far off. [8]*

3 DeepFake and misinformation

So far in our article, we've taken a look at artificial intelligence and explored the notion of deepfake, which uses deep learning techniques to generate models that tend to substitute for reality. Aside from all the good this technology can do, particularly in the film industry to create certain comic tricks, deepfake, in the wrong hands, serves more as a disinformation engine than anything else. In fact, to be informed in the past, before the explosion of the Internet, people had to go through the traditional media of the time, such as paper newspapers, radio and television. The institutions in charge of these media had a firm grip on them, which meant that the information published was verified and the sources were reliable. Today, however, *there's a global pandemic, thanks to the information circulating on television, your tablet or your phone. The Internet makes information available anywhere, at any time, in just two clicks [9].* With the Internet, people have access to a huge amount of information from a variety of sources. And this abundance of information makes it extremely difficult to distinguish the true from the false. *Information is a decision-making tool, or even a means of influencing reality [10].* If information is what allows an individual to make a decision, and with the advent of the internet and the impossibility of controlling what individuals publish as information on their networks, also considering that a significant portion of the global population no longer has the habit of obtaining information from reputable media dedicated to this purpose, then deepfakes, in themselves, contribute more to misinformation than anything else. This is because anyone can have access to create or alter existing information and publish it.

DeepFakes are often used to create falsified speeches by public figures, as we'll see in a case study. It can also generate fake videos of historical events...etc. And one of the immediate consequences of this is that individuals may be reluctant to believe when confronted with real information, which slows down their decision-making considerably.

4 Possible Solutions

If any user is a victim of misinformation, it can be assumed that he has exposed himself to an unofficially accredited source of information. He has consumed information that has not been verified, and this has probably induced him to make erroneous decisions as a result of this false information. But it's very important to note that,*under most circumstances, fake news come to users when they are in a relaxed mode for information consumption instead of a critical-thinking mode for work or for study. [11].*This means that many people are just vulnerable to information, and don't approach what they perceive with a critical mind. Criticism means questioning the source, making a logical analysis of what you perceive as information, both visually and in terms of content. This situation reveals an alarming reality about how our psychology can influence the reception of information. If a person is in a relaxed state, such as eating or drinking, he or she is more vulnerable to passively absorbing information without questioning its veracity. It therefore becomes clear that one of the first methods of combating disinformation in an environment polluted by deepFakes would be to have everyone think critically when confronted with information.

on the other hand, The ethical aspect of AI should be taken into consideration. *AI continues demonstrating its positive impact on society while sometimes with ethically questionable consequences. Not doing AI responsibly is starting to have devastating effect on humanity, not only on data protection, privacy, and bias but also on labor rights and climate justice. Building and maintaining public trust in AI has been identified as the key to successful and sustainable innovation . Thus, the issue of ethical AI or responsible AI has gathered high-level attention. Nearly one hundred principles and guidelines for ethical AI have been issued by private companies, research institutions, and public organizations and some consensus around high-level principles has emerged.[12]*

5 Case studies

blablalalbalalbaba

6 Conclusion

I should my conclusion here

References

[1] “Internet, l’autoroute de la désinformation ?.” https://lejournal.cnrs.fr/articles/internet-lautoroute-de-la-desinformation.

[2] S. Negi, M. Jayachandran, and S. Upadhyay, “Deep fake : An Understanding of Fake Images and Videos,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 183–189, May 2021.

[3] “IBM,” *Wikipédia*, Nov. 2023.

[4] “Qu’est-ce qu’un réseau de neurones ? — IBM.” https://www.ibm.com/fr-fr/topics/neural-networks.

[5] “What is a Generative Adversarial Network (GAN)? — Definition from TechTarget.” https://www.techtarget.com/searchenterpriseai/definition/generative-adversarial-network-GAN.

[6] I. R. Khan, S. Aisha, D. Kumar, and T. Mufti, “A Systematic Review on Deepfake Technology,” in *Proceedings of Data Analytics and Management* (A. Khanna, Z. Polkowski, and O. Castillo, eds.), Lecture Notes in Networks and Systems, (Singapore), pp. 669–685, Springer Nature, 2023.

[7] S. van der Linden, “Misinformation: Susceptibility, spread, and interventions to immunize the public,” *Nature Medicine*, vol. 28, pp. 460–467, Mar. 2022.

[8] jlbeyer, “Deep Fakes, Fake News, and What Comes Next,” Mar. 2019.

[9] Laetitia, “Comment s’informait-on avant internet ?.” https://curiokids.net/comment-sinformait-on-avant-internet-lespace-arthur-masson-texplique-tout/, May 2021.

[10] I. Boydens, “L’océan des données et le canal des normes:,” *Annales des Mines - Responsabilité et environnement*, vol. N° 67, pp. 22–29, July 2012.

[11] H. Kanoh, “Why do people believe in fake news over the Internet? An understanding from the perspective of existence of the habit of eating and drinking,” *Procedia Computer Science*, vol. 126, pp. 1704–1709, Jan. 2018.

[12] L. Zhu, X. Xu, Q. Lu, G. Governatori, and J. Whittle, “AI and Ethics—Operationalizing Responsible AI,” in *Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership* (F. Chen and J. Zhou, eds.), pp. 15–33, Cham: Springer International Publishing, 2022.