Professor Evgeny Pashentsev spoke on the malicious use of artificial intelligence and new challenges for information and psychological security of Russia at the Military Academy of the General Staff of the Armed Forces of the Russian Federation

On May 16, 2019, the Military Academy of the General Staff of the Armed Forces of the Russian Federation hosted a research seminar on "Multipolarity as a factor of world stability and security of the Russian Federation". The event was attended by representatives of the State Duma of the Federal Assembly of the Russian Federation, the Joint Staff of the Collective Security Treaty Organization (CSTO), higher education institutions and research organizations of the Ministry of Defence of the Russian Federation. Among the speakers were the representatives of the Military Academy of the General staff of the Armed Forces, Lomonosov Moscow State University, Financial University under the Government of the Russian Federation, Institute of Contemporary International Studies of the Diplomatic Academy of the Ministry of Foreign Affairs of the Russian Federation, Institute of Security Problems of the CIS, Academy of Management of the Ministry of Internal Affairs of the Russian Federation and other universities and organizations. During the research seminar, contemporary concepts of the multipolar world, the processes of transformation of the main centers of power, factors affecting the security of the Russian Federation, and other issues were discussed.

In his report "Malicious use of artificial intelligence and new challenges for information and psychological security of Russia" DSc Professor, leading researcher of ICIS of DA, Professor of Lomonosov Moscow State University. Evgeny Pashentsev presented his vision of current and future challenges for psychological security of Russia in the context of the growing malicious use of artificial intelligence (AI).

The overall extremely tense situation in the world poses an undoubted threat to Russia's national security. An important element of this threat is the rapidly growing risks of AI being used to manipulate public opinion at the international level. Prof. Pashentsev defined international psychological security **(**IPS) as protecting the system of international relations from negative

information and psychological influences associated with various factors of international development. The latter include targeted efforts by various state, non-state and supranational actors to achieve partial/complete, local/global, short-term/long-term, and latent/open destabilization of the international situation in order to gain competitive advantages, even through the physical elimination of the enemy.

International actors engaging in hybrid warfare exert negative *direct* and *indirect* impacts on the enemy's public consciousness and, often, on themselves, their allies and neutral actors. The malicious use of artificial intelligence (MUAI) can become a very serious threat to the information and psychological security of Russia, as our country interacts with the outside world through many diverse ties at the state, group and individual levels. It is impossible and impractical to fully control them, and even more so to break them, because Russia itself will suffer first of all from this drastic rupture of ties. Targeted use of AI can dramatically increase the effectiveness of psychological operations against Russia, which requires a systematic analysis of this problem.

According to Evgeny Pashentsev MUAI can allow hostile actors to be more successful than so far in:

• *provoking a public reaction to* a *non-existent factor of social development* in the interests of the customer of psychological impact. The target audience sees something that doesn't really exist.

• *presenting* a *false interpretation of the existing factor* of social development and thus provoking the desired target reaction. The audience sees what exists, but in a false light.

• significantly and dangerously *strengthening (weakening) public reaction to the real factor* of social development. The audience sees what exists but reacts inadequately.

Prof. Pashentsev suggests the *following MUAI classification* according to the degree of implementability:

• current MUAI practices;

• existing MUAI capabilities that have not been used in practice yet (this probability is associated with a wide range of new rapidly developing AI capabilities — not all of them are immediately included in the range of implemented MUAI capabilities);

• future MUAI capabilities based on current developments and future research (assessments should be given for the short, medium and long term);

• unidentified risks, also known as "the unknown in the unknown." Not all AI developments can be accurately assessed. Readiness to meet unexpected hidden risks is crucial.

It is important and necessary to use independent teams of different specialists and AI systems to assess MUAI capabilities.

Prof Pashentsev also proposes the following MUAI classifications:

• by territorial coverage: local, regional, global;

• by the degree of damage: insignificant, significant, major, catastrophic;

• by the speed of propagation: slow, fast, rapid;

• by the form of propagation: open, hidden.

Evgeny Pashentsev named the most important factors favorable for MUAI that can cause serious destabilizing effects on the social and political development of Russia and the system of international relations, including the sphere of IPS: growth of integrated, all-encompassing systems with active or leading AI use, creation of 'deepfakes', 'Fake People' technology, agenda setting and amplification through AI, prognostic weapons etc. (Bazarkina and Pashentsev, 2019). Many threats by MUAI are neutralized with AI.

For example, AI allows increase dramatically data processing speed and respond faster to people's expectations, which makes phishing more dangerous. Progress in automated spear phishing has demonstrated that automatically generated text can be effective at fooling humans, and indeed very simple approaches can be convincing to humans, especially when the text pertains to certain topics such as entertainment (Brundage, et al., 2018, p. 3, 46). Main methods of using artificial intelligence hackers are phishing, spear phishing and whaling, i.e. phishing focused on senior managers responsible for financial decision-making. However, *Swisscom Innovations* developed and trained an artificial intelligence based phishing detection system. It predicts reliably whether a formerly unknown website contains phishing or not (Bürgi, 2016). Another programme, *Lookout Phishing AI* continuously scans the Internet looking for malicious websites. *Lookout Phishing AI* detects the early signals of phishing, protects end users from visiting such sites as they come up, and alerts the targeted organizations (Richards, J., 2019).

Evgeny Pashentsev proposed some specific measures to respond to the MUAI in the field of psychological security of Russia.

In conclusion, the speaker stressed that the task today is to repel threats from the real and constantly developing "weak" artificial intelligence, which is a threat not in itself, but because of the actions of antisocial external and internal actors that turn it into a threat to the national security of Russia. In the not so distant future, there may be problems associated with "strong intelligence", the possibility of which in the coming decades, forecast more and more researchers.

References:

Bazarkina, D., Pashentsev, E., 2019. Artificial Intelligence and New Threats to International Psychological Security. Russia in Global Affairs, Issue 1, pp.147-170.

Brundage, et al., 2018. The malicious use of artificial intelligence: forecasting, prevention, and mitigation. Oxford, AZ: Future of Humanity Institute, University of Oxford.

Bürgi, U., 2016. Using Artificial Intelligence to Fight Phishing. Swisscom [online]. Available at: <https://ict.swisscom.ch/2016/11/using-artificial-intelligence-to-fight-phishing/> [Accessed 22 June 2019].

Doyle, A., et al., 2014. Forecasting significant societal events using the EMBERS streaming predicative analytics system. Big Data, Vol. 4, pp. 185–195.

Horowitz, M. C., et al., 2018. Artificial intelligence and international security. Washington: Center for a New American Security (CNAS).

Karras, T., Laine, S., and Aila, T., 2018. A style-based generator architecture for generative adversarial networks. arXiv of Cornell University [online]. Available at: <https://arxiv.org/pdf/1812.04948.pdf> [Accessed 31 January 2019].

Larina, E., and Ovchinskiy, V., 2018. Iskusstvenny? intellekt. Bol'shie dannye. Prestupnost' [Artificial intelligence. Big Data. Crime]. Moscow: Knizhnyj mir.

Richards, J., 2019. What is Lookout Phishing AI? Lookout Blog. <https://blog.lookout.com/lookout-phishing-ai> [Accessed 22 June 2019].

The Times of Israel, 2018. 'I Never Said That!' The High-Tech Deception of 'Deepfake' Videos. The Times of Israel [online]. Available at: <https://www.timesofisrael.com/i-never-said-that-the-high-tech-deception-of-deepfake-videos/> [Accessed 31 January 2019].

Waddel, K., 2018. The impending war over deepfakes. Axios [online]. Available at: <https://www.axios.com/the-impending-war-over-deepfakes-b3427757-2ed7-4fbc-9edb-45e461eb87ba.html> [Accessed 31 January 2019].