Adam Majchrzak
University of Gdansk
ORCID: 0000-0003-2376-2721

Russian disinformation and the use of images generated by artificial intelligence (*deepfake*) in the first year of the invasion of Ukraine

Abstract
The article addresses the issue of the use of images generated with the use of artificial intelligence as part of disinformation in the first year of the Russian invasion of Ukraine (24.02.2022-24.02.2023). Based on the review of the literature, reports and media coverage in Polish, English, Ukrainian and Russian, examples of the use of *deepfake* in Russian disinformation were highlighted and how such technology was used and what was its significance in a theoretical context.

Keywords: disinformation, Russian disinformation, Russian invasion, artificial intelligence, deepfake.

Introduction

Disinformation is a process that involves the methodical dissemination of false and misleading[1] content in such a way as to achieve specific economic, political or military gains[2]. In political science terms – as defined by the North Atlantic Treaty Organization (NATO) – disinformation is "the deliberate creation and dissemination of false and/or manipulated information with the intent to deceive and/or mislead", and is intended to "deepen divisions within and between allied states and undermine people's confidence in elected governments"[3]. It is a process that can be used by individuals, groups, companies, criminals (non-state actors) and states as part of their policies[4]. In Western research discourse, much attention has long been paid to the problem

[1] NASK Academy, *Disinformation – what it is and how to verify it*, https://akademia.nask.pl/blog/dezinformacja—czym-jest-i-jak-ja-zweryfikowac_i23.html (accessed 5.04.2023).
[2] J. Borecki, *Disinformation as a threat to private and public enterprises*, https://warsaw-institute.org/pl/dezinformacja-jako-zagrozenie-dla-prywatnych-publicznych-przedsiebiorstw/ (accessed 5.04.2023).
[3] NATO, *NATO's approach to countering disinformation: a focus on COVID-19*, https://www.nato.int/cps/en/natohq/177273.htm (accessed 5.04.2023).
[4] Global Egnagement Center, *Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors*, https://www.state.gov/gendered-disinformation-tactics-themes-and-trends-by-foreign-malign-actors/ (accessed 5.04.2023).

of Russian disinformation which, as a tool of hybrid warfare, undermines the functioning and order of Western democratic systems[5]. Disinformation is believed to be rooted in Russian history and strategic culture[6]. With the launch of the full-scale Russian invasion of Ukraine on 24 February 2022, this topic has become particularly relevant to almost all European societies[7]. Based on reports from the first year of the invasion, it can be shown that the Russian Federation has used a wide set of disinformation measures as part of its warfare, including false and incomplete information and conspiracy theories[8], to, among other things, blame itself for causing the conflict[9], discredit Ukraine's allies[10], or to generate resentment towards Ukrainians in Europe[11]. However, there are many more threats and probably not all of them have been detected at this point – disinformation is a very malleable creature and it is often difficult to predict exactly what the reaction to certain messages will be.

Modern Russian disinformation uses a broad set of means, including, but not limited to, new media and technologies[12], as well as – of particular relevance to recent technological developments – artificial intelligence (AI for short)[13]. According to the definition proposed by the European Parliament, artificial intelligence is: "the ability of machines to exhibit human skills such as reasoning, learning, planning and creativity"[14]. In turn, according to the World Economic Forum (WEF), it is "a field of science and a type of technology characterized by the development and use of machines capable of performing

[5] OECD, *Disinformation and Russia's war of aggression against Ukraine. Threats and governance responses*, https://www.oecd.org/ukraine-hub/policy-responses/disinformation-and-russia-s-war-of-aggression-against-ukraine-37186bde/ (accessed 5.04.2023).
[6] R. Kupiecki, F. Bryjka, T. Chłoń, *Dezinformacja międzynarodowa. Pojęcie rozpoznanie, przeciwdziałanie,* Wydawnictwo Naukowe Scholar, Warsaw 2022.
[7] P. Zakhovsky, *Ukraine: the first day of the Russian invasion*, https: //www.osw.waw.pl/pl/publikacje/analizy/2022-02-25/ukraina-pierwsza-doba-rosyjskiej-inwazji (accessed 5.04.2023).
[8] I. Yablokov, *Russian disinformation finds fertile ground in the West*, https://www.nature.com/articles/s41562-022-01399-3 (accessed 5.04.2023).
[9] A. Maternik, *Disinformation in Russian: weaken Ukraine and ridicule the West*, https://demagog.org.pl/analizy_i_raporty/dezinformacja-po-rosyjsku-oslabic-ukraine-i-osmieszyc-zachod/ (accessed 5.04.2023).
[10] Ł. Jasina, *On Russian disinformation: one year since the full-scale invasion of Ukraine – commentary by the Foreign Ministry Spokesman*, https://www.gov.pl/web/dyplomacja/o-rosyjskiej-dezinformacji-rok-od-pelnoskalowej-inwazji-na-ukraine—komentarz-rzecznika-prasowego-msz (accessed 5.04.2023).
[11] Demagogue Association, *There is a war in Ukraine. This is not "denazification"!* https://demagog.org.pl/fake_news/w-ukrainie-trwa-wojna-to-nie-denazyfikacja/ (accessed 5.04.2023).
[12] M. Scholtens, *Russian Disinformation Profits from Changing Social Media Landscape*, https://www.cartercenter.org/news/features/blogs/2022/russian-disinformation-profits-from-changing-social-media-landscape.html (accessed 5.04.2023).
[13] E. Ajao, *AI and disinformation in the Russia-Ukraine war*, https://www.techtarget.com/searchenterpriseai/feature/AI-and-disinformation-in-the-Russia-Ukraine-war (accessed 5.04.2023).
[14] European Parliament, *Artificial intelligence: what is it and what are its applications?* https://www.europarl.europa.eu/news/pl/headlines/society/20200827STO85804/sztuczna-inteligencja-co-to-jest-i-jakie-ma-zastosowania (accessed 5.04.2023).

tasks that would normally require human intelligence"[15]. This technology has been developing rapidly for several years[16], and the application of AI can bring a number of simplifications to many areas of life. The proliferation of tools such as chatbots that generate automatic answers to simple and advanced questions based on AI (e.g. Chatbot GPT)[17], and arbitrary graphics generators (e.g. DALL-E and Midjourney)[18], will likely enable humanity to automate many processes. Nevertheless, the widespread availability of such technology will create an increasing risk of using AI in mass disinformation, and the time of the Russian invasion is, in a way, a 'testing ground' for the use of AI in war disinformation. AI-related technology can be a useful tool, but in contrast – when it comes to the unethical dimension of use – it can act as a weapon of mass destruction. Recently, images generated partly or entirely using AI seem to be of particular interest in this regard. Thanks to modern AI, it is possible to generate completely fictitious people or to substitute faces in images and films, which raises a serious risk of using the technology in disinformation[19]. It is difficult to imagine the total potential of such techniques, but attempts are underway to identify areas where messages with AI-generated images could be used.

Disinformation studies have already documented at least a few instances of their use of AI-generated graphics to deliberately mislead audiences[20]. Based on general premises, it is hypothesized that disinformation messages using *deepfakes have* already appeared in the information space to serve Russian interests, but their effectiveness has been limited due to active measures taken against the false messages. The primary research objective of this thesis is to theorize how artificial intelligence was used as part of Russian disinformation in the first year after the escalation of the Ukraine-Russia conflict (24.02.2022 to 24.02.2023). The stated aim was achieved by answering the following two research questions: How was Russian disinformation carried out using images generated by artificial intelligence? Also: for what purpose were images generated by artificial intelligence used as part of disinformation? Proper answers to the questions posed were made possible by a review of scientific literature and official, analytical and media information sources in Polish, Ukrainian, Russian and English from 24.02.2022 to 24.02.2023. Outlining and systematizing the use of AI in Russian disinformation will allow predicting the future use of AI for further activities of a similar disinformative nature.

---

[15] N. Routley, *What is generative AI? An AI explains*, https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/ (accessed 5.04.2023).

[16] R. Anyoha, *The History of Artificial Intelligence*, https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ (accessed 5.04.2023).

[17] OpenAI, *Introducing ChatGPT*, https://openai.com/blog/chatgpt (accessed 5.04.2023).

[18] J. Noguera, *DALL-E 2 and Midjourney can be a boon for industrial designers*, https://theconversation.com/dall-e-2-and-midjourney-can-be-a-boon-for-industrial-designers-199267 (accessed 5.04.2023).

[19] N. Barney, *Deepfake AI (deep fake)*, https://www.techtarget.com/whatis/definition/deepfake (accessed 5.04.2023).

[20] M. Cholewa, *'Pope full of drip'. AI-created photos trend online*, https://demagog.org.pl/analizy_i_raporty/papiez-pelen-dripu-zdjecia-tworzone-przez-ai-trenduja-w-sieci/ (accessed 10.05.2023).

The risk of using artificial intelligence in Russian disinformation

During the first year of the full-scale invasion, the Russian Federation undertook a number of activities aimed at disinformation and the dissemination of propaganda against Ukraine and against other countries allied to it[21]. Messages from online fact-checking organizations and entities that investigate disinformation show that Russia has repeatedly been responsible for false messages that may have served to polarize society[22], build resentment between Ukraine and European Union societies[23], and undermine trust in Ukraine's public institutions and Western countries[24]. Among other things, there were false reports online of leaked ballots in favor of detaching Lviv from Ukraine and annexing it to Poland, which would harm the country's sovereignty[25]. On another occasion, Russian broadcasts accused Ukraine of mystifying the Bucza massacre[26] or destroying a hospital in Mariupol[27]. Pro-Russian narratives also included conspiracy theories that the West and Ukraine had initiated the conflict. Such messages were intended to dilute Russia's responsibility for the war[28]. These and similar false materials were disseminated in a coordinated manner and were created in an appropriate way depending on the specific political situation and the situation on the frontline.

According to a breakdown by the Global Engagement Center at the US State Department, Russia's disinformation activities are guided by an ecosystem of 'five pillars of Russian disinformation and propaganda'. The pillars consist of information created by: official state centers (1), state-funded media centers (2), proxy sources (3), social media (4) and techniques that enable disinformation through cyberspace (5)[29]. Effective

[21] Ł. Jasina, *On Russian disinformation: one year on – a commentary from MFA Spokesperson*, https://www.gov.pl/web/diplomacy/on-russian-disinformation-one-year-on—a-commentary-from-mfa-spokesperson (accessed 10.04.2023).
[22] Demagogue Association, *Russia saved Europe from contamination? Propaganda fake news!* https://demagog.org.pl/fake_news/rosja-uratowala-europe-przed-skazeniem-propagandowy-fake-news/ (accessed 10.04.2023).
[23] I. Tomaszewska, *How language is manipulated about refugees and the war in Ukraine*, https://demagog.org.pl/analizy_i_raporty/jak-manipuluje-sie-jezykiem-na-temat-uchodzcow-i-wojny-w-ukrainie/ (accessed 10.04.2023).
[24] Polish Institute of International Affairs – PISM, *Disinformation in wartime – threats and counteraction*, https://www.pism.pl/konferencje/dezinformacja-czasu-wojny-zagrozenia-i-przeciwdzialanie (accessed 10.04.2023).
[25] Demagogue Association, *Lviv annexed to Poland....*
[26] Demagogue Association, *Bucza crimes staged? Russian disinformation!!!*, https://demagog.org.pl/fake_news/zbrodnie-w-buczy-inscenizacja-rosyjska-dezinformacja/ (accessed 10.04.2023).
[27] Demagogue Association, *Mariupol hospital shelling was a set-up? Fake news!* https://demagog.org.pl/fake_news/ostrzal-szpitala-w-mariupolu-byl-ustawka-fake-news/ (accessed 10.04.2023).
[28] J. Reid, *'They started the war': Russia's Putin blames West and Ukraine for provoking conflict*, https://www.cnbc.com/2023/02/21/russias-putin-blames-west-and-ukraine-for-provoking-conflict.html (accessed 10.04.2023).
[29] Global Engagement Center, *GEC Special Report: August 2020 Pillars of Russia's Disinformation and Propaganda Ecosystem*, https://www.state.gov/wp-content/uploads/2020/08/

disinformation implies the parallel use of all or part of the pillars simultaneously, making disinformative messages more credible in the eyes of the audience[30]. In the latter case, we can speak of events such as hacking, taking control of websites, cloning websites, creating forgeries (e.g. fabrication of photographs or photomontages) or disrupting the continuity of selected state centers and media in the online sphere[31]. According to the classification adopted, disinformation practiced by official centers is the easiest to recognize, while disinformation with unclear funding or clandestine nature is the least easy to recognize.

In the context of the fifth pillar, it is worth noting the phenomenon of *deepfake,* an image processing technique that involves superimposing human faces onto moving and still images using the action of artificial intelligence[32]. The origins of this technology date back to 2014, when Ian Goodfellow and colleagues developed the *Generative Adversarial Network* (GAN)[33]. In the simplest terms, a GAN is two adversarial neural networks, one of which tries to generate an image that is authentic to the human eye (the generator) and the other assesses its authenticity (the discriminator)[34]. In the classic view, *deepfake* consisted of the swapping of faces in recordings[35], but over time the term has come to be extended to include any AI-generated images (not only videos, but also graphics). With AI, it is possible to replace any face on available footage or generate a completely new face from scratch. Such a technique can be used to steal an image, a voice, or to destroy the achievements and reputation of certain individuals[36]. It could cause the greatest difficulties when falsifying the identity of important public figures, such as heads of state[37]. The difficulty of recognizing such forgeries depends on the amount of effort expended; some *deepfake* material may have clear imperfections and others will be virtually unrecognizable on general inspection. Over time, *deepfakes are* likely to get better and better. Should the Russian Federation decide to use such deepfakes on a large scale, it would seem appropriate to assign the use of *deepfake* technology precisely to the last pillar of disinformation and propaganda, i.e. cyber disinformation. Subsequently, such a message could be reinforced by the other pillars of Russian disinformation and propaganda.

Shortly after the start of the Russian invasion, a lot of attention was paid in Ukrainian media discourse to the issue of the use of AI-generated images – especially in February/

Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf (accessed 10.04.2023).

[30] *Ibid.*

[31] *Ibid.*

[32] M. Somers, *Deepfakes, explained*, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained (accessed 12.04.2023).

[33] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al, *Generative Adversarial Networks*, https://arxiv.org/abs/1406.2661 (accessed 12.04.2023).

[34] *Ibid.*

[35] J. Peele, *Obama Deep Fake,* https://ars.electronica.art/center/en/obama-deep-fake/ (accessed 12.04.2023).

[36] R. Kupiecki, F. Bryjka, T. Chłoń, *Dezinformacja międzynarodowa...*

[37] *Ibid.*

March 2022[38]. It was then that the slogan *дінфейк* (translated as *deepfake*) was most popular, and this has been the case for the last five years[39]. Nevertheless, in the first year of the full-scale Russian invasion of Ukraine, there were not many documented instances of AI-generated images being used as part of pro-Russian disinformation. It is difficult to assess the scale on which similar techniques will be used in the future, but the use of AI-generated false images in Russian disinformation in the first year of the invasion may have had several compelling justifications from the perspective of Russian interests, including: testing the capabilities of *deepfake* technology in disinformation (1), learning about disinformation potential understood as the reach and effectiveness of *deepfakes* (2), and assessing the adversary's ability to prevent and minimize the effects of spreading an AI-generated false message (3). There have never been convenient conditions to test such AI capabilities in disinformation warfare, as AI graphics generators were still in the early stages of development and possible *deepfakes* usually walked behind technological innovations.

Use of AI-generated images in Russian disinformation

Notwithstanding the number of instances of AI-generated images being used for disinformation in the first year of the Russian invasion of Ukraine, one in particular deserves special mention. In March 2022, a video with an image of Ukrainian President Volodymyr Zelenskiy appeared on social media, arguing that the Ukrainian military had failed and an official decision had been made to capitulate[40]. If such a version were to be believed, the decision would have taken place in contradiction to the politician's earlier declarations about continuing the fight against the invaders and supporting the citizens to fight on the ground[41]. The footage was quite static, as it showed the motionless figure of the politician at his desk[42]. No torso or arm movements were visible in the video. The president's head appeared disproportionately larger than the rest of his body, and there were characteristic blurs around and on his face, as well as unnatural

[38] T. Yavorovych, *Росія готує відеофейк із Зеленським про начебто капітуляцію України – розвідка*, https: //suspilne.media/213171-rosia-gotue-videofejk-iz-zelenskim-pro-nacebto-kapitulaciu-ukraini-rozvidka/ (accessed 12.03.2023).

[39] The popularity of the keyword 'діпфейк' in Ukrainian over the last five years can be checked in the Google Trends tool at: https://trends.google.pl/home.

[40] J. Wakefield, *Deepfake presidents used in Russia-Ukraine war*, https://www.bbc.com/news/technology-60780142 (accessed 12.03.2023).

[41] S. Braithwaite, *Zelensky refuses US offer to evacuate, saying 'I need ammunition, not a ride'*, https://edition.cnn.com/2022/02/26/europe/ukraine-zelensky-evacuation-intl/index.html (accessed 12.04.2023).

[42] Figures depicted in static poses are characteristic of classic deepfakes. Deepfakes with images of Barack Obama in 2018 and Mark Zuckerberg in 2019 were presented in a very similar pose. This presentation of people is due to the ever-present imperfection of technology in generating rapid movements; with faster changes in facial and torso positions, a fake video would be vulnerable to being exposed more quickly.

leaps in facial expression[43]. The incident was groundbreaking in its own way, as it is most likely one of the first instances of *deepfakes being* used as part of wartime disinformation in the world.

As part of its hypothetical disinformation potential, the fake AI-generated video of the Ukrainian president could have introduced additional information chaos into society (1), had a psychological impact and lowered the morale of the Ukrainian military by sowing panic (2) and undermined confidence in the Ukrainian president (3). Such material could also have been used to test the Ukrainian response (4). An important feature of such *a deepfake* was its massiveness – the message was intended to target the entire Ukrainian public. In practice, it was ineffective. What is the reason for this? In this case, the Ukrainian authorities were prepared for the possible appearance of such *a deepfake* in the public space. As early as 2 March 2022, they warned that there was a clear risk of provocation and the use of this technology in Russian disinformation[44]. Some of the findings of the services suggested that an image of the president would be used[45]. Consequently, when the *deepfake* emerged, President Volodymyr Zelenski assessed the fabricated material as a "childish provocation"[46].

The number of all cases of *deepfakes* being used for Russian disinformation may not be fully known, because the literature and media describe only those cases that were directed to a mass audience and were detected or were directed only to a narrow group of people, but decided to be declassified by the authorities that came into possession of the forged material. For example, the Security Service of Ukraine (HUR) in September 2022 described a *deepfake* case involving an impersonation of Ukrainian Prime Minister Denys Shmyhal[47]. Russian special services used his image to remotely connect with Haluk Bayraktar – the head of a Turkish company responsible for producing military drones[48]. In this case, the *deepfake* also depicted a static figure, but his head appeared more proportionate[49]. However, if we are talking about the type

---

[43] J. Hsu, *Deepfake detector spots fake videos of Ukraine's president Zelenskyy*, https://www.newscientist.com/article/2350644-deepfake-detector-spots-fake-videos-of-ukraines-president-zelenskyy/ (accessed 12.04.2023).

[44] V. Orlova, *Росіяни готуються запустити новий діпфейк із Зеленським: що цього разу вигадали у Кремлі*, https://www.unian.ua/war/dipfeyk-zelenskogo-kreml-gotuye-noviy-dipfeyk-iz-zelenskim-novini-vtorgnennya-rosiji-v-ukrajinu-11789001.html (accessed 12.04.2023).

[45] *Радіо Свобода Україна, Росія може створити діпфейк з Зеленським про капітуляцію України – Центр стратегічних комунікацій*, https: //www.radiosvoboda.org/a/news-rosia-dipfeik-pro-zelenskoho/31732835.html (accessed 12.04.2023).

[46] A. Kazmierska, W. Brzezinski, *Deepfake at war: faked recordings with Zelenski and Putin*, https://www.tygodnikpowszechny.pl/deepfake-na-wojnie-sfalszowane-nagrania-z-zelenskim-i-putinem-172209 (accessed 12.04.2023).

[47] P. Kozlowski, *Bayraktar phone call and Shmyhal deepfake. HUR foiled a provocation by Russian services,* https://technologia.dziennik.pl/aktualnosci/artykuly/8564298,hur-prowokacja-rosyjskie-sluzby-telefon-bayraktar-deepfake-wojna-ukraina-rosja.html (accessed 13.04.2023).

[48] N. Sorokinie, *Росіяни зателефонували до Туреччини, видаючи себе за Дениса Шмигаля: навіщо і що з цього вийшло (ВІДЕО)*, https://donbas24.news/news/rosiyani-zatelefonuvali-do-tureccini-vidayuci-sebe-za-denisa-smigalya-navishho-i-shho-z-cyogo-viislo-video (accessed 13.04.2023).

[49] *Ibid.*

of message, unlike the *deepfake* with the image of Volodymyr Zelenskiy, this time the message was supposed to be addressed to one specific person and not to the whole society. Again, the *deepfake* proved ineffective, as the conversation was intercepted by Ukrainian intelligence and the topic was publicized by media around the world[50]. Consequently, this was another example of the failed use of AI-generated images. In this case, the Russian Federation could have used such a *deepfake* to gain possession of sensitive and confidential information on arms supplies (1) or to discredit cooperation between Ukraine and Turkey (2) and, as in the case of *the deepfake* with the President, to assess the disinformation potential of the fabricated image (3), as well as the recipient's ability to detect it (4). Even if this was another unsuccessful attempt to use AI for disinformation after the incident, the HUR assessed that Russia has and will continue to make such attempts to develop its capabilities to use the technology in the future[51]. Thus, it cannot be said that such actions were completely pointless from the point of view of those responsible for creating disinformation.

In the first year of the Russian invasion of Ukraine, there were not many other similar incidents that could be linked to Russian disinformation activity. There could be various justifications for this – it could take a long time to generate another and more authentic *deepfake and*, in addition, Russian disinformation centers could use this time more efficiently and with less effort. It would be much simpler to prepare false information, for example, using ready-made texts and out-of-context recordings, rather than exerting effort in generating *deepfakes*.

Videos using stolen identities from 24.02.2022-24.02.2023, however, are not the only instances of AI-generated imagery being used in Russian disinformation. Even before the conflict escalated in February 2022, the Russian Federation had a history of creating false identities used within profiles of so-called trollkonts and bots[52], which disseminated selected information on social media. During the conflict, artificial intelligence was also sometimes used to create completely false identities through which the Russian point of view was reinforced online. The case indicated was publicized by, among others, NBCNews reporter Ben Collins[53]. His account indicated that a character such as Vladimir Bondarenko, for example, who presented himself as a blogger from Kiev, was in fact a completely fictitious person whose job was to make Russian narratives about the war credible in the eyes of Ukrainians and Russians. Irina Kerimova – also

[50] *Ibid.*

[51] Polish Press Agency, *Ukrainian authorities: we foiled the Russian provocation with the Bayraktar phone call and deep fake technique*, https://www.pap.pl/aktualnosci/news%2C1447515%2Cwladze-ukrainy-udaremnilismy-rosyjska-prowokacje-z-telefonem-do-bayraktara (accessed 13.04.2023).

[52] J.C. Wong, *Russian agency created fake leftwing news outlet with fictional editors, Facebook says*, https://www.theguardian.com/technology/2020/sep/01/facebook-russia-internet-research-agency-fake-news (accessed 13.04.2023).

[53] B. Collins, J. L. Kent, *Facebook, Twitter remove disinformation accounts targeting Ukrainians*, https://www.nbcnews.com/tech/internet/facebook-twitter-remove-disinformation-accounts-targeting-ukrainians-rcna17880 (accessed 13.04.2023).

fictitious – a guitar teacher from Kharkiv, was similarly acting[54]. Their face was created entirely by an AI tool. Similar to the one that can be used at www.this-person-does-not-exist.com. The indicated website allows the generation of fake faces of various people of different ages[55]. The use of such accounts may have led primarily to the amplification of Russian narratives by creating the impression of authenticity and universality of selected opinions outside Russia (1). In addition, they may have been used to create false identities in order to infiltrate public debate and gauge the sentiments of specific groups under cover (2). It is difficult to identify the number of all accounts that use false faces and remain active on social media, but social media platforms make attempts to block them[56]. For example, setting up a profile with a 'photo' of a non-existent person on Facebook can end up with the account being blocked. However, it becomes more difficult if such a photo is further modified, making it difficult to identify a fake profile. At the same time, it is similarly challenging to unambiguously identify the purposes for which false identities generated by AI are used – they are more unpredictable than individual *deepfakes,* which are more incidental and serve a specific purpose at a specific time.

Summary

The study tested the hypothesis that 'disinformative messages using *deepfakes* appeared in the information space to serve Russian interests, but their effectiveness was limited due to active measures taken against the false information'. According to available sources, it can be observed that several *deepfakes* have indeed appeared in the information space, which were created as part of Russian disinformation. However, due to the active action of Ukrainian state bodies and services, as well as individual signals from journalists, the disinformation potential of such messages minimized (at least in known cases). On the other hand, it is difficult to determine the effectiveness of such actions in the case of fake accounts, which used the image of generated figures, on social media, as not every such profile could be detected.

In assessing the course of Russian disinformation in the period 24.02.2022-24.02.2023 and answering the research question: "how did Russian disinformation using AI-generated images proceed?", it is necessary to point out that the use of AI-generated images was not widespread within Russian disinformation, but several significant incidents were

[54] J. Rokicka, *Faces of disinformation. This is what fake Facebook and Twitter accounts targeting Ukrainians looked like,* https://cyberdefence24.pl/social-media/twarze-dezinformacji-tak-wygladaly-falszywe-konta-na-facebooku-i-twitterze-skierowane-do-ukraincow (accessed 13.04.2023).

[55] The website www.this-person-does-not-exist.com features a fake face generator that works with StyleGAN technology. The generated faces can resemble real people of different ages, genders and with completely different features.

[56] Q. Wong, C. Reichert, *Facebook removes bogus accounts that used AI to create fake profile pictures*, https://www.cnet.com/news/privacy/facebook-removed-fake-accounts-that-used-ai-to-create-fake-profile-pictures/ (accessed 13.04.2023).

recorded, which set a certain precedent for the development of false online messages disseminated in a methodical manner. Such incidents could be a starting point for the development of more dangerous forms of disseminating manipulated AI-generated visual messages. Above all, more difficult to detect and more powerful. This will most likely also be encouraged by the development of AI technology, which will translate into its easier, more accurate, faster and cheaper use.

In the search for an answer to the second question, "for what purpose were AI-generated images used as part of disinformation?", it is important to point out that in the first year of the invasion, the Russian Federation used generated visual content most likely to: extend information chaos in society (1), psychologically affect the morale of the Ukrainian military (2), undermine trust in the head of state and government administration (3), come into possession of confidential and sensitive information (4), discredit international cooperation (5), reinforce selected Russian narratives (6), as well as to test the opponent's reaction to the use of fabricated images (7) and infiltrate the information environment (8). In the near future, we should expect to see AI-generated images used in similar areas on a larger scale and in a more refined form, aided by the experience gained, the progressive development of artificial intelligence and its dissemination. It should not be forgotten that disinformation involves a process of continuous learning – one message may be ineffective, but it provides lessons for the future, through which disinformation activities can be optimized. The use of AI-generated images in the first year of the Russian invasion is most likely a harbinger of further operations involving the use of *deepfake* technology in disinformation. For this reason, those and centers responsible for developing AI tools that generate authentic-looking images should prepare their tools to appropriately label the resulting material and make available the means to quickly identify it should it be used unethically.

Bibliography

Ajao E., *AI and disinformation in the Russia-Ukraine war*, https://www.techtarget.com/searchenterpriseai/feature/AI-and-disinformation-in-the-Russia-Ukraine-war (accessed 5.04.2023).

NASK Academy, *Disinformation – what it is and how to verify it*, https://akademia.nask.pl/blog/dezinformacja—-czym-jest-i-jak-ja-zweryfikowac_i23.html (accessed 5.04.2023).

Anyoha R., *The History of Artificial Intelligence,* https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/ (accessed 5.04.2023).

Barney N., *Deepfake AI (deep fake),* https://www.techtarget.com/whatis/definition/deepfake (accessed 5.04.2023).

Borecki J., *Disinformation as a threat to private and public enterprises*, https://warsawinstitute.org/pl/dezinformacja-jako-zagrozenie-dla-prywatnych-publicznych-przedsiebiorstw/ (accessed 5.04.2023).

Braithwaite S., *Zelensky refuses US offer to evacuate, saying 'I need ammunition, not a ride',* https://edition.cnn.com/2022/02/26/europe/ukraine-zelensky-evacuation-intl/index.html (accessed 12.04.2023).

Cholewa M., *'Pope full of drip'. AI-created photos trend online*, https://demagog.org.pl/analizy_i_raporty/papiez-pelen-dripu-zdjecia-tworzone-przez-ai-trenduja-w-sieci/ (accessed 10.05.2023).

Collins B., Kent J.L., Facebook, *Twitter remove disinformation accounts targeting Ukrainians*, https://www.nbcnews.com/tech/internet/facebook-twitter-remove-disinformation-accounts-targeting-ukrainians-rcna17880 (accessed 13.04.2023).

Global Engagement Center, *GEC Special Report: August 2020 Pillars of Russia's Disinformation and Propaganda Ecosystem*, https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf (accessed 10.04.2023).

Global Egnagement Center, *Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors*, https://www.state.gov/gendered-disinformation-tactics-themes-and-trends-by-foreign-malign-actors/ (accessed 5.04.2023).

Goodfellow I.J., Pouget-Abadie J., Mirza M. et al, *Generative Adversarial Networks*, https://arxiv.org/abs/1406.2661 (accessed 12.04.2023).

Hsu J., *Deepfake detector spots fake videos of Ukraine's president Zelenskyy*, https://www.newscientist.com/article/2350644-deepfake-detector-spots-fake-videos-of-ukraines-president-zelenskyy/ (accessed 12.04.2023).

Jasina Ł, *On Russian disinformation: one year on – a commentary from MFA Spokesperson*, https://www.gov.pl/web/diplomacy/on-russian-disinformation-one-year-on—a-commentary-from-mfa-spokesperson (accessed 10.04.2023).

Jasina Ł., *On Russian disinformation: one year since the full-scale invasion of Ukraine – commentary by the Foreign Ministry Spokesman*, https://www.gov.pl/web/dyplomacja/o-rosyjskiej-dezinformacji-rok-od-pelnoskalowej-inwazji-na-ukraine—komentarz-rzecznika-prasowego-msz (accessed 5.04.2023).

Yavorovych T., *Росія готує відеофейк із Зеленським про начебто капітуляцію України – розвідка*, https://suspilne.media/213171-rosia-gotue-videofejk-iz-zelenskim-pro-nacebto-kapitulaciu-ukraini-rozvidka/ (accessed 12.03.2023).

Kazmierska A., Brzezinski W., *Deepfake at war: faked recordings with Zelenski and Putin*, https://www.tygodnikpowszechny.pl/deepfake-na-wojnie-sfalszowane-nagrania-z-zelenskim-i-putinem-172209 (accessed 12.04.2023).

Kupiecki R., Bryjka F., Chłoń T., *Dezinformacja międzynarodowa. Pojęcie rozpoznanie, przeciwdziałanie,* Wydawnictwo Naukowe Scholar, Warsaw 2022.

Kozlowski P., *Bayraktar phone call and Shmyhal deepfake. HUR foiled a provocation by Russian services,* https://technologia.dziennik.pl/aktualnosci/artykuly/8564298,hur-prowokacja-rosyjskie-sluzby-telefon-bayraktar-deepfake-wojna-ukraina-rosja.html (accessed 13.04.2023).

Maternik A., *Disinformation in Russian: weaken Ukraine and ridicule the West*, https://demagog.org.pl/analizy_i_raporty/dezinformacja-po-rosyjsku-oslabic-ukraine-i-osmieszyc-zachod/ (accessed 5.04.2023).

NATO, *NATO's approach to countering disinformation: a focus on COVID-19*, https://www.nato.int/cps/en/natohq/177273.htm (accessed 5.04.2023).

Noguera J., *DALL-E 2 and Midjourney can be a boon for industrial designers*, https://theconversation.com/dall-e-2-and-midjourney-can-be-a-boon-for-industrial-designers-199267 (accessed 5.04.2023).

OECD, *Disinformation and Russia's war of aggression against Ukraine. Threats and governance responses*, https://www.oecd.org/ukraine-hub/policy-responses/disinformation-and-russia-s-war-of-aggression-against-ukraine-37186bde/ (accessed 5.04.2023).

OpenAI, *Introducing ChatGPT*, https://openai.com/blog/chatgpt (accessed 5.04.2023).

Orlova V., *Росіяни готуються запустити новий дінфейк із Зеленським: що цього разу вигадали у Кремлі*, https://www.unian.ua/war/dipfeyk-zelenskogo-kreml-gotuye-noviy-dipfeyk-iz-zelenskim-novini-vtorgnennya-rosiji-v-ukrajinu-11789001.html (accessed 12.04.2023).

European Parliament, *Artificial intelligence: what is it and what are its applications?* https://www.europarl.europa.eu/news/pl/headlines/society/20200827STO85804/sztuczna-inteligencja-co-to-jest-i-jakie-ma-zastosowania (accessed 5.04.2023).

Peele J., *Obama Deep Fake,* https://ars.electronica.art/center/en/obama-deep-fake/ (accessed 12.04.2023).

Polish Press Agency, *Ukrainian authorities: we foiled the Russian provocation with the Bayraktar phone call and deep fake technique*, https://www.pap.pl/aktualnosci/news%2C1447515%2Cwladze-ukrainy-udaremnilismy-rosyjska-prowokacje-z-telefonem-do-bayraktara (accessed 13.04.2023).

Polish Institute of International Affairs – PISM, *Disinformation in wartime – threats and counteraction*, https://www.pism.pl/konferencje/dezinformacja-czasu-wojny-zagrozenia-i-przeciwdzialanie (accessed 10.04.2023).

Reid J., *'They started the war': Russia's Putin blames West and Ukraine for provoking conflict*, https://www.cnbc.com/2023/02/21/russias-putin-blames-west-and-ukraine-for-provoking-conflict.html (accessed 10.04.2023).

Rokicka J., *Faces of disinformation. This is what fake Facebook and Twitter accounts targeting Ukrainians looked like,* https://cyberdefence24.pl/social-media/twarze-dezinformacji-tak-wygladaly-falszywe-konta-na-facebooku-i-twitterze-skierowane-do-ukraincow (accessed 13.04.2023).

Routley N., *What is generative AI? An AI explains*, https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/ (accessed 5.04.2023).

Scholtens M., *Russian Disinformation Profits from Changing Social Media Landscape*, https://www.cartercenter.org/news/features/blogs/2022/russian-disinformation-profits-from-changing-social-media-landscape.html (accessed 5.04.2023).

Somers M., *Deepfakes, explained*, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained (accessed 12.04.2023).

Sorokinie N., *Росіяни зателефонували до Туреччини, видаючи себе за Дениса Шмигаля: навіщо і що з цього вийшло (ВІДЕО)*, https://donbas24.news/news/rosiyani-zatelefonuvali-do-tureccini-vidayuci-sebe-za-denisa-smigalya-navishho-i-shho-z-cyogo-viislo-video (accessed 13.04.2023).

Demagogue Association, *Lviv annexed to Poland? Beware of a fabricated document!* https://demagog.org.pl/fake_news/lwow-przylaczany-do-polski-uwaga-na-sfabrykowany-dokument/ (accessed 10.04.2023).

Demagogue Association, *Mariupol hospital shelling was a set-up? Fake news!* https://demagog.org.pl/fake_news/ostrzal-szpitala-w-mariupolu-byl-ustawka-fake-news/ (accessed 10.04.2023).

Demagogue Association, *Russia saved Europe from contamination? Propaganda fake news!* https://demagog.org.pl/fake_news/rosja-uratowala-europe-przed-skazeniem-propagandowy-fake-news/ (accessed 10.04.2023).

Demagogue Association, *There is a war in Ukraine. This is not "denazification"!* https://demagog.org.pl/fake_news/w-ukrainie-trwa-wojna-to-nie-denazyfikacja/ (accessed 5.04.2023).

Demagogue Association, *Bucza crimes staged? Russian disinformation!!!*, https://demagog.org.pl/fake_news/zbrodnie-w-buczy-inscenizacja-rosyjska-dezinformacja/ (accessed 10.04.2023).

Tomaszewska I., *How language is manipulated about refugees and the war in Ukraine*, https://demagog.org.pl/analizy_i_raporty/jak-manipuluje-sie-jezykiem-na-temat-uchodzcow-i-wojny-w-ukrainie/ (accessed 10.04.2023).

Wakefield J., *Deepfake presidents used in Russia-Ukraine war*, https://www.bbc.com/news/technology-60780142 (accessed 12.03.2023).

Wong J.C., *Russian agency created fake leftwing news outlet with fictional editors, Facebook says*, https://www.theguardian.com/technology/2020/sep/01/facebook-russia-internet-research-agency-fake-news (accessed 13.04.2023).

Wong Q., Reichert C., *Facebook removes bogus accounts that used AI to create fake profile pictures*, https://www.cnet.com/news/privacy/facebook-removed-fake-accounts-that-used-ai-to-create-fake-profile-pictures/ (accessed 13.04.2023).

Yablokov I., *Russian disinformation finds fertile ground in the West*, https://www.nature.com/articles/s41562-022-01399-3 (accessed 5.04.2023).

Zakhovsky P., *Ukraine: the first day of the Russian invasion*, https://www.osw.waw.pl/pl/publikacje/analizy/2022-02-25/ukraina-pierwsza-doba-rosyjskiej-inwazji (accessed 5.04.2023).

Радіо Свобода Україна, *Росія може створити дінфейк з Зеленським про капітуляцію України – Центр стратегічних комунікацій*, https://www.radiosvoboda.org/a/news-rosia-dipfeik-pro-zelenskoho/31732835.html (accessed 12.04.2023).