Artificial Intelligence Technologies Using in Social Engineering Attacks

Natalia Ryabchuk[1][0000-0003-1531-4398], Nina Grishko[1] [0000-0001-7227-157X], Vladislav Grishko[2] [0000-0002-9616-4506], Andriy Rudenko[2] [0000-0001-2585-6598], Valentyn Petryk[3] [0000-0003-2301-0722] ,Ideyat Bapiyev[3] [0000-0001-8468-8938] and Solomia Fedushko[0000-0001-7548-5856]

[1] College of Engineering and Management of the National Aviation University, Kyiv, Ukraine
[2] National Aviation University, Kyiv, Ukraine
[3] National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
[4] Satbayev University, Almaty, Kazakhstan
[5] Lviv Polytechnic National University, S. Bandera str. 12, 79013 Lviv, Ukraine

natali_push@ukr.net, solomiia.s.fedushko@lpnu.ua

Abstract. This article aims to familiarize readers with the concept of social engineering. The main areas of work of attackers in the field of social engineering are also listed and reviewed. The methods of psychological influence on a person are considered, the basic techniques used by attackers are described. The article considers the concept of generative-competitive neural networks, describes a short history of their creation. The article also explores the concept of deepfake. Academic and amateur developments related to deepfake are reviewed. The principles of the operation of generative-competitive neural networks are considered. The article also outlines the main threats that cybercriminals can pose using computer technology and knowledge of the psychology of human behavior. In conclusion, the main methods of dealing with intruders are presented.

Keywords: social engineering, areas of social engineering, methods of psychological impact on a person, generative-competitive neural network, deepfake.

1 Definition of Social Engineering

Social engineering is a set of various psychological techniques and fraudulent methods, the purpose of which is to obtain confidential information about a person by fraud. Confidential information is usernames / passwords, personal intimate data, incriminating evidence, bank card numbers and anything that can cause financial or reputation losses [1]. The concept of "social engineering" came from the field of hacking. Typically, hackers search for vulnerabilities in computer systems to obtain various information and personal benefits.

Computers work according to certain laws. Knowing them, hackers find points of entry into the system. But in any computer system, the most vulnerable place is the user, i.e. person [2]. Using the experience accumulated by mankind in psychology, manipulation and mechanisms of influence, hackers began to “crack people.”

In social engineering there are many different directions. The most famous areas are:

1. Carding (Fig. 1). This type of fraud includes various actions and fraud with bank cards, details, etc.

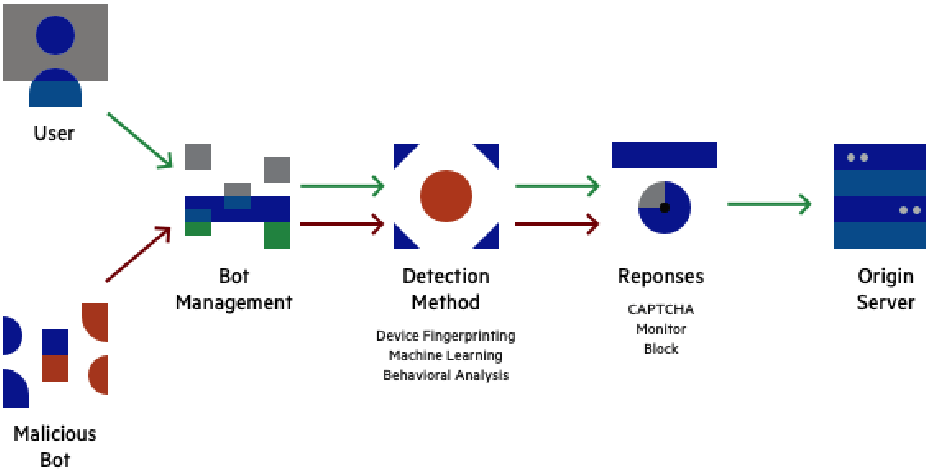

Fig. 1. Carding scheme

2. Phishing (Fig. 2). The main essence of this method is capturing your usernames and passwords from important sites, accounts, bank accounts, etc. by sending emails with malicious emails, or by sending you to fraudulent sites.

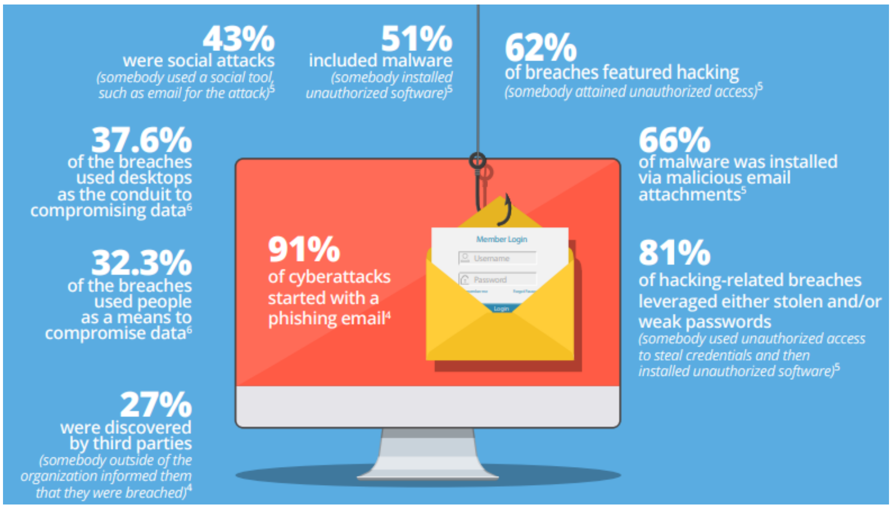

Fig. 2. Phishing statistics

3. Farming. This type of fraud is more dangerous than phishing, although it is directly related to it. This is a covert user redirection to a false IP address.

4. Hacking social networks (Fig. 3). This method gives attackers access to the personal information of users.

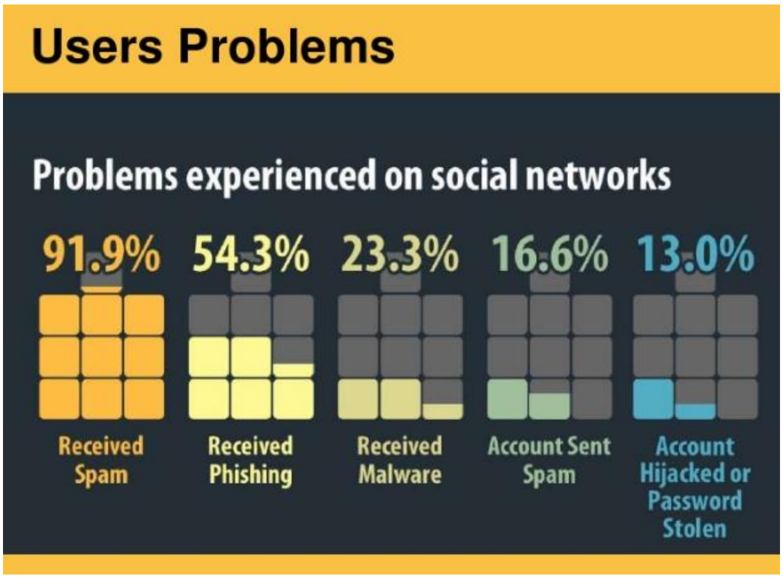

Fig. 3. Social networks security problems

5. SMS attacks (Fig. 4). At the moment, this method is dangerous for operations with bank cards.

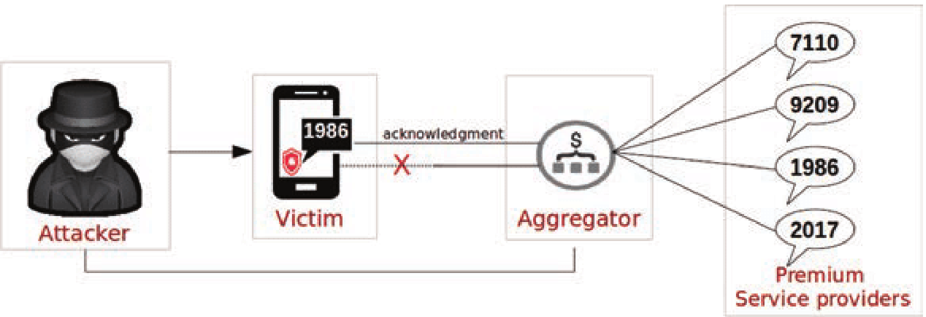

Fig. 4. SMS attack scheme

There are many different ways in the arsenal of professional attackers. Usually all of these methods are mixed if the hacker’s target is single. If the attack is massive, then the individual approach with the personal participation of the attacker is somewhat fading into the background.

To destroy the reputation of a famous person, a well-known enterprise, preparation for a raider seizure, etc. It is often necessary for attackers to use a variety of psychological tricks. Among them are the following [3]:

1. Mechanisms of influence.

The first and most important tool that most suits the needs of fraudsters is the psychology of influence and an understanding of the mechanisms that allow us to find gaps in people's beliefs and values. The most fundamental work on this subject is called The Psychology of Influence, authored by Robert Cialdini.

In the psychology of influence, one fundamental rule is distinguished: there are basic mechanisms of influence, and ALL people in the world are subject to at least one or two of them. It all depends on how developed your critical thinking is and what your level of intelligence is. The higher these indicators, the lower the chance of “getting hit,” but if you run into a competent performer, everything becomes more complicated.

The peculiarity of influence mechanisms for social engineering is that they are good for use both individually and for influencing the masses of people.

2. Managing the emotional state of a person

One of the main skills that any professional fraudster possesses is managing human emotions. There is a golden rule in negotiations and in any dialogue: the one who can cause any emotion from the interlocutor will always manage the conversation. Attackers are well aware of this rule and, unfortunately, actively use it. The main tool in this topic is provocation.

3. Psychology of the masses and sociology

This knowledge is very well correlated with the mechanisms of influence, because many of them can be used and applied en masse. Sociology for a hacker is an excellent hint and database, analyzing which, he can understand the most massive "vulnerabilities" of people up to specific regions.

4. Profiling

If in the previous paragraph we spoke more about mass character, then such a tool as profiling (drawing up a person’s profile - a psychological portrait) is a purely individually-oriented tool. This is an extremely complex science and requires great qualifications and knowledge.

Recently a new kind of deepfake scam has spread. Deepfakes (Fig. 5) can also be used to create fake news and malicious hoaxes.

Deepfake (“deep learning” and “fake”) is a method of synthesizing images of people based on artificial intelligence. It is used to combine and overlay existing images and videos onto source images or videos using a machine learning technique known as the adversarial network.

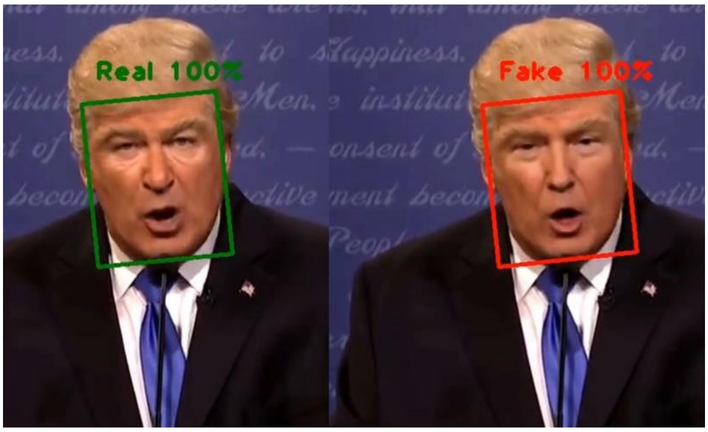

Fig. 5. Example of deepfake realization

To a large extent, deepfake developed under two conditions: research in academic institutions and amateur development in online communities.

2 History of GOS (Generative-Adversarial Neural Networks)

Adversarial machine learning has a different application than generative modeling, and can be applied to models other than neural networks. The general idea of learning through competition between players dates back to at least 1959 thanks to the influential work of Arthur Samuel, demonstrating that algorithms can learn to play checkers through competitive self-reproduction [4].

Ian Goodfellow is recognized by several sources as the inventor of GOS in 2014. This document included the first working implementation of a generative model based on adversarial networks, as well as a theoretical analysis of the game, establishing that the method is reliable. However, there is some independent debate about who invented which aspects of the GOS (Fig. 6).

Some peer-reviewed articles on GOS from academic sources, for example, attribute the idea to Ian Goodfellow and do not mention Jürgen Schmidhuber. Ian Goodfellow's GSN article in its own peer-reviewed article mentions an uncontrolled Schmidhuber technique called predictability minimization (PM), which states that PM is not a minimax game. Schmidhuber disputes this claim. He also articulates GOS as special cases of his Adversarial Curiosity (1990). Schmidhuber interrupted the presentation of Goodfellow in 2016, demanding more respect for his previous work. Goodfellow himself claims that the direct source of inspiration for the GOS was a comparative noise estimate, which uses the same loss function as the GOS, and which Goodfellow studied during his Ph.D. dissertation in 2010-2014.

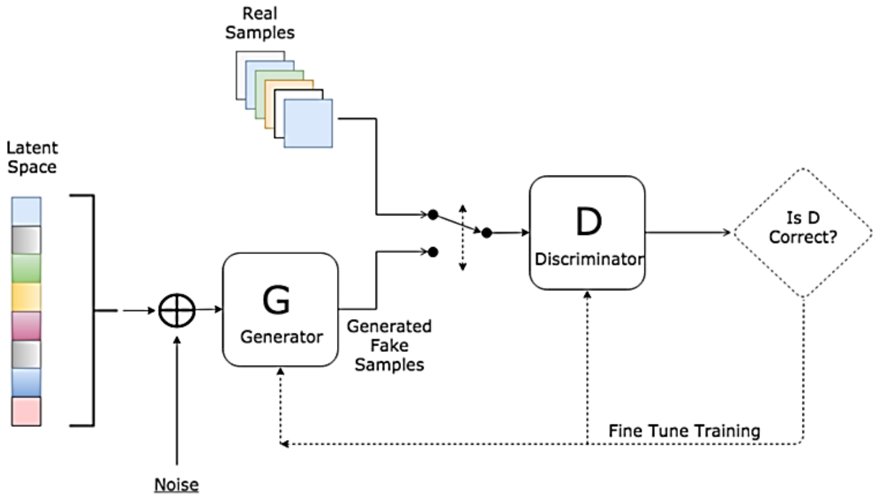

Fig. 6. General adversarial network

Other people had similar ideas, but they did not develop them in this way. The idea of warring networks was published on Olli Nimitalo's blog in 2010. This idea was never implemented and did not include stochasticity in the generator, and thus was not a generative model. This is now known as conditional GOS or cGOS. An idea similar to GOS was used to model the behavior of animals by Lee, Gauchi, and Gross in 2013.

In 2017, GOS was used to improve the image, focusing on realistic textures rather than pixel accuracy, providing higher image quality at high magnification. In 2017, the first persons were created. They were exhibited in February 2018 at the Grand Palais. Persons created by StyleGAN in 2019 are compared to deepfake.

Starting in 2017, GOS technology began to appear in the arena of fine art with the advent of a newly developed implementation, which was said to have crossed the threshold of ability to create unique and attractive abstract paintings and therefore was called “KSN” (creative - adversarial network). The GOS system was used to create a picture of Edmond de Belami in 2018, which was sold for $ 432,500. At the beginning of 2019, members of the original SPF team discussed further progress in this system, and also examined the general prospects of art with artificial intelligence.

In May 2019, researchers from Samsung demonstrated a GOS-based system that produces a video of a person talking to only one photograph of that person.

In August 2019, a large data set was created consisting of 12,197 MIDI songs, each with paired lyrics and melody alignment, to generate a neural melody from lyrics using the conditional GAN-LSTM (see Sources in GitHub AI Melody Generation from Lyrics).

3 Academic developments related to deepfake

Scientific research related to deepfake [5-8] lies primarily in the field of computer vision, a computer science unit, often based on artificial intelligence, which focuses on computer processing of digital images and video.

The first landmark project was the Video Rewrite program, published in 1997, which modified existing videos of a talking person to portray that person speaking words contained in another audio track. This was the first system that fully automated this type of facial resuscitation, and she did this using machine learning methods to establish a connection between the sounds created by the subject of the video and the shape of their faces.

Modern academic projects have focused on creating more realistic videos and creating simpler, faster, and more affordable methods. In Obama’s Synthesis program, published in 2017, the videos of former President Barack Obama have been modified to depict how he pronounces words contained in a separate audio track. As a main research contribution, the project lists its photorealistic technique for synthesizing mouth shapes from audio.

The Face2Face program, published in 2016, changes the video recording of a person’s face to depict them in real time, imitating the facial expressions of another person. As the main research contribution, the project lists the first method for reproducing facial expressions in real time using a camera that does not record depth, which allows this technique to be performed using conventional consumer cameras.

4 Amateur deepfake related

The term deepfakes originated around the end of 2017 from a Reddit user named “deepfakes”. He, like other members of the Reddit r / deepfakes community, shared the deep fakes they created. Many videos were the faces of celebrities attached to the bodies of porn actresses, while non-pornographic materials included many videos with the face of actor Nicolas Cage included in various films.

In December 2017, Samantha Cole published an article on r / deepfakes in Vice, which first drew widespread attention to deepfake, which are distributed in the online community. Six weeks later, Cole wrote in the next article about a large increase in the number of fake porn videos created using artificial intelligence.

In February 2018, r / deepfakes was banned by Reddit for distributing involuntary pornography; other websites also banned the use of deepfakes for involuntary pornography, including the social media platform Twitter and porn site PornHub. Other online communities have remained, including Reddit communities that do not disseminate pornography, such as r / SFWdeepfakes (short for safe for work deepfakes), in which community members share deepfakes featuring celebrities, politicians, and other non-pornographic people scenarios. Other online communities continue to share pornography on platforms that do not prohibit deepfake pornography.

5 The principle of the GOS

Two neural networks compete with each other in a game (in terms of game theory, often, but not always, in the form of a zero-sum game). With a training set, this method is trained to generate new data with the same statistics as the training set. For example, a GOS trained in photographs can generate new photographs that look at least outwardly authentic to human observers and have many realistic characteristics. Although GOS was originally proposed as a form of a generative model for non-teacher learning, they were also useful for supervised learning, fully supervised learning, and reinforced learning.

The generative network generates candidates, and the discriminatory network evaluates them. The competition works in terms of data dissemination. As a rule, the generating network learns to display the data distribution of interest from hidden space, while the discriminating network distinguishes the candidates issued by the generator from the true data distribution. The task of the learning generative network is to increase the frequency of errors in the discriminatory network (that is, to "deceive" the discriminator network by creating new candidates that, according to the discriminator, are not synthesized (they are part of the true distribution of data)).

The well-known data set serves as the initial training data for the discriminator. Training involves presenting samples from a set of training data to achieve acceptable accuracy. The generator trains depending on whether it manages to trick the discriminator. Usually a random input is introduced into the generator, which is selected from a predetermined hidden space (for example, a multidimensional normal distribution). After that, the candidates synthesized by the generator are evaluated by the discriminator. Backpropagation is used on both networks, so the generator produces better images, while the discriminator becomes more skilled in marking synthetic images. The generator is usually a deconvolutional neural network, and the discriminator is a convolutional neural network.

Thus, to create a DeepFake video, you need to get a large number of different images of the victim, analyze them using GAN algorithms, and then use this special software to apply these faces to the video. In addition, you can add text to the video that will be "voiced" by the victim - the person in the video.

The main danger of DeepFake technology is that tools that allow you to do all of the above have become available to ordinary people. Such applications require powerful computers to work, but today, in the era of powerful cryptocurrency mining servers installed in ordinary apartments, this is not a problem.

New York Times journalist Kevin Roose tested FakeApp and described his experiences. To create the video, he had to use a remote server rented through the Google Cloud Platform, which provided sufficient computing power. But even in this case, the remote server generated models for more than 8 hours. Renting a server cost $ 85.96. On a regular laptop, this task can take days or even weeks. In the experiment, more than 400 photos of the author of the article and more than a thousand photos of his double (actor Ryan Gosling) were used for the first DeepFake video. Although the video was very blurry

The experiment of the journalist was also interesting in that both his participant, user Reddit, and the developer FakeApp, who commented on the application, remained anonymous - they did not give their real names and communicated using aliases and impersonal email addresses.

We are very close to living in a world in which fakes really become virtually indistinguishable from each other. DeepFakes make fakes so realistic that our hearing and eyesight cannot tell a lie from a truth.

The danger of DeepFake lies not only in political propaganda and competition of the highest level of cynicism, in which you can force your opponent to speak outright lies, which the majority will believe. DeepFake is also a very dangerous tool for manipulating public opinion, with which you can easily sow panic and fear.

However, at the moment there are already some methods that allow you to identify DeepFake, and recently tools for identifying them have appeared.

Developers from Norwegian University of Technology concluded that face-switching technology, also known as DeepFake, could be used to good effect. They suggested using it to preserve the anonymity of people when creating videos and photos. Today, for this face, people are usually blurred or covered with a black rectangle, which, among other things, disrupts the distribution of data in the image and complicates their subsequent processing by algorithms.

As reported, to demonstrate the viability of the idea, the developers created a solution called DeepPrivacy, based on several neural networks popular in the field of image processing. First, the original image with a person or several people is fed into the S3FD neural network, which marks rectangular areas with faces in the image. The Mask R-CNN neural network then marks the key points for the detected faces: eyes, ears, shoulders and nose. Calculation of face parameters is necessary so that subsequently the superimposed face has a realistic position relative to the body, repeating the original.

After the R-CNN mask has calculated the key points, these parameters are encoded into a small image. Then the areas with faces on the original frame become gray, and then the pixels in it change to random colors. After that, such an anonymous image that does not allow to uniquely identify a person is fed into the U-Net neural network generator. During the operation of the neural network on one of the layers, the image created from the key points of the face is added to the main image, which allows the algorithm to accurately and realistic place a new face in the frame.

Thanks to this scheme, the researchers indicate that the generative neural network responsible for synthesizing a region with a face does not receive the original face, which allows to bypass the European General Data Protection Regulation (GDPR).

To train DeepPrivacy, developers created their own Flickr Diverse Faces (FDF) dataset, based on the YFCC-100M image dataset. The new data set includes 1.47 million images containing the faces of people, and they were shot under normal conditions and located at different angles to the camera, and sometimes partially covered by other objects. Each face in the photos is highlighted with a rectangular stroke, key points are also highlighted on it. The authors took 17 days to train the neural network model.

6 Conclusions

The purpose of this article was to introduce social engineering and systematization of techniques used by attackers, including new software tools for controlling the psychology of the masses. In order not to become victims of cybercriminals, you need to use anti-virus programs, do not trust social networks with a lot of personal information, create complex passwords, don't follow suspicious links, don't trust mailing lists, conduct training on basic information security rules at enterprises and, most importantly, with healthy cynicism to treat the "news" received from the network.

References

1. Gartner: site. - URL: http://www.gartner.com/newsroom/id/565125
2. Michele Fincher's presentation in SE Village at DEF CON 23: - URL: http://www.social-engineer.org/resources/82015-defcon23-se-village-chris-hadnagy/
3. Richard Feynman. Joy of knowledge (The Pleasure of Finding Things Out). 2013 . 348 p.
4. Popov A.A. Ergonomics of user interfaces in information systems. 2016.312 p.
5. S. Gnatyuk, Critical Aviation Information Systems Cybersecurity, Meeting Security Challenges Through Data Analytics and Decision Support, NATO Science for Peace and Security Series, D: Information and Communication Security. IOS Press Ebooks, Vol.47, №3, pp. 308-316, 2016.
6. Gnatyuk S., Akhmetova J., Sydorenko V., Polishchuk Yu., Petryk V. Quantitative Evaluation Method for Mass Media Manipulative Influence on Public Opinion, CEUR Workshop Proceedings, Vol. 2362, pp. 71-83, 2019.
7. A. Peleschyshyn, T. Klynina, S. Gnatyuk, Legal Mechanism of Counteracting Information Aggression in Social Networks: from Theory to Practice, CEUR Workshop Proceedings, 2019, Vol. 2392, pp. 111-121.
8. S. Gnatyuk, M. Aleksander, P. Vorona, Yu. Polishchuk, J. Akhmetova, Network-centric Approach to Destructive Manipulative Influence Evaluation in Social Media, CEUR Workshop Proceedings, Vol. 2392, pp. 273-285, 2019.
9. M. Brundage et al., The malicious use of artificial intelligence: Forecasting prevention and mitigation, arXiv:1802.07228, Feb. 2018.
10. Molodetska K., Brodskiy Yu., Fedushko S. Model of Assessment of Information-Psychological Influence in Social Networking Services Based on Information Insurance. CEUR Workshop Proceedings. Vol 2616: Proceedings of the 2nd International Workshop on Control, Optimisation and Analytical Processing of Social Networks (COAPSN-2020), Lviv, Ukraine, May 21, 2020. p.187-198. http://ceur-ws.org/Vol-2616/paper16.pdf
11. Davydova I., Marina O., Slianyk A., Syerov Y. Social Networks in Developing the Internet Strategy for Libraries in Ukraine. CEUR Workshop Proceedings. 2019. Vol 2392: Proceedings of the 1st International Workshop on Control, Optimisation and Analytical Processing of Social Networks, COAPSN-2019. P. 122–133.
12. Tkachenko, R., Izonin, I.: Model and Principles for the Implementation of Neural-Like Structures based on Geometric Data Transformations. In: Hu, Z.B., Petoukhov, S., (eds) Advances in Computer Science for Engineering and Education. ICCSEEA2018. Advances in Intelligent Systems and Computing. Springer, Cham, vol.754, pp.578-587, 2019. https://doi.org/10.1007/978-3-319-91008-6_58