Deepfakes as Digital Propaganda: The Russian Case in the War in Ukraine

Yuliia Bronovytska
Department of Communication Studies
Tzu Chi University, Taiwan

Abstract: In the digital age, the development of AI tools not only simplifies our lives but also poses dangers that we need to address. One such tool, “deepfake”, has become a powerful tool in modern warfare and disinformation campaigns. The Russia-Ukraine war is a vivid example of the use of deepfakes in warfare. This paper explores the mechanisms, reach, and dangers of Russian deepfakes, focusing on how they manipulate society’s perceptions and advance strategic disinformation narratives. By analyzing the case studies of Russian deepfakes, the paper aims to explore the creation techniques and dissemination of fabricated content in the context of the war in Ukraine. It also considers the measures taken by the international community to control the use of artificial intelligence. In the context of the war in Ukraine, deepfakes have emerged as a potentially destructive force of artificial intelligence, making it essential to address their implications for global information integrity and security.

Keywords: Artificial Intelligence, deepfake, disinformation, propaganda, Russia, Ukraine

Introduction

The present times are often referred to in scientific publications as the era of fake news, disinformation, or information chaos (Załoga, 2022). With the development of artificial intelligence technologies, verifying the authenticity of information and trusting media agencies is becoming increasingly difficult. In recent years, the growing availability of deepfake technologies has been a cause for concern. The existence of fake video or audio creates an environment in which the lines between fact and fiction are dangerously blurred. The use of deepfakes becomes especially dangerous in times of war, where information warfare is as important as on the frontline.

The Russian-Ukrainian war is the first example of the use of deepfake technology in wartime to manipulate facts and spread disinformation. Russia, known for its complex disinformation campaigns, is increasingly using deepfakes in its attempts to manipulate public opinion, sow distrust, and destabilize its adversaries. By depicting Ukrainian leaders making false statements and creating videos that discredit the Ukrainian Armed Forces, Russia is trying to undermine Ukraine's credibility and justify its actions in the international arena.

This article explores the use of deepfakes as a form of digital propaganda in the context of the Russian-Ukrainian war. It considers the scientific viewpoint on the potential danger of deepfakes, investigates the responses of countries and tech companies, and analyzes specific cases of Russian deepfakes. By highlighting the role of deepfakes in modern propaganda, this article aims to contribute to a broader understanding of how AI technologies become a new potent tool in information warfare.

Deepfake: The Growing Danger of Synthetic Media

In 2017, a Reddit user first used the term “deepfake” to refer to pornographic material that purportedly depicted the features of well-known women. Deepfakes use neural networks that process vast data to learn how to imitate a person's voice, intonations, mannerisms, and facial expressions (Westerlund, 2019). In other words, deepfakes use facial mapping technology and artificial intelligence to replace a person’s face in a video with another person’s face. Despite the intricacy of the technology involved in producing deepfakes, it is concealed by widely accessible, easy-to-use tools and services like DeepFaceLab (iperov, 2020) or FaceSwap (deepfakes, 2020). Computer users can use these tools with conventional home PCs equipped with gaming graphics cards and without deep technical expertise.

Deepfake technology is causing concern nowadays. The sophistication of deepfakes is rapidly evolving, and, likely, they will eventually be undetectable to the untrained eye (Maras & Alexandrou, 2019). Breen (2021, p. 123) argues that deepfakes “take disinformation to the next level” by “further complicating the ability to decipher true information”. Woolley and Joseff (2020, p. 23) claim that “in the political domain, manipulated and synthesized audio and visual content might be used to affect diplomatic negotiations, incite conflicts, and manipulate elections” while, no less importantly, “in the social domain, they could be employed to exacerbate polarization and demographic divisions or erode trust in institutions, among other outcomes”. Huston and Bahm (2020) pinpoint, “while the harmful use of misinformation has been around for centuries, technology now allows this to happen at a speed and scale never before seen”. Furthermore, deepfakes target social media platforms, where conspiracies, rumors, and misinformation spread easily, as users tend to go with the crowd (Westerlund, 2019). Studies show that people are prone to the echo chamber effect, i.e., they tend to believe information that reflects their beliefs, even if they suspect it is not true (Shen et al., 2019). The increasing number of deepfake videos may exacerbate socio-political divisions in society due to cognitive bias. The Russia-Ukraine war was the first real-world example showing that video created by artificial intelligence could become another form of weapon in war.

As AI-generated disinformation poses a growing danger, governments and large corporations have started to develop some regulatory responses to deepfakes. For example, prominent tech executives have called for a temporary halt to advanced AI development, and Microsoft’s and Google’s CEOs have publicly warned about the threats posed by deepfakes (Birrer & Just, 2024). Platforms have implemented deepfake policies and developed technologies to identify, categorize, or eliminate deepfakes. Additionally, to combat dishonest AI meddling in 2024 elections worldwide, 20 major tech companies signed a unified “Tech Accord” to set expectations for how signatories will manage risks arising from misleading AI electoral content created on their platforms (AI Elections Accord, 2024).

China enacted strict laws in January 2023 that require that content that has been altered must have the subject's consent, include digital signatures or watermarks, and give users a way to "refute rumors" (Kopecky, 2024). In the United States, several bills at the state and federal level target specific types of harmful use of deepfakes that address two issues: election interference and pornography. In addition, the United States aims to ban misleading fake content in political advertising. South Korea recently introduced a similar restriction, facing criticism that it could be used to control elections (Birrer & Just, 2024). In March 2024, the European Union enacted a law on artificial intelligence, the first such regulatory document in the world. The EU has appointed a special body to monitor compliance with the law - the AI Office, as well as three advisory bodies - the European Artificial Intelligence Board, a council of independent experts that can signal risks that it has noticed, and an advisory forum that will include a wide range of stakeholders (AI Act, 2024).

Acknowledging the potential dangers of deepfakes in spreading disinformation will help countries and tech companies develop countermeasures and mitigate the harmful impact on the public. Legal actions can also serve as a deterrence and a foundation for bringing charges against those involved in such behavior.

Fabricated Realities: How Russia Uses Deepfakes in Information Warfare

In Ukraine, the use of artificial intelligence, and in particular, deepfakes, has increased several times since the Russian full-scale invasion on the 24th of February, 2022. Moreover, the fact that the same tactic has already been used in other wars and conflicts suggests that deepfake manipulation will continue to be a new front in all upcoming conflicts (Klepper, 2023).

Russian propagandists are using AI to spread false information allegedly from Ukrainian authorities. One of the famous cases is the deepfake video of Ukrainian president Volodymyr Zelensky, which appeared on Facebook and Telegram in March 2022, in which the Ukrainian President allegedly called on the Ukrainian military to lay down their weapons. The deepfake was also shown by the TV channel Ukraine 24 after the Russian hacking of the news feed (Pearson & Zinets, 2022). The video immediately aroused suspicion among experts and the public, as its quality left much to be desired - Zelensky's face looked unnatural and his movements did not match his actual gestures. The sound was also synthetic, indicating that the video was fake. Nevertheless, this event was described by experts as one of the first cases of a fabricated politician being intentionally used to spread disinformation (Twomey et al., 2023). The main goal of the video was to demoralize the Ukrainian military and citizens.

"By restraining Russian actions, our defenders are leading the Russian leadership to the idea: talk is necessary - address by President of Ukraine Volodymyr Zelenskyy." by President Of Ukraine is licensed under CC0 1.0 Universal

Over time, deepfake technology has improved. New examples emerged when fakes became even more realistic. Russian propagandists began to use these technologies to create better fakes. For instance, on the 7th of November, 2023, the pro-Russian Telegram channel Radio Truha published a deepfake video of then-Ukrainian Commander-in-chief General Valerii Zaluzhnyi. In the video, Zaluzhnyi allegedly calls for a military coup, urging soldiers to march on Kyiv and disobey President Zelensky's "criminal orders," and claiming that the media would remain silent due to their allegiance to Zelensky. The Center for Countering Disinformation refuted this video and emphasized that it was distributed on TikTok, X, and Telegram (ЦЕНТР ПРОТИДІЇ ДЕЗІНФОРМАЦІЇ [@Center for Countering Disinformation], 2023). As of November 23, 2024, the video had received 385,200 views and was shared 11,800 times (ТРУХА⚡[@RadioTruha], 2024). Although Valerii Zaluzhnyi looked very unnatural in the video, the quality of the video demonstrates how deepfakes have evolved since Zelensky's deepfake. Moreover, this video was more successful in creating public confusion because it was much harder to recognize as fake.

"Session with Ukraine - Military Committee in Chiefs of Defence Session" by NATO North Atlantic Treaty Organization is licensed under CC BY-NC-ND 2.0

Another notable example was a fake video with AI-generated audio of Oleksii Danilov, the former secretary of the National Defense and Security Council of Ukraine. The Russian state-controlled TV channel NTV in a special newscast dedicated to the terrorist attack on Crocus City Hall, which took place on the evening of March 22, 2024, in the Moscow region, showed a fake video in which Oleksii Danilov allegedly confirmed Ukraine's involvement in the attack and said that Moscow was "having a lot of fun" in reference to the attack (Robinson, Sardarizadeh & Brown, 2024). Fortunately, the video was debunked as a deepfake before it could be used as a pretext for a military response about which vowed Russian ex-president Dmitry Medvedev. BBC Verify identified that the video was a montage of two interviews with Danilov using an AI-generated voice (Robinson, Sardarizadeh & Brown, 2024).

Discrediting the Ukrainian Armed Forces is also an important part of Russia's disinformation campaign against Ukraine. Russians use the technology of deepfakes to produce fake videos for social media (mostly Telegram) to influence Western audiences to weaken support for Ukraine among its partners. They distribute deepfakes in which Ukrainian soldiers allegedly insult Russian-speaking children, mock their brothers-in-arms, and talk about overthrowing the government. Such fakes are created under the control of The Main Directorate of the General Staff of the Armed Forces of the Russian Federation (Denkovych, 2024). Russia's main tactic in this case is to make video fakes emotional to cause outrage and make people spread lies that can then go viral. One of the vivid examples is about a man with Down's syndrome who was mobilized into the Ukrainian army with the call sign Vokha and is bullied by his brothers-in-arms (Телебачення Торонто [@UkrainianTorontoTelevision], 2024).

Russian deepfakes have become an active part of Russia's information warfare against Ukraine, which is constantly improving. As can be observed from the above, the quality of Russian fake videos has improved since the beginning of the war. Although many identifiable flaws exist in the deepfake videos, the increase in their number and realism indicates the importance of developing media literacy in society. It is worth noting that the media literacy level in Ukraine is not high. According to the survey "Media Literacy Index of Ukrainians 2020-2023 (fourth wave)" conducted by the NGO Detector Media, some Ukrainians still have low media literacy - 3%. Below average level is 21% of Ukrainian citizens, above average - 62%. Only 14% of Ukrainians have a high level of media literacy (Detector Media, 2024). In times of war, this is dangerous, so both the government and NGOs are trying to promote media education in Ukraine. An example of a successful media literacy initiative was the Comprehensive Information and Education Campaign to Counter Disinformation, which ran from June 2023 to August 2024. The Ministry of Culture and Information Policy and the Ministry of Education and Science of Ukraine also offer a large number of training courses and materials (Osvitoria Media, 2024).

Enhancing media literacy provides individuals with essential critical thinking abilities to differentiate between trustworthy information and AI-generated content, thereby preventing misinformation from undermining national security or social unity. Additionally, it enables citizens to effectively access, interpret, and analyze information, which is vital for protecting democratic processes and enabling informed decision-making which is one of the critical factors for society during the war. Media literacy is becoming a key skill in today's world, where manipulation and disinformation spread very quickly.

Conclusion

The disturbing potential of artificial intelligence as a tool for digital propaganda and disinformation is highlighted by the usage of deepfakes during the Russia-Ukraine conflict. As demonstrated by the case studies, deepfakes serve as a powerful instrument to manipulate public opinion, sow confusion, and undermine the credibility of Ukrainian leadership and media both domestically and internationally. The speed at which this technology is evolving emphasizes the importance of promoting media literacy, creating reliable detection systems, and developing international cooperation to counter the spread of deepfake content.

The war in Ukraine underscores how technological advancements like deepfakes are transforming modern warfare and the dynamics of information operations. In response to the growing threat, several countries and tech companies have already taken regulatory steps. The United States and

China have introduced legislation to regulate AI-generated content, requiring transparency and consent, while the European Union's 2024 AI law sets a precedent for global standards. Meanwhile, digital giants like Google and Microsoft collaborate through accords and improve detection systems. It is essential to continue to research the development of deepfakes and further explore the threats posed by these manipulative tools.

References

AI Act. (2024, October 14). Shaping Europe's Digital Future. *European Commission*. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

AI Elections Accord. (2024). A tech accord to combat deceptive use of AI in 2024 elections. https://www.aielectionsaccord.com/

Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*, *0(0)*. https://doi.org/10.1177/14614448241253138

Breen, D. C. (2021). Silent no more: How deepfakes will force courts to reconsider video admission standards. *Journal of High Technology Law, 21(1)*, 122–161. https://bpb-us-e1.wpmucdn.com/sites.suffolk.edu/dist/5/1153/files/2021/01/Breen.pdf

deepfakes. (2020, August 14). GitHub - deepfakes/faceswap: Deepfakes Software For All. [GitHub Repository]. *GitHub*. https://github.com/deepfakes/faceswap

Denkovych, Y. (2024, July 9). *Стали відомі імена працівників ГРУ РФ, які створюють діпфейки для дискредитації ЗСУ – фото [The names of GRU RF employees creating deepfakes to discredit the AFU have become known – photo]*. TCH.ua. https://tsn.ua/ato/stali-vidomi-imena-pracivnikiv-gru-rf-yaki-stvoryuyut-dipfeyki-dlya-diskreditaciyi-zsu-foto-2617527.html

Detector Media. (2024). *Індекс медіаграмотності українців 2020-2023 (четверта хвиля): Аналітичний звіт за результатами комплексного дослідження [Media Literacy Index of Ukrainians 2020-2023 (fourth wave): Analytical report based on the results of a comprehensive study]*. https://detector.media/doc/images/news/archive/2021/225738/zvit-novyy.pdf

Huston, R. P. & Bahm, M. E. (2020). Deepfakes 2.0: The new era of "truth decay". *Just Security*. https://www.justsecurity.org/69677/deepfakes-2-0-the-new-era-of-truth-decay

iperov. (2020, April 9). GitHub - iperov/DeepFaceLab: DeepFaceLab is the leading software for creating deepfakes. [GitHub Repository]. *GitHub*. https://github.com/iperov/DeepFaceLab?tab=readme-ov-file

Klepper, D. (2023, November 28). Deepfakes from Gaza war increase fears about AI's power to mislead. *AP News*. https://apnews.com/article/artificial-intelligence-hamas-israel-misinformation-ai-gaza-a1bb303b637ffbbb9cbc3aa1e000db47

Kopecky, S. (2024). Challenges of Deepfakes. In: Arai, K. (eds) *Intelligent Computing. SAI 2024. Lecture Notes in Networks and Systems, vol 1016 (pp 158–166). Springer, Cham.* https://doi.org/10.1007/978-3-031-62281-6_11

Maras, M.-H. & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof, 23(3), 255-262.* https://doi.org/10.1177/1365712718807226

Osvitoria Media. (2024). *Медіаграмотність як базова навичка: де та як підвищувати свій рівень? [Media literacy as a basic skill: where and how to improve your level?]. Osvitoria.org.* https://osvitoria.media/opinions/mediagramotnist-yak-bazova-navychka-de-ta-yak-pidvyshhuvaty-svij-riven/

Pearson, J. & Zinets, N. (2022, March 16). Deepfake footage purports to show Ukrainian president capitulating. *Reuters.* https://www.reuters.com/world/europe/deepfake-footage-purports-show-ukrainian-president-capitulating-2022-03-16/

Robinson, O., Sardarizadeh, S. & Brown, P. (2024, March 26). Moscow attack: Debunking the false claims. *BBC News.* https://www.bbc.com/news/world-europe-68657383

Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y. & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. *New Media & Society, 21(2)*, 438-463. https://doi.org/10.1177/1461444818799526

Twomey, J., Ching, D., Aylett, M. P., Quayle, M., Linehan, C. & Murphy, G. (2023). Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLOS ONE 18(10):* e0291668. https://doi.org/10.1371/journal.pone.0291668

Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review, 9(11)*: 39-52. https://timreview.ca/sites/default/files/article_PDF/TIMReview_November2019%20-%20D%20-%20Final.pdf

Woolley, S. & Joseff, K. (2020). Demand for deceit: How the way we think drives disinformation. Washington: *National Endowment for Democracy and International Forum for Democratic Studies.* https://www.ned.org/wp-content/uploads/2020/01/Demand-for-Deceit.pdf

Załoga, W. (2022). Disinformation in the Age of the Digital Revolution in the Aspect of State Security. *Wiedza Obronna, 280(3),* 43–62. https://doi.org/10.34752/2022-c280

Телебачення Торонто [@UkrainianTorontoTelevision]. (2024, July 7). *ХТО СТВОРЮЄ ВІРУСНІ ДІПФЕЙКИ ПРО ЗСУ: розслідування +ENG SUB. [WHO CREATES VIRAL DEEPFAKES ABOUT THE AFU: Investigation +ENG SUB]*. [YouTube video]. YouTube. Retrieved from https://youtu.be/rxRMzB_1Yvs?si=psMlZo8cesBCw9lE

ТРУХА ⚡ (@RadioTruha). (2023, November 7). ⚡*Залужный начал военный переворот на Украине. [⚡Zaluzhnyi started a military coup in Ukraine]. [Video attached] [Telegram post]. Telegram.* Retrieved from https://t.me/RadioTruha/260

ЦЕНТР ПРОТИДІЇ ДЕЗІНФОРМАЦІЇ [@Center for Countering Disinformation]. (2023, November 8). #ЦПД_застерігає: ❌ *Ворожі ТG-канали координовано поширюють діпфейки відеозвернення Головнокомандувача ЗСУ Валерія Залужного… [#CCD_warns: ❌ Hostile TG channels coordinated distribution of deepfakes of video message of the Chief of the Armed Forces of Ukraine Valeriy Zaluzhnyi]*. [Telegram post]. Telegram. Retrieved from https://t.me/CenterCounteringDisinformation/7744