Model United Nations of Bucharest
2023

The United Nations Commission on Science and Technology for Development

Regulating the Deepfake Technology

-Research Report-

Anjo Babel

-Chairperson-

Tudor Papuc

-Chairperson-

Theodor Micula

-Chairperson-

Table of contents

1. Introduction

Over the years, technology has become one of the most popular industries, and has evolved in ways that no human being could have ever imagined. As a result of this great and fast expansion, scientists began to work even more in order to achieve higher levels. Nowadays, experts have achieved the creation of advanced technologies that could both aid and harm our planet: Artificial Intelligence (AI). Using this highly capable invention, scientists have been discovering new branches of AI, including “deepfake”. The term deepfake refers to a deep learning algorithm that is able to perceive digital content, make changes to it, and even create such from scratch. These transmutations are made so well that a human cannot easily detect whether a particular type of content is real or not.

As a consequence of worldwide access to technology and the internet, people have managed to acquire this sort of technology and have started to use it, both for good and destructive purposes. For example, there are a number of cases in which anonymous internet users have created and published digital content -created using deepfake technology - involving different popular celebrities and even politicians. Because of this, it is important to acknowledge that these pieces of media have the potential of creating chaos and manipulating public perception, both on an individual and global scale.

2. Key Terms

AI- Artificial intelligence (AI), also known as machine intelligence, is a branch of computer science that focuses on building and managing technology that can learn to autonomously make decisions and carry out actions on behalf of a human being;

GAN- Generative Adversarial Networks, or GANs for short, are an approach to generative modelling using deep learning methods, such as convolutional neural networks.;

WIPO- The World Intellectual Property Organization (WIPO) is the global forum for intellectual property (IP) services, policy, information, and cooperation.

CBRN - Acronym for chemical, biological, radiological, and nuclear.

3. History

The manipulation of photos has been an issue since the 19th century. During that time, photographers used mirrors to depict different viewing angles or paint in order to obtain various effects. In most cases, this technique was applied to create even more impressive and more detailed pictures, but, especially during World War I, at the beginning of the 20th century, it served the purpose of propaganda and a depiction of a "fake world".

Nevertheless, as technology evolved and moving pictures were developed, manipulation of videos and movies also emerged. At the end of the 20th century, the first pieces of film editing software had been published to ease the process of editing video content. Combined with the multiple possibilities offered by the internet and the media, it became very easy to use this technological progress with the purpose of spreading fake news.

Although prototypes of deepfake software already existed for some time, the developer and computer scientist Ian Goodfellow created two separate AIs to artificially generate realistic pictures of high accuracy in 2014 - this represented a key moment regarding the evolution of deepfake technology. Moreover, several releases of deepfake-technology-made digital content involving public figures drew attention to the threat that deepfakes may pose to the maintenance of global stability.

Nowadays, not only pictures and movies can be modified, but it is also possible to copy the whole voices of people and apply deep fakes in real-time, as there is a number of deepfake-related apps that can be easily accessed by any internet user.

4. Key Issues

At the moment, the issue regarding deepfake technology is well-known among most nations of the world and, although the development of this branch of AI may have its advantages, its drawbacks can have catastrophic consequences. In most cases, the problem that is represented by deepfake has a negative impact on civilians: there have been multiple illegal publications of digital content involving public figures on various social media platforms, apps that register millions of users every day.

Because of this, many public figures have been wrongly accused of committing offenses, and, subsequently, were legally sanctioned. Seeing as how many people have been troubled by the same factor, the problem of deepfake has become an international manner.

Moreover, deepfake can also have an impact at the economic level. For example, in 2019 the CEO of a U.K.-based energy firm received a phone call from the leader of the firm's German parent company, in which the transfer of €220,000 to a supplier in Hungary was requested; however, it was later discovered that the amount of money had been moved to unknown accounts. In an interview, the CEO stated that he recognized the "slight German accent and the melody" of his supervisor's voice, but, as it was later discovered, the phone call was made using deepfake technology.

These technological means have also been used for political purposes: such an example may be provided by an event that occurred in March 2022, during the first part of the Russian invasion of Ukraine, when a fake video of Ukrainian President Volodymyr Zelenskyy emerged among the news that circulated in Ukrainian media. In the clip, the Ukrainian soldiers are being told to" lay down their weapons", and, as a response to the fake video, Volodymyr Zelenskyy posted a video to his Telegram Channel saying that "We are defending our land, our children, our families. So, we don't plan to lay down any arms. Until our victory.".

Despite all of this, deepfake is yet to reach its full destructive potential, and, therefore, there is still enough time to implement measures that aim at finding a solution to this issue.

5. Major Parties Involved

- *People's Republic of China*

China is one of the most active combaters of deep fake technology, but, at the same time, it is a country that records a significant number of reports regarding deepfake-related activities. In the past five years, the government has passed and implemented numerous legislations which have the purpose of combating deep fake technologies on a national level: such an example is represented by the "Provisions on the Administration of Deep Synthesis of Internet Information services" law, passed in January 2023. These laws have been implemented with a satisfactory degree of success due to the control the government has over its cybernetic space, and while this degree of control may be questioned for ethical reasons, it has proven to be effective in the fight against deepfake technologies in China.

- *Ukraine*

Since the Russian invasion of Ukraine started in February 2022, Ukraine has been overwhelmed with misinformation in the context of the war, deepfakes being no exception. One of the most well-known instances of deepfake usage in the war is the one mentioned in the Key Issues section; however, it is not the only instance of deepfake technology being used currently. The Ukrainian approach to solving the issue of deepfakes focuses on individually identifying deepfakes and removing them. Nonetheless, this strategy can be successful on a smaller scale level. Last but not least, the Ukrainian government has also invested considerable resources in cyber security, a decision that has helped Ukraine in handling the effects of information warfare, and, subsequently, in combatting deepfake technology.

- *Russian Federation*

Russia has been associated on numerous occasions with deepfake attacks by the international community. In the context of the Russian Invasion of Ukraine, many cyber-attacks that Ukraine faced were launched from Russian-controlled territories; nevertheless, the Russian Federation has also faced this sort of attack during this period. To be more specific, a deepfake video of Russian president Vladimir Putin announcing that he had negotiated a peace treaty with Ukraine was released in the Russian press. Similarly to China, in Russia, the government plays an important part in what concerns the surveillance of the activity that takes place in the cyber-space; however, no legislation that explicitly refers to deepfakes has been introduced nor implemented.

- *United States of America*

The US is another main combatant in the international effort to mitigate the effects of developing deep fake technology, a great example of this being the H.R.3230 – DEEP FAKES Accountability Act, introduced during the 116th congress, and the S.2559 – Deep fake Task Force Act, introduced during the 117th congress. As it can be seen from the previously mentioned acts, in the past few years the US has recognized the threat posed by deep fake technology, devised and implemented a legal framework, and also created a task force in order to enforce the previously mentioned legal framework. The US model is proven to be effective, yet as deep fake technologies improve, the method to mitigate the effects of such technologies must improve as well.

- European Union

The European Union and its member states are one of the most successful institutions in the fight against rising deep fake technologies, on both national and international levels. This process started with an extremely detailed study followed by debates which included European countries, and had as a conclusion the creation of a Code of Practice, aimed at regulating the activity of platforms and companies that might help or promote deep fake technology. The code states that "companies can be fined up to 6% of their global revenue if found guilty of non-compliance". On a national level, laws such as the German Network Enforcement Act and the French Anti Fake-News law (which was revised and currently has a special article for deepfakes) have been successful in combatting deepfake technologies; thus, the collaboration between member states will lead to the improvement of the discussed problem (the Netherlands, Belgium and Sweden are already involved in projects regarding the elaboration and adoption of new laws of this type).

6. Timeline

2014: Ian Goodfellow and his colleagues developed the first "generative adversarial network (GAN)" which is a class of machine learning. By using two separate ais to create a realistic picture of somebody, one to create the image, and one to compare it to real pictures and to verify proper versions, deepfakes became possible.

2017: The University of Washington made a deepfake video including former US President Barack Obama, a video that was broadcasted by several digital media platforms. Some were impressed and amazed by the capabilities of AI technology, while others were worried about the fact that determining the differences between reality and artificial reality has become almost impossible. In the following years, many other public figures became targets of deepfake-related activities.

2021: The EU enacted a policy regulating deepfakes, making it mandatory to label deepfakes as such. Meanwhile in the USA, only three states limited the use of deepfake technology, due to a lack of awareness. China banned it completely.

Even if there are no specific rules in some countries regarding this, deepfakes often violate other laws, such as personal rights, infringe property rights, or are categorized as causing misinformation and are, therefore, illegal.

2021: the UNIDIR (United Nations Institute for Disarmament Research) also met and discussed aspects that are strongly linked to the theme of "Risks and Threats of Deepfakes, Trust and International Security". The UN never passed a resolution regarding deepfake technology, although the institution is aware of this problem as stated in several reports on this matter.

Consequentially, many tech companies banned deepfakes completely. The platform Reddit was the first to ban deepfakes in January 2020. Twitter soon made it mandatory to label deepfakes in February 2020, and the former Facebook - today, Meta – followed, with a complete ban in 2021.

7. Evaluation of Previous Attempts

Deepfake technology development is a relatively recent problem; thus, one might assume that not many efforts have been made towards solving this issue. However, many actions have been taken at national levels in countries such as the United States or the People's Republic of China, as stated in the Major Parties Involved section, with various degrees of success. Decisions have already been implemented, yet as successful measures in combating deepfake technologies are discovered, the previously-mentioned technologies also continue to; therefore, it is clear that an internationally-organized effort is needed in order to partially - if not fully - resolve the issue of developing deep fake technologies. The UN has recognized this and, despite the fact that no actual resolutions passed on the topic of deepfake technologies development, measures are being taken. Concretely, since the UNRIC A/75/982 report of the secretary general, which states that "The Internet has altered our societies as profoundly as the printing press did" and that "Societies should be encouraged to develop a common, empirically backed consensus on the public good of facts, science and knowledge […] and a global code of conduct that promotes integrity in public information could be explored", actions have been taken to raise awareness about fake news and the fast development of deepfake technologies, actions such as the World Health Organization communicate from the 10th of February 2023 and the UNICRI (the United Nations Interregional Crime and Justice Research Institute) Handbook to combat CBRN disinformation. Raising awareness is a good starting point, but in order to solve the issues, viable solutions must be found and implemented as soon as possible.

8. Possible Solutions

First of all, it should be mentioned that deepfakes are mostly spread on the internet, a globally-accessible media platform. That is why national countermeasures and regulations are not always effective, and, thus, why international cooperation would help in increasing the rate of success concerning the handling of deepfake technology consequences.

1. Raising awareness of deepfakes for those potentially affected is a preventive measure that could increase the detection rate of deepfakes as well as reduce the number of victims falling for deep-faked fraud. This sort of action could take place as early as in school. Official government websites, organizations such as the UN, influencers, or radio and tv shows could raise awareness and provide low-threshold information.

2. Some countries, as well as the European Union and many tech companies, made labelling deepfakes compulsory. This might be seen as a very lenient measure, since there is no restriction for producing or publishing deepfake content. Nonetheless, this labelling must be clear and understandable in order to avoid any other further unpleasant consequences.

Besides that, it remains difficult to check videos of deepfake effects, since programs for creating deepfakes are widely available and easy to use. Consequently, any identified acts of violation of current legislation must be sanctioned accordingly.

3. One of the most controversial measures would be a complete ban on deepfakes. By doing this, it is expected that number of deepfakes would rapidly decrease, and the internet would be safer. Recently developed AI software would help in scanning the internet for deepfakes and detecting it. Hence, punishments for incompliance with the law would be much more severe. However, this decision would also have a number of negative consequences, as a complete banning may involve the violation of certain principles, such as the ones that are embodied by the idea of freedom of speech.

There are various actions that could be taken to regulate the use of deepfake technology, but it remains to be seen which solution will prove to be optimal for solving this issue.

9. Bibliography

Johnson, Dave. “What Is a Deepfake? Everything You Need to Know about the AI-Powered Fake Media.” Business Insider, 10 Aug. 2022, www.businessinsider.com/guides/tech/what-is-deepfake

Schroer, Alyssa. “What Is Artificial Intelligence? How Does AI Work? | Built In.” Builtin, 2017, https://builtin.com/artificial-intelligence

Digital Guide IONOS, 2020. *Deepfakes: Fälschungen der nächsten Generation.* [Online]

Available at: https://www.ionos.de/digitalguide/online-marketing/social-media/deepfakes/

[10 02 2023].

eyerys, 2017. *A Reddit User Starts 'Deepfake'.* [Online]

Available at: https://www.eyerys.com/articles/timeline/reddit-user-starts-deepfake#event-a-href-articles-timeline-internet-captivated-when-netizens-realized-older-woman-who-took-prince-harry-virginitythe-internet-captivated-when-netizens-realized-039-the-older-woman039

[10 02 2023].

Organization for Social Media Safety, 2023. *Deepfake Technology.* [Online]

Available at: https://www.socialmediasafety.org/advocacy/deepfake-technology/#:~:text=Deepfake%20technology%20first%20appeared%20in,to%20create%20realistic%20fake%20videos.

[10 02 2023].

Slater, S., 2021. *History of Photo Manipulation.* [Online]

Available at: https://clippingthephotos.com/history-of-photo-manipulation/#:~:text=The%20birth%20of%20%E2%80%9Cphotography%20manipulation,mirrors%20to%20create%20different%20perspectives.

[10 02 2023].

Zeitchik, S., 2022. *Ready or not, mass video deepfakes are coming.* [Online]

Available at: https://www.washingtonpost.com/technology/2022/08/30/deep-fake-video-on-agt/

[10 02 2023].

Allyn, Bobby. "Deepfake Video of Zelenskyy Could Be "Tip of the Iceberg" in Info War, Experts Warn." NPR, 16 Mar. 2022, www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia.

Çolak, Betül. "Legal Issues of Deepfakes." Www.internetjustsociety.org, 8 Feb. 2021, www.internetjustsociety.org/legal-issues-of-deepfakes.

"Deepfakes Are on the Rise — How Should Government Respond?" Www.govtech.com, www.govtech.com/policy/Deepfakes-Are-on-the-Rise-How-Should-Government-Respond.html.

Somers, Meredith. "Deepfakes, Explained." MIT Sloan, 21 July 2020, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

Baig, Rachel. "Fact Check: The Deepfakes in the Disinformation War between Russia and Ukraine | DW | 18.03.2022." DW.COM, 18 Mar. 2022, www.dw.com/en/fact-check-the-deepfakes-in-the-disinformation-war-between-russia-and-ukraine/a-61166433.

Roth, Andrew. "European MPs Targeted by Deepfake Video Calls Imitating Russian Opposition." The Guardian, 22 Apr. 2021, www.theguardian.com/world/2021/apr/22/european-mps-targeted-by-deepfake-video-calls-imitating-russian-opposition.

Wakefield, Jane. "Deepfake Presidents Used in Russia-Ukraine War." BBC News, 18 Mar. 2022, www.bbc.com/news/technology-60780142.

BBC News, 2017. Fake Obama created using AI video tool - BBC News. [Online]

Available at: https://www.youtube.com/watch?v=AmUC4m6w1wo

[11 02 2023].

Bianco, A. A. a. B., 2021. The 2021 Innovations Dialogue Conference Report: Deepfakes. [Online]

Available at: https://unidir.org/sites/default/files/2021-12/UNIDIR_2021_Innovations_Dialogue.pdf

[11 02 2023].

Çolak, B. B., 2021. DISINFORMATION Legal Issues of Deepfakes. [Online]

Available at: https://www.internetjustsociety.org/legal-issues-of-deepfakes

[11 02 2023].

European Parliamentary Research Service, 2021. Tackling deepfakes in European policy. [Online]

Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf

[11 02 2023].

Federal Office for Information Security, 2023. Deep Fakes – Threats and Countermeasures. [Online]

Available at: https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html

[11 02 2023].

Hsu, T., 2023. As Deepfakes Flourish, Countries Struggle With Response. [Online]

Available at: https://www.nytimes.com/2023/01/22/business/media/deepfake-regulation-difficulty.html#:~:text=Meta%2C%20TikTok%2C%20YouTube%20and%20Reddit,are%20intended%20to%20be%20misleading.

[11 02 2023].

Organization for Social Media Safety, 2023. Deepfake Technology. [Online]

Available at: https://www.socialmediasafety.org/advocacy/deepfake-technology/#:~:text=Deepfake%20technology%20first%20appeared%20in,to%20create%20realistic%20fake%20videos.

[10 02 2023].

q5id, 2022. Are Deepfakes Illegal? Here are the Implications. [Online]

Available at: https://q5id.com/blog/are-deepfakes-illegal-here-are-the-implications

[11 02 2023].

Springer, 2022. Deepfakes generation and detection. [Online]

Available at: https://www.researchgate.net/figure/Timeline-of-the-evolution-of-Deepfakes_fig3_361086563#:~:text=Deepfakes%20are%20created%20using%20the,important%20to%20review%20detection%20methods.

[11 02 2023].

Chee, Foo Yun. "Exclusive: Google, Facebook, Twitter to Tackle Deepfakes or Risk EU Fines." Reuters, 14 June 2022, www.reuters.com/technology/google-facebook-twitter-will-have-tackle-deepfakes-or-risk-eu-fines-sources-2022-06-13/.

"IGF 2020 WS #211 Collective Human Rights Approach to Deepfake Applications | Internet Governance Forum." Intgovforum.org, https://intgovforum.org/en/content/igf-2020-ws-211-collective-human-rights-approach-to-deepfake-applications.

"H.R.3230 - 116th Congress (2019-2020): Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019." Congress.gov, 2019, www.congress.gov/bill/116th-congress/house-bill/3230

"UN's Rights Council Adopts "Fake News" Resolution, States Urged to Take Tackle Hate Speech." UN News, 1 Apr. 2022, https://news.un.org/en/story/2022/04/1115412.

"UNRIC Library Backgrounder: Combat Misinformation - Selected Online Resources on Misinformation, Disinformation and Hate Speech." United Nations Western Europe, 8 Feb. 2022, https://unric.org/en/unric-library-backgrounder-combat-misinformation/.

Bocetta, S., 2019. Deepfakes are a problem, what's the solution?. [Online]

Available at: https://cybersecurity.att.com/blogs/security-essentials/deepfakes-are-a-problem-whats-the-solution

[12 02 2023].

Federal Office for Information Security, 2023. Deep Fakes – Threats and Countermeasures. [Online]

Available at: https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html

[11 02 2023].

Fischbach, J., 2020. We Can Solve the Problem of Deepfakes and Disinformation. [Online]

Available at: https://www.protegopress.com/we-can-solve-the-problem-of-deepfakes-and-disinformation/

[12 02 2023].

iproov, 2022. How To Protect Against Deepfakes – Statistics and Solutions. [Online]

Available at: https://www.iproov.com/blog/deepfakes-statistics-solutions-biometric-protection

[12 02 2023].

Mesa-Cucalon, N., 2021. Deepfakes: Effective Solutions for Rapidly Emerging Issues. [Online]

Available at: https://medium.com/analytics-vidhya/deepfakes-effective-solutions-for-rapidly-emerging-issues-8b1685feef56

[12 02 2023].

10. Appendices

I. https://unric.org/en/unric-library-backgrounder-combat-misinformation/

II. https://amp.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them

III. https://lieber.westpoint.edu/deepfake-technology-age-information-warfare

IV. https://www.cnbc.com/amp/2022/12/23/china-is-bringing-in-first-of-its-kind-regulation-on-deepfakes.html

V. https://amp.cnn.com/cnn/2022/06/14/tech/social-media-deepfakes-eu-fines/index.html

VI. https://crsreports.congress.gov/product/pdf/IF/IF11333

VII. https://unidir.org/events/2021-innovations-dialogue

VIII. https://www.theguardian.com/world/2021/apr/22/european-mps-targeted-by-deepfake-video-calls-imitating-russian-opposition