Artificial Intelligence and the International Information and Psychological Security

Konstantin A. Pantserev and Konstantin A. Golubev

Abstract The paper considers the problem of malicious use of technologies that are based on artificial intelligence (AI). The authors presume that the need to develop advanced technologies is seen by states as essential to ensuring their global leadership and technological sovereignty. Particular focus is on AI-based technologies whose capabilities are growing at unprecedented rates. AI has already become part and parcel of intelligent machine translation and transport systems. AI-based technologies are widely used in medical diagnostics, e-Commerce, online training and even in the production of news and information. Meanwhile, world's top search engines have offered their users voice assistants that significantly simplify and accelerate search for relevant information. Yet, evidently most technological innovations that are meant to make our lives easier could potentially be used for malicious purposes. Therefore, the rapid growth of our dependency on hybrid computer intelligent systems renders national critical infrastructure extremely vulnerable to attacks by those who would like to use AI-based technologies to cause significant harm to a nation, which in turn poses a serious challenge to psychological and information security of people around the world. The paper discusses ways of malicious use of AI and offers possible instruments of mitigating the threat that advanced technologies are posing.

Keywords Artificial intelligence · Strategic communication · International information and psychological security · Information and psychological warfare

1 Introduction

Developing advanced technologies is deemed a sine-qua-non to ensure one's global leadership in today's world. Particular focus is on technologies that are based on artificial intelligence (AI). The latter's capabilities have been growing at unprecedented rates. Nowadays, AI-algorithms are widely used in intelligent machine translation systems, medical diagnostics, electronic commerce, on-line education, intelligent

K. A. Pantserev (✉) · K. A. Golubev
Saint-Petersburg State University, 7/9 Universitetskaya Emb, 199034 Saint-Petersburg, Russia
e-mail: k.pantserev@spbu.ru

H. Jahankhani et al. (eds.), *Cybersecurity, Privacy and Freedom Protection in the Connected World*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-68534-8_1

transport systems and even in the production of news and information. Meanwhile, world's top search engines have offered their users voice assistants that significantly simplify and accelerate search for relevant information.

Often a time, developers of AI-based solutions receive financial support from their national governments. According to official data, over 30 countries have adopted national strategies and roadmaps related to AI. Those include the USA, China, France, Japan, Russia, and UAE to name just a few [1].

Yet, evidently most technological innovations that are meant to make our lives easier could potentially be used for malicious purposes. Therefore, the rapid growth of our dependency on hybrid computer intelligent systems renders national critical infrastructure extremely vulnerable to attacks by those who would like to use AI-based technologies to cause significant harm to a nation, which in turn poses a serious challenge to psychological and information security of people around the world.

2 Artificial Intelligence: From Imaginary Threats to Real Ones

As a result of the advances in computer science, complex algorithms are becoming an integral part of people's daily routines. Yet, one has to be conscious that those technologies in fact could present a real threat to individual and collective security at both national and international levels. Thus, researchers who are involved in the development of AI are divided over the issue of assessing the threat to the future of humanity in the age of smart societies. Some (e.g. Bill Gates, Mark Zuckerberg) view technological breakthroughs in AI as an opportunity for people to leave behind monotonous work and engage in self-actualization. Yet, others (e.g. Moshe Vardi, Steven Hocking) admonish that a rapid development of such technologies could trigger mass unemployment or even World War III. Therefore, we should get ready to deal with the risks coming from advanced technologies, trying to mitigate those before they turn into a real threat. That is why it seems crucial to elaborate effective mechanisms of legislature at both national and international levels to regulate the application of AI-technologies.

Particularly daunting are the rapid rates at which contemporary information technologies develop, so clearly neither national legislature, nor the system of international law, let alone the existing mechanisms of control, can keep up with those. This is the main challenge of the digital age. Most technological innovations developed over the recent years are supposed to make people's lives easier, yet the lack of effective control mechanisms, as well as that of a proper regulatory framework, dramatically increases the risk of malicious use of such technologies.

As D. Bazarkina and E. Pashentsev argue, possible malicious use of artificial intelligence "can cause serious destabilizing effects on the social and political development of the country and the system of international relations" [2]. However, given that AI-based technologies are a relatively recent phenomenon, there are few if any

real cases of malicious use of artificial intelligence either by actual actors of international psychological warfare or by terrorists. Yet, it seems almost inevitable that destructive elements of all kinds, when they get familiar with such technologies, will put the latter to malicious use before long. Therefore, we need to be prescient about the ways of possible malicious use of AI in order to elaborate pre-emptive measures of neutralizing those threats. Nonetheless, some threats have already materialized.

Thus we suggest dividing all threats connected to artificial intelligence into two types, namely latent challenges that threaten international psychological security and actual ones that threaten individual psychological security. Latent ways of malicious use of artificial intelligence have been studied by professors Bazarkina and Pashentsev in their work "Artificial Intelligence and New Threats to International Psychological Security." In it they highlight the following potential threats that have to do with an active use of advanced technologies:

– Malicious takeover of integrated, all-encompassing systems that either actively or primarily use AI;
– Delivery of explosives by or causing crashes with commercial AI systems such as drones or autonomous vehicles [2].

To that list should be added such threats as:

– Autonomous weapon systems programmed for killing. Nowadays, most leading nations are racing to develop various intelligent armament systems. One can speculate that in the nearest future the nuclear arms race will give way to that of developing military hybrid intelligent systems. One can only imagine what would happen if a national army lost control over such intelligent systems or if such systems fell into the hands of terrorists.
– Social manipulative practices. Currently, social media aided by AI-algorithms can effectively target prospective consumers. Likewise, using access to personal data of millions of people, knowing their needs, strengths and weaknesses, AI-algorithms could be used maliciously also to engage in mind manipulation and direct propaganda at specific audiences.
– Invasion of privacy. This threat is already with us. Now ill-minded individuals can track every step of online users, as well as calculate the exact times of their doing daily chores. In addition, surveillance cameras are being installed almost everywhere, taking advantage of facial recognition algorithms that easily identify each and every one.
– Mistakes by operators. The human element will remain crucial even if there appear smart machines capable of learning on their own. AI is valuable to us primarily because of its high productivity and efficiency. However, if we fail to define the objectives clearly for an AI-system, their optimal attainment could have unintended consequences.
– Lack of data. As is well known, AI-algorithms are based on processing data and information. The more data is fed into the system, the more accurate the result will be. Yet, if there is insufficient data to perform a particular task or if the data

has been corrupted by perpetrators, it could cause fatal glitches within the entire AI-system, which could cause unpredictable outcomes.

Finally, we should pinpoint yet another risk factor that might be the most significant one—our increased dependency on advanced technologies that are integrated into every aspect of daily life of each individual, overseeing the functioning of numerous applications and even most critical infrastructure. Clearly such technologies will be a magnet to ill-minded people and terrorists of all sorts.

In particular, nowadays one can observe a rapid growth of complex intelligent automated systems. Such systems span a wide range of useful applications such as the organisation and optimisation of traffic or the management of large facilities and infrastructure. In the meantime, one should bear in mind that all those intelligent systems can become an easy target to high-tech terrorist attacks. One can only imagine the consequences of an interception of control over the transport management system of a major city such as New York, Moscow, Paris, Sidney, Beijing, Shanghai, Tokyo, Seoul or over intelligent navigation systems that are used by vessels and aircraft. Undoubtedly such incidents could result in numerous casualties or cause panic and create a climate that from information and psychological perspectives would be conducive to further hostile actions.

Another plausible scenario is for future high-tech AI-based systems that manage energy grids of large industrial regions to be hacked by an adversary or terrorists, resulting in unfathomable damage.

The aforementioned examples are only imaginary threats; nevertheless, society must come to grips with such challenges. In our opinion, as of now, the real threat comes not from hypothetic terrorists trying to gain control over hybrid intelligent automated systems but from relatively experienced perpetrators who possess sufficient knowledge of AI-algorithms. For example, they can devise intelligent bots whose mission is to misinform ordinary people, create unfavourable public opinion or manipulate news agenda [3]. Undoubtedly such bots can be widely used in defamation campaigns against individuals, groups and entire nations.

The greatest danger still, as of now, is posed by those fake videos created with the help of AI, whose fabricated yet realistic footage is capable of confusing the general public regarding what in fact is happening around the world.

In and of itself, the technology (Generative Adversarial Networks) of making fake videos, also known as deepfakes, applies certain algorithms to synthesise facial movements of humans based on AI. Most importantly, one needs neither fancy skills nor knowledge in AI or machine learning to create such videos. All tasks are performed by a computer application which can be freely downloaded from the Internet.

One of such applications is FakeApp. It can superimpose a person's face onto someone else's. It has a rather intuitive interface and a detailed help file about how to produce deepfake videos. What is required is a sufficient amount of authentic video footage of the person one intends to recreate so that the application could render it to be as realistic as possible.

The application allows users with the most basic skills in computers to produce replicas of any person that walk and talk like their prototype and pronounce any nonsense one can only think of.

This latter threat has already become part of our reality. Starting in December 2017, there appeared on the Web a number of fake porno video clips starring well-known Hollywood actresses such as Gal Gadot, Chloë Moretz, Jessica Alba and Scarlett Johansson. Those films by no means pose a threat to international psychological security but they demonstrate a possibility to threaten personal psychological security because few people would want to be cast in such films. Yet, what is worse, there are already fake videos out there featuring such salient political figures as Vladimir Putin, Donald Trump and Barak Obama.

So far, fake footage featuring world leaders has been put out either for fun or to demonstrate the capability of this new technology. However, some experts caution that in the future deepfakes could become so realistic that this could have a detrimental impact on world politics as a whole [4].

One of the most extraordinary examples of deepfake videos was the one created by American filmmaker Jordan Peel back in April 2018. It featured the former US President Barak Obama badmouthing the incumbent President Donald Trump. The idea behind the film was not to misinform the audience but to draw attention to the potential danger of this new advanced technology [5].

The Russian President Vladimir Putin has also been a target of such deepfakes. One of such video clips featured Mr. Putin boasting that it was him who put Donald Trump into the Oval Office [6].

Last but not least, there was a phony video that showed U.S. President Donald Trump speaking at the White House, declaring that the United States was going to withdraw from the Paris climate agreement and calling on Belgium to follow suit. That video was made and distributed by the Flemish Socialist Party. So far, this has been the first and only real case of using deepfake technology to pursue political aims. Still, it was a relatively innocuous endeavour since its creators did not intend to misinform their audiences. On the contrary, at the end of it, the fictitious character of Donald Trump revealed that the video was fake. In fact, the Flemish Socialist Party simply wanted to draw public attention once again to the problem of climate change by calling on the people to sign a petition to invest more money into alternative sources of energy and electric cars and to shut down the nuclear power plant Doel in Flanders [7].

Still, the above cases clearly demonstrate the harmful potential of this new technology that allows any person with basic computer skills, using a desktop application available to be downloaded from the Web, to create a deepfake video of any person and to put any words in his or her mouth. It seems highly likely that terrorist groups will shortly take advantage of this promising technology and start using it in order to incite ethnic hatred or to recruit new acolytes to their ranks. This last point brings us to a conclusion that in the current age of fake news and disinformation society will have to grapple with some serious challenges to national and international security.

We are already suffering from information overload due to huge volumes of information created not only by professional journalists but ordinary social media users

who very often disseminate unwittingly or on purpose unverified stories or even outright lies. The advent of technology capable of creating deepfake videos presents yet another challenge—we can no longer trust what we see or hear for any video that one comes across on the Web, no matter how veritable it may seem, may in fact be fake, contributing further to "widespread uncertainty" which "may enable deceitful politicians to deflect accusations of lying by claiming that nothing can be proved and believed" [8].

Worse still, according to some experts, in the nearest future, as deepfake technology progresses, one simply will not be able to tell a real video from a fake one [9]. The future, then, may turn out rather terrifying. One can imagine the chaos that would result if a fake video came out with the US President announcing an "impending nuclear missile attack on North Korea" [10]. This could absolutely perplex audiences as to what is happening to their world. And if people are not able to trust what they see or hear, they will choose for themselves what they want to believe [11]. As Chesney and Citron argue:

"Imagine a video depicting the Israeli prime minister in private conversation with a colleague, seemingly revealing a plan to carry out a series of political assassinations in Tehran. Or an audio clip of Iranian officials planning a covert operation to kill Sunni leaders in a particular province of Iraq. Or a video showing an American general in Afghanistan burning a Koran. In a world already primed for violence, such recordings would have a powerful potential for incitement. Now imagine that these recordings could be faked using tools available to almost anyone with a laptop and access to the Internet—and that the resulting fakes are so convincing that they are impossible to distinguish from the real thing. Advances in digital technologies could soon make this nightmare a reality. Thanks to the rise of "deepfakes"—highly realistic and difficult-to-detect digital manipulations of audio or video—it is becoming easier than ever to portray someone saying or doing something he or she never said or did" [12].

The cases cited above lead one to ponder about how we should respond to those challenges. What clearly comes out is that it is necessary to work out effective mechanisms that could put a lid on uncontrolled spread of deepfakes of this kind so as to neutralize their toxic content. In the meantime, computer science experts are working hard to come up with specific algorithms that would be able to detect deepfakes. Once in effect, such algorithms could be used by social media platforms such as Facebook and Twitter to inspect and mark up uploaded videos before they could be viewed by other users.

Yet, as of now, there are no efficient algorithms that could detect deepfakes with 100% certainty. Moreover, it should be noted that deepfake technology is getting better with each detection cycle. And this is just one side of the coin. The other one is that there exists no law to regulate the process of creation and distribution of deepfakes.

When working on the appropriate legislation, aiming to stop the spread of toxic deepfakes, we would run into a serious problem since any unjustified prohibition to create or distribute fake videos would be interpreted as a violation of the basic principle of freedom of speech and expression. Therefore, we believe it important that

the legislature draw a distinction between malicious use of deepfakes that are aimed to create some toxic content and the sort of satire, creative effort and self-expression. Until this conundrum is solved, there will be no law to regulate the process of creation and distribution of deepfakes. This means that in the short term one should expect a flurry of highly realistic and difficult-to-detect deepfakes to come out, raising the risk of destabilisation of international system and threatening the global psychological security.

3 Conclusion

Nowadays, most advanced nations increasingly focus on developing AI-based systems. This trend in science and technology becomes a key priority for the national development of every country to ensure technological sovereignty in the contemporary digital age. Yet, the rapid integration of AI into our daily life increases the risk of malicious use of such technologies, which can put personal, national and international psychological security in jeopardy.

In our opinion the real and present danger, as of now, lies with deepfakes. In the contemporary information environment, it has become the rule of thumb to double-check any information published on the Web. However, the emergence of deepfakes puts disinformation at a qualitatively new level where it could be used maliciously not only by cyber prankers but also by various perpetrators, terrorists and other destructive elements. The numerous cases of deepfake videos out there has shown clearly that this new technology could be used to blackmail and terrify people who in fact have done nothing wrong.

Thus it becomes crucial for the international community to elaborate effective mechanisms to control the spread of deepfakes with toxic content. Nowadays, computer science experts are trying to elaborate appropriate algorithms to detect deepfakes. The problem, however, is that there currently exists no methodology to ensure 100 percent detection rate. Worse still, the practical implementation of such algorithms would be stumbling over some serious legislative obstacles for as long as this field remains free of legislative regulation.

Given all that, it seems extremely important to elaborate an appropriate legal base as quickly as possible to deal with deepfakes so as to distinguish between malicious use of this technology and the innocuous one—that of satire, creativity and self-expression. Until then, we can expect some toxic content created with the help of AI-based deepfake technology to threaten the stability of political systems around the world.

References

1. Dutton T (2016) An overview of national AI strategies. Politics + AI, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd. Last accessed 15 Aug 2016
2. Bazarkina D, Pashentsev E (2019) Artificial intelligence and new threats to international psychological security. Russia in Global Affairs 1 (2019). https://eng.globalaffairs.ru/articles/artificial-intelligence-and-new-threats-to-international-psychological-security. Last accessed 17 Aug 2016
3. Horowitz MC, Scharre P, Allen GC, Frederick K, Cho A, Saravalle E (2018) Artificial intelligence and international security. Washington: center for a new American security (2018). http://www.cnas.org/publications/reports/artificial-intelligence-and-international-security. Last accessed 17 Aug 2016
4. Palmer A (2018) Experts warn digitally-altered 'deepfakes' videos of Donald Trump, Vladimir Putin, and other world leaders could be used to manipulate global politics by 2020. Mail Online. http://www.dailymail.co.uk/sciencetech/article-5492713/Experts-warn-deepfakes-videos-politicians-manipulated.htm. Last accessed 19 Aug 2016
5. Pantserev KA (2020) The malicious use of AI-based deepfake technology as the new threat to psychological security and political stability. In: Jahankhani H, Kendzierskyj S, Chelvachandran N, Jimenez JI (eds) Cyber defence in the age of AI, smart societies and augmented humanity. Springer Nature Switzerland AG, pp 37–55
6. Putin: Face Replacement. https://www.youtube.com/watch?time_continue=3&v=hKxFqxCaQcM&feature=emb_logo. Last accessed 20 Aug 2016
7. Von der Burchard H (201) Belgian socialist party circulates 'deep fake' Donald Trump video (2018). https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video. Last accessed 20 Aug 2016
8. Vaccari C, Chadwick A (2020) Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Social Media + Society
9. Browne R (2018) Anti-election meddling group makes A.I.-powered Trump impersonator to warn about 'deepfakes'. https://www.cnbc.com/2018/12/07/deepfake-ai-trump-impersonator-highlights-election-fake-news-threat.html. Last accessed 23 Aug 2016
10. Harris D (2018) Deepfakes: false pornography is here and the law cannot protect you. Duke Law Technol Rev 17(1)
11. Dack S (2019) Deep fakes, fake news, and what comes next. https://jsis.washington.edu/news/deep-fakes-fake-news-and-what-comes-next. Last accessed 23 Aug 2016
12. Chesney R, Citron D (2019) Deepfakes and the new disinformation war: the coming age of post truth geopolitics. Foreign Affairs