

PSYRGGU 2019
**Psychology of subculture: Phenomenology and Contemporary
Tendencies of Development**

**THE USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN
INFORMATION AND PSYCHOLOGICAL WARFARE**

Evgeny A. Mikheev (a)*, Timofey A. Nestik (b)
*Corresponding author

(a) Institute of psychology, Russian academy of sciences, Moscow, Russia, mih-news@mail.ru
(b) Institute of psychology, Russian academy of sciences, Moscow, Russia, nestik@ipras.ru

Abstract

In this article scientific developments of scientists in the field of information and psychological confrontation are investigated. An attempt is made to investigate the use of artificial intelligence technologies in the Internet, the attitude of users to AI, ways to counter the spread of misinformation and fake news. The social perception of AI threats is not related to the technology itself, but to its use by people – employers, hackers, the state. A special attention is paid to psychological consequences of the so-called «deep fakes» (digital audio, video fakes) that are generated by the systems, based on the use of the artificial intelligence. The efforts of digital communities to protect themselves against the “deep fakes” lead to the construction of the so-called “echo chambers”. To reduce the exposure of online communities and subcultures to “deep fake” effects, we need the humanitarian technologies that develop group reflexivity, critical thinking, and culture of cooperation. The efforts are made to increase the level of data privacy, unbiased algorithms, safety of autonomous systems, equal access to the AI technologies, and social justice during automation, but the psychological mechanisms of ethically sound AI coding and using are to be explored. The further research is needed that address socio-psychological, ethical, and cultural aspects of AI adoption by contributors and users of digital platforms.

© 2019 Published by Future Academy www.FutureAcademy.org.UK

Keywords: Artificial intelligence, disinformation, information-psychological influence, echo chamber, fake news.



1. Introduction

Analysis of political and socio-psychological problems of information and psychological impact (IPI) is an urgent task of scientific research and an important object for study (Sosnin, Kitova, Nestik, & Yurevich, 2017) not only in Russia but also in other countries. There are many scientific problems that are closely related to information security and information and psychological warfare: purposeful and unintended impact on the emotional state of a person (Emel'yanova, 2016), the reaction of users to the post-event discourse (Pavlova & Grebenschikova, 2017), the attitudes to information threats and individual socio-psychological characteristics (Mikheev & Nestik, 2018), the impact of disinformation on interpersonal communication (Krasnikov, 2006) and defensive group mechanisms (Nestik, 2014).

2. Problem Statement

We assume that the using of AI is connected with different socio-psychological, ethical, and cultural aspects. The social perception of AI threats is not related to the technology itself, but to its use by people – employers, hackers, the state.

2.1. The place of AI technologies in the system of information and psychological warfare

The Information and psychological warfare (IPW) is a system of information and psychological effects on the information resources of the enemy, the consciousness and feelings of its soldiers and the population, as well as a set of measures to protect their own information and psychological resources (mil.ru).

The objects of the IPW are the population, armies and governments of warring, friendly and neutral countries. The sphere of its competence can be control systems, communication channels and electronic communications, databases and data banks, mass media, as well as consciousness and mentality of people.

The essence of the IPW as a specific form of modern warfare can be represented as a war with the use of information technology and psychological methods of influence, i.e. a kind of "information and psychological weapons". And the role of information in such a war becomes global: it is both a weapon, a resource and a goal. The psychological component plays the role of a kind of channel (method) or means of delivery, which determines the effectiveness of the creation, transmission, reception, perception, processing, storage and use of this information.

One of the promising ways to improve the efficiency of counteraction to "information and psychological weapons" is the development and implementation of artificial intelligence technologies. Obviously, the AI expands the ability to collect and analyze data, search for sources of information and mask important information from the enemy (Losev, 2017). At the same time, AI is considered as a tool of counteracting fake news, protecting civil society from the information warfare.

AI can have a significant impact on the content, nature and intensity of information dissemination in the mass media, both traditional media (TV, Newspapers, radio) and new media (social networks, blogs). AI is also able to generate harmful psychological content, including misinformation and promotion of the so-called "fake news". These forms are already actively used and pose a threat to the whole of humanity and the national security of sovereign states. According to Edelman Trust Barometer,

73% of respondents were concerned about fake news in 2019 alone (An Urgent Desire for Change, 2019). Many people (78% of respondents) believe in the need to increase the level of funding for research of artificial intelligence technologies for the development of QMS (Newman, 2019). Psychological studies conducted in Russia by T.Nestik have shown that the social perception of AI threats is not related to the technology itself, but to its use by people – employers, hackers, the state, etc. The core of social representation about the development of artificial intelligence includes unemployment, intellectual and spiritual degradation of people, AI's getting out of control and the world captured by machines, total control and invasion of privacy, irresponsible use of AI by people, and the use of AI as a weapon of war. Perceived AI threats are not so much connected with the reliability and predictability of the technology itself, as with its use by the state and other people. This indicates that the introduction of AI in everyday life will exacerbate the attention of citizens to cases of injustice and discrimination (Nestik & Zhuravlev, 2018).

2.2. Ways and methods of creating and promoting «deep fakes»

Currently, the mass media is already actively using machine learning technology based on the AI methodology. An example is the complexes and systems of analysis of Big Data, allowing to monitor the social media, the development of socio-political processes and make sophisticated models with a high probability of predictions.

Artificial intelligence techniques, such as generative-adversarial network (GAN), are used to create fake digital content – the so-called “deep fakes”. At SCAC, opposing neural networks work together to create the most realistic fake content from audio, video, and images (Memes That Kill: The Future Of Information Warfare, 2018).

In fact, one neural network in the GAN acts as an irritant that forces another network to find more accurate solutions. The network analyzes and corrects its decisions until it receives the most realistic video or image, the content of which has never been in reality. Neural networks also make it easier to create fake sounds. Neural networks can convert elements of a sound source into statistical data, and this data can be used to create original fake audio clips.

An experiment to create a digital "videofake" was successfully implemented in 2017 at Stanford University. In the course of it, scientists used a neural network to transform the audio file into the base points of the mouth of the "digital fake", as well as to train and synchronize the points of the mouth of the original and new video images. Some programs to create a "fake" video today are already in the public domain and anyone can use them. Specially trained neural networks allow you to create fictional video content indistinguishable from the present. Fake digital twins of politicians or important for a particular person people may pronounce the given text, to appeal that never would their real-life counterparts (Suwajanakorn, Seitz, & Kemelmacher-Shlizerman, 2017). It is now possible to control these video devices in real time (Thies, Zollhöfer, Stamminger, Theobalt, & Nießner, 2016). The use of AI allows to transfer information warfare in a fully automated mode, when neural networks themselves download metadata "targets" and analyze their psychological profile on digital traces in search of vulnerabilities, generate artificial video content based on these psychological profiles, organize an army of bots for its stuffing into social networks, target messages for those users who are most likely to send this information

to their friends, and then they conduct an automated assessment of the destructive impact of the information campaign on the society of the enemy country (Memes That Kill: The Future Of Information Warfare, 2018).

One of the main channels of distribution of "deep fake" are platforms. In this regard, there is a growing interest in the regulation of not only mass media sites, official mass media accounts in social networks, but also these platforms (news aggregators), which are based on algorithmic recommendations of content created by users and mass media (Newman, 2019). Such platforms do not belong to any type of mass media and their status, mode of operation and, accordingly, ways of controlling them are not fully clear. This uncertainty has contributed to the mass promotion across platforms and the destructive "fake" content, "astroturfing", creating panic and anxiety in society, frustration and uncertainty.

One of the trends that also contributes to the promotion of destructive content is the development of online video opportunities on the basis of social platforms. This is due, among other things, to the idea of Facebook to create a new television, characterized by greater sociality and interactivity. In the future, the platforms may crowd out YouTube and become the main channel for promoting "fake" audiovisual content.

Information and psychological impacts of "deep fakes" can result in damaging psychological safety of the individual, increase the person's vulnerability to manipulation, and undermine social trust, individual resilience, as well as person's relations to the world and to herself (Mikheev & Nestik, 2018).

3. Research Questions

It is necessary to understand the socio-psychological, ethical, and cultural aspects of AI adoption by contributors and users of digital platforms. Research tasks included:

- understanding the place of AI technologies in the system of information and psychological warfare;
- analysis of the ways and methods of creating, promoting "digital audio, video fakes";
- analysis of psychological consequences of AI-generated "deep fakes";
- analysis of the counteraction to "digital" fakes.

4. Purpose of the Study

The purposes of this study are:

- to demonstrate that the social perception of AI threats is not related to the technology itself, but to its use by people – employers, hackers, the state;
- to justify through the analysis of scientific literature and media materials the need to find solutions to social-psychological problems of using AI;
- demonstrate the opportunities and limits of AI as a tool of counteracting psychological warfare.

5. Research Methods

In this methodological study the following methods were used:

- philosophical and structural-functional analysis;
- interdisciplinary and comparative analysis.

6. Findings

We propose to raise the level of awareness of the population about the socio-psychological problems associated with the use of AI, methods of protection and counteraction. The main provisions of this program are developed by us in a series of publications (Nestik & Zhuravlev, 2018, Mikheev & Nestik, 2018, and they are as follows.

6.1. From technical aspects of AI use to psychological issues

The use of AI is associated with three areas in which the security of the individual, society and the state is ensured: digital, physical and political (Brundage et al., 2018). In the digital sphere, AI is used to improve the effectiveness of cyberattacks, including those that require large resources (targeted phishing). Also, AI can be used to search for human vulnerabilities (speech synthesis), software (automatic hacking), AI systems (providing false data). In the physical sphere the use of AI is due to conduct drone attacks and other physical systems (e.g., stand-alone deployment, such as combat or complex automatic control "swarm" microdrones). In the political sphere, AI can be used to solve the problems of monitoring (analysis of social processes), persuasion ("digital" propaganda), misinformation or deception (creation of "digital audio, video fakes").

In research on military capabilities of the AI involved many structures of the military and the U.S. intelligence community, in particular the Management of advanced studies of the Ministry of defense (DARPA), the research laboratory of the air force (AFOSR), Research laboratory for Land forces (ARL), Institute of behavioural and social Sciences of Land forces (ARI), Management of research work of the naval forces (ONR). National laboratories think tanks and universities are also doing a great deal of work. The most significant project in the defense Ministry, testing AI technology in military Affairs on a permanent basis, is the unit for the conduct of "algorithmic warfare" (Project Maven). It was established on April 26, 2017 under the leadership of the above-mentioned R. Wark in order to accelerate the testing of machine learning and other AI technologies in the activities of the national armed forces. In the spring of 2018 in America launched the process of establishing the joint Artificial Intelligence Center (Joint Artificial Intelligence Center), which consolidates the efforts of the national military community to develop in the field of AI. Currently, in addition to the individual initiatives in this area, in one form or another such technology is already integrated in about 600 programs of the Ministry of defence (Vilovatykh, 2019).

The development of digital technologies, including the full-scale introduction of machine learning into everyday life, deepens the cultural gap between those who are ready for uncertainty and choice, and those who are trying to avoid having to choose something. Artificial intelligence enables an individual to shift responsibility for his actions to an impersonal algorithm and its developers. This is already happening in the field of targeted Internet advertising and news, where content personalization puts a person in the "bubble" of her own interests, eliminating the need to search for information on her own.

The use of “deep fake” aggravates these psychological effects by undermining social trust and triggering defensive reactions to uncertainty of the world. At the individual level of socio-psychological analysis AI-generated “deep fakes” intensify the biases of overconfidence and reduce cognitive complexity, force the person to look for strengthening her self-esteem and defending her positive identity by placing her-self in a personal information bubble. At the group level the digital fakes increase the conformity, closure effect and polarization. The efforts of digital communities to protect themselves against the “deep fakes” lead to the construction of the so-called «echo chambers». At the societal level they destroy trust to social institutions, support the populism and restrictive policies motivated by collective fears.

6.2. Counteraction to «digital» fakes

Counteraction to «fake content» today includes identification, blocking, expert commenting, bringing to administrative or criminal responsibility of distributors. At the same time, a characteristic feature of the fight against misinformation in social networks in 2019 is the shift of focus to closed networks and communities, i.e. «echo chambers» (Newman, 2019).

In recognition of "digital fakes" using "Google Earth" search engine "Wolfram Alpha". They find cross-references to the pages with the image of the terrain and containing information about weather conditions, which allows to compare the elements of the environment (weather and climatic conditions) captured on the analyzed video with the real situation.

Eulerian Video Magnification technology is also used to recognize «fake video» created with the help of artificial intelligence. It is based on deep image detail and recognition of the smallest details, such as the presence or absence of heart rate in humans, changes in skin color due to blood flow, etc.

Many international organizations identify and monitor information generated, disseminated and promoted through AI. For example, the Institute of electrical and electronics engineering (Institute of Electrical and Electronic Engineers (IEEE)), which is working to increase the level of data privacy, transparency, Autonomous systems, ethical issues of the use of robotic systems, assessment of the reliability of AI and Autonomous systems. In the structure of the International organization for standards (international Standards Organization (ISO)), a Department has been created that deals with the issues of safety and reliability of AI (Millar et al., 2018).

However, the AI itself is insufficient to protect internet-users' psychological safety against “deep fakes”. By itself, it will only strengthen negative psychological defense mechanisms. To reduce the exposure of online communities and subcultures to “deep fake” effects, we need the humanitarian technologies that develop group reflexivity, critical thinking, and culture of cooperation.

7. Conclusion

Thus, today the technology of artificial intelligence is actively introduced into the system of elements of information and psychological warfare. This fact indicates that the further research is needed to address socio-psychological, ethical, and cultural aspects of AI adoption by contributors and users of digital platforms.

Acknowledgments

The research is supported by a grant of the Russian Science Foundation (project № 18-18-00439).

References

- An Urgent Desire for Change (2019). *Edelman Trust Barometer. Executive Summary*, 6. Retrieved from https://www.edelman.com/sites/g/files/aatuss191/files/2019-01/2019_Edelman_Trust_Barometer_Executive_Summary.pdf
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinke B., ... & Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Retrieved from <http://ecai.raai.org/lib/exe/fetch.php?media=malicioususeofai.pdf>
- Emel'yanova, T.P. (2016). The phenomenon of collective feelings in the psychology of large social groups. *Sotsial'naya i ekonomicheskaya psikhologiya*, 1, 3—22. Retrieved from <http://soc-econom-psychology.ru/cntnt/bloks/dop-menu/archive/g16/t1-1/s16-1-01.html>
- Krasnikov, M.A. (2006). *Regulatory function of misinformation in the process of interpersonal communication*. Moscow: IP RAS.
- Losev, A. (2017). Military artificial intelligence. *Arsenal otechestva*, 6(32). Retrieved from <http://arsenal-otechestva.ru/article/990-voennyj-iskusstvennyj-intellekt>
- Memes That Kill: The Future Of Information Warfare (2018). *CB Insights*. Retrieved from <https://www.cbinsights.com/research/future-of-information-warfare/>
- Mikheev, E.A., & Nestik, T.A. (2018). Disinformation in social networks: current state and perspective research directions. *Social Psychology and Society*, 9(2), 5–20. <http://dx.doi:10.17759/sps.2018090201>
- Millar, J., Barron, B., Hori, K., Finlay, R., Kotsuki, K., & Kerr, L. (2018). Accountability in AI. In *G7 Multistakeholder Conference on Artificial Intelligence* (pp. 8-10). Canada, Montreal: CIFAR.
- Nestik, T.A. (2014). Group reflexivity as a factor of relationship formation to the collective past. In *Social psychology of time* (pp. 264-291). Moscow: IP RAS. Retrieved from <http://spkurdyumov.ru/uploads/2016/03/socialnaya-psixologiya-vremeni.pdf>
- Nestik, T.A., & Zhuravlev, A.L. (2018). Collective emotions and misinformation in the era of global risks. In T.P. Emel'yanova, & E.A. Mikheev (Eds.), *Psychology of global risks*, (pp. 120-137). Moscow: IP RAS. Retrieved from <http://globalrisks.ru/engine/documents/document99.pdf>
- Newman, N. (2019). Journalism, Media and Technology Trends and Predictions. In *Digital News Project*. (pp. 9-19). Reuters Institute. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Newman_Predictions_2019_FINAL_2.pdf
- Pavlova, N.D., & Grebenshchikova, T.A. (2017). Intent analysis of post-event discourse on the Internet. *Psikhologicheskie Issledovaniya*, 10(52), 8. Retrieved from <http://psystudy.ru>.
- Sosnin, V.A., Kitova, D.A., Nestik, T.A., & Yurevich, A.V. (2017). Mass consciousness and behavior as objects of research in social psychology. *Sotsial'naya i ekonomicheskaya psikhologiya*, 4(8), 71-105. Retrieved from <http://soc-econom-psychology.ru>.
- Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). *Synthesizing Obama: Learning Lip Sync from Audio*. SIGGRAPH. USA, Seattle: University of Washington. Retrieved from <http://grail.cs.washington.edu/projects/AudioToObama>.
- Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2Face: Real-time Face Capture and Reenactment of RGB Videos. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2387–2395). Retrieved from <http://www.niessnerlab.org/projects/thies2016face.html>.
- Vilovatykh, A.V. (2019). Artificial intelligence as a factor of military policy of the future. *Problemy natsional'noi strategii*, 1(52), 177- 192. Retrieved from <https://riss.ru/bookstore/journal/2019-g/problemy-natsionalnoj-strategii-1-52/>