ENGINEER

international scientific journal

ISSUE 3, 2025 Vol. 3
E-ISSN
3030-3893
ISSN
3060-5172

SLIB.UZ
Scientific library of Uzbekistan

A bridge between science and innovation

TOSHKENT DAVLAT TRANSPORT UNIVERSITETI
Tashkent state transport university

ENGINEER

A bridge between science and innovation

E-ISSN: 3030-3893

ISSN: 3060-5172

VOLUME 3, ISSUE 3

SEPTEMBER, 2025

engineer.tstu.uz

TASHKENT STATE TRANSPORT UNIVERSITY

ENGINEER

INTERNATIONAL SCIENTIFIC JOURNAL
VOLUME 3, ISSUE 3 SEPTEMBER, 2025

Founder of the international scientific journal "Engineer" – Tashkent State Transport University, 100167, Republic of Uzbekistan, Tashkent, Temiryo'lchilar str., 1, office: 465, e-mail: publication@tstu.uz.

The "Engineer" publishes the most significant results of scientific and applied research carried out in universities of transport profile, as well as other higher educational institutions, research institutes, and centers of the Republic of Uzbekistan and foreign countries.

The journal is published 4 times a year and contains publications in the following main areas:

- Engineering;
- General Engineering;
- Aerospace Engineering;
- Automotive Engineering;
- Civil and Structural Engineering;
- Computational Mechanics;
- Control and Systems Engineering;
- Electrical and Electronic Engineering;
- Industrial and Manufacturing Engineering;
- Mechanical Engineering;
- Mechanics of Materials;
- Safety, Risk, Reliability and Quality;
- Media Technology;
- Building and Construction;
- Architecture.

Tashkent State Transport University had the opportunity to publish the international scientific journal "Engineer" based on the **Certificate No. 1183** of the Information and Mass Communications Agency under the Administration of the President of the Republic of Uzbekistan. **E-ISSN: 3030-3893, ISSN: 3060-5172.** Articles in the journal are published in English language.

Cyber attacks using Artificial Intelligence systems

U. Begimov[1][a], T. Buriboev[1][b]
[1]Alfraganus university, Tashkent, Uzbekistan

Abstract: This article discusses one aspect of the use of Artificial Intelligence in cybersecurity. It is about cyberattacks that can be carried out using Artificial Intelligence (AI) systems. AI- enabled cyberattacks can be defined as any hacking operation that relies on the use of AI mechanisms. Another term used is offensive AI. AI-based cyberattacks are undoubtedly changing the cybersecurity landscape. First of all, it is necessary to talk about the speed of implementation of attacks and their scaling. AI-based cyberattacks involve the use of advanced machine learning algorithms to identify vulnerabilities, predict patterns, and exploit weaknesses. Efficiency and rapid data analysis enhance the ability of hackers to gain a tactical advantage, resulting in rapid intrusions or destruction of data. Traditional cybersecurity methods are no longer sufficient to combat sophisticated attacks, as AI-enabled cyberattacks adapt and evolve in real time. In addition, the introduction of AI systems in cyber defense gives rise to new risks. AI systems themselves become targets of adversarial attacks.The article discusses general issues of organizing cyber attacks using AI, provides taxonomy and examples of such attacks.

Keywords: machine learning, deep learning, cybersecurity, cyber attacks

1. Introduction

The use of Artificial Intelligence (AI) in cybersecurity has several aspects. Following the gradation proposed by Microsoft, the following areas can be distinguished:

- AI in cyberattacks (offensive or attacking AI);
- AI in defense against cyberattacks. The most well-known area of application today with the largest number of examples of use
- Cybersecurity of AI systems themselves (attacks on AI systems). The most actively developing area
- Malicious influences (e.g. deepfakes)

The article was received on June 12, 2024. D.E. Namiot - Lomonosov Moscow State University As usual.

AI systems are understood as machine learning models. In this article, we would like to focus on the use of AI in cyberattacks. Obviously, due to the specifics, not everything in this area is published. But it is also obvious that this area has received a great additional impetus for development with the growing popularity of large language models. The idea that it is possible, for example, to automate programming immediately prompts interested parties to think about automating the creation of malware, the ability of generative models to create “human” tests gives rise to thoughts about phishing, etc. Since a very rapidly developing industry is considered, the course materials are revised annually. The current version (2024) was created with the support of the Cybersecurity Department of Sberbank PJSC. The rest of the article is structured as follows. Section II discusses general provisions. Section III is devoted to the taxonomy of offensive AI. Section IV considers an example of solving a captcha. And Section V contains the conclusion.

2. Research methodology

Procedure for paper submission. In the era of artificial intelligence, attackers are using AI-based techniques to hack cyber defense programs. These AI-based cyber attacks are undoubtedly changing the cyber security landscape. First of all, it is necessary to talk about the speed of execution of attacks and their scalability. AI-based cyber attacks involve the use of advanced machine learning algorithms to identify vulnerabilities, predict patterns, and exploit weaknesses. Efficiency and rapid data analysis enhance the ability of hackers to gain a tactical advantage, resulting in rapid intrusions or destruction of data. Traditional cyber security methods are no longer sufficient to combat sophisticated attacks, as AI-based cyber attacks adapt and evolve in real time.

The traditional defense scheme for IT organizations in the early 2000s included perimeter protection and malware concerns. Organizations during those periods also focused on software security, but since software applications were minimal, methods to protect against external attacks were the priority. Later, software applications emerged to help solve user-based performance issues, and organizations built advanced perimeter defense devices such as intelligent firewalls, routers, and switches to counter external network attacks.Software and hardware attacks can pose a constant threat to businesses. However, there are effective ways to counter these threats. One such way is to use a system dependency model. This model combines predictive analysis, response time, attack type, containment, and cyber defense into a single system rather than treating them as separate entities. The system dependency model helps predict attack patterns and counter intrusions, especially for SOC (Security operations center) personnel. Each team member has an advantage due to the visual indicators and threat data provided by network security devices. However, AI-enabled cyber attacks require SOC personnel to re-evaluate their cyber defense strategy. Today’s situation is different because AI- driven cyberattacks are software-driven and adapt to configuration changes. No cyberdefender can resist the real-time changes, analysis, and adaptability of AI-driven attacks. Because AI platforms use machine learning to identify network behavior patterns and vulnerable targets, they can adapt and change their attack method.

Artificial intelligence (machine and deep learning) is increasingly used in cyber defense. At the same time, all

[a] https://orcid.org/0000-0002-6983-6709

[b] https://orcid.org/0009-0001-6700-4095

such defense tools can be targets of adversarial attacks. Such attacks involve modifications of data at different stages of the machine learning pipeline, are relatively easy to implement, and, in most cases, cannot be completely excluded. Accordingly, poisoning attacks, backdoors, and, of course, evasion attacks, which concern AI-based defense tools, are typical applications of AI (machine learning) in cyber attacks. NIST in its taxonomy of adversarial attacks separately considers adversarial attacks in the field of cybersecurity. Historically, the first adversarial attacks began in this domain. The first known poisoning attack was developed to generate worm signatures back in 2006.The aforementioned work considered systems that automatically determine signatures (signs) of software worms. That is, in fact, rules for malware signatures. The attack proposed by the authors polluted (noised) the traffic used by automatic signature generators during their extraction. The attack was aimed at misleading signature generation algorithms by introducing well-designed noise that prevented the generation of useful signatures. It was shown that it is possible to introduce noise without prior knowledge of the classification technique used. The use of artificial intelligence brings its own risks that differ from those traditionally considered in cybersecurity. There are many different classifications regarding this, one of which is given in. There are 14 risks of AI listed:

1. Lack of transparency and explainability of AI
2. Job loss due to AI automation
3. Social manipulation by AI algorithms
4. Oversight functions performed by AI technology
5. Lack of data privacy when using AI tools
6. Bias due to AI
7. Socioeconomic inequality as a result of AI
8. Weakening of ethics due to AI
9. Autonomous weapons based on AI
10. Financial crises caused by AI algorithms
11. Loss of human influence
12. Uncontrolled AI
13. Increased criminal activity
14. Wider economic and political instability

Lack of privacy is perhaps one of the most serious problems, which can also be relatively easily exploited through adversarial attacks targeting IP [10]. Vulnerability mitigation programs need to be changed, but there are also issues of classification. Imagine a data breach on an AI platform. Although the risk is software-based, should it be classified as a software risk or an AI-based risk? The largest collection of AI risks is contained in the MIT project: AI Risk repository. Its description is in.

In addition to adaptability and real-time analysis, AI-based cyberattacks can also cause more disruptions during a small time window. This is due to the way the incident response team works. When AI-based attacks occur, it is possible to bypass or hide traffic patterns (changing the system log analysis process or removing data that allows for defensive actions). Cybersecurity systems will need other algorithms that identify AI-based cyberattacks.

AI has created problems in which security algorithms must become, first of all, predictive and fast and accu- rate. The traditional IT landscape contains many risks related to privacy, perimeter protection, software applica- tions or data leakage. These risks create loopholes and weaken the organization's defensive posture. Counter- measure tactics are to eliminate risks and improve the level of cybersecurity. The introduction of AI into the risk and vulnerability ecosystem is transforming security compliance and cyber defense. As AI leverages behavioral analytics, machine learning, and real-time analysis, busi- nesses must learn about risks based on patterns and com- putational errors. This is where continuous monitoring and AI will work best. Organizations must also determine how IT system audits, risk assessments, etc., configura- tion changes and remediation deadlines should evolve.

Cybersecurity transformation also requires the development and implementation of controls. Typical frameworks such as NIST 800-53 or OWASP are structured based on application, cloud, data, identity and infrastructure. An open question is whether AI should be implemented in the same control frameworks or whether current controls should be modified? This, among other things, will de- termine the attack surfaces of AI.

Taxonomy of offensive AI. AI-enabled cyberattacks can be defined as any hacking operation that relies on the use of AI mechanisms. Another term used is offensive AI.Everything in scientific papers begins with some classification. Let us note right away that cyberattacks are a rather sensitive area, so not everything is openly published. However, Figure 1 shows one possible classification of AI attacks:

Fig. 1. AI attacks

This list obviously lacks adversarial attacks on machine learning models, which are widely used in information systems, cyber-physical systems, and Internet of Things systems. Other comments include the following: Keyboard sniffing is part of a more general problem called side-channel attacks, where AI is widely used. Phishing, in principle, can be classified as a social engineering attack. Deepfakes also include voice cloning. Separating voice cloning into a separate category is possible if we are talk- ing about biometric authentication, for example. This is traditionally separated from deepfakes. Classically, deepfake (from deep learning + fake) was originally understood as as a method of synthesizing an image or voice that imitates a person and is based on artificial intelligence. Deepfake technologies can also be used to create fake news and any malicious deception. Deepfakes are usually singled out as a separate area of using AI in cybersecurity, and they are considered in this paper.

Despite these remarks, at least this list gives an idea of what AI attacks are. Of the elements omitted in this classification, it would be worth adding automation of attacks. In our opinion, this is a separate area of using AI in cyberattacks. For example, the so-called AI-driven pentesting. Examples of such automation of pentesting are, for example, startups XBOW and RunSybil.

Figure 2, which is taken from a highly cited work, provides a classification of attacks described in the scientific literature by types of impact. The attacks here are distributed

across six stages of the cybersecurity chain (kill chain). Fig. 2. Attacks by stages of the kill chain

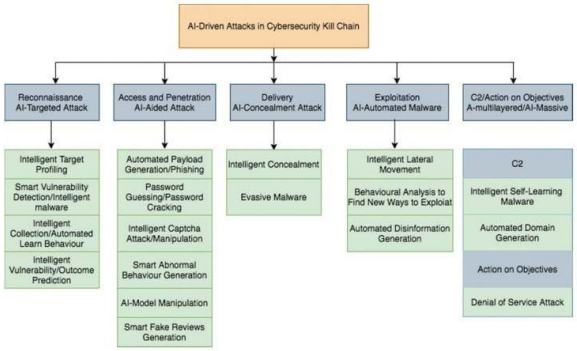

Fig. 2. AI attacks

Six types of AI-driven attacks were identified in the access and penetration stage (AI-assisted attack), four types of AI-driven attacks were identified in the access reconnaissance stage (AI-targeted attack), three types of AI-driven attacks were identified in the exploitation stage (AI-automated attack), and two types of AI- driven attacks were identified in the delivery stage (AI-concealment) and C2 stage (Command Control) respectively. In contrast, one type of AI-driven attack was identified in the targeting stage (AI-malware). The access and penetration stage had the most publications (6), followed by the reconnaissance stage (4), the exploitation stage had three publications, and the delivery and C2 stages had two. In contrast, the target-based phase (AI malware) had the fewest publications (1). Fig. 3. LLM in attack.

Year	MITRE Tactic(s)	Application	Model(s)
2023	Execution	Generating code to perform actions that could be malicious	GPT-3
2022	Initial Access	Generate phishing emails to bypass spam filters	GPT-2, GPT-3, RoBERTa
2022	Execution - Command & Control	Use of LLMs as plug-ins to act as a proxy	GPT-4
2023	Initial Access - Collection	Generate Phishing Website via ChatGBT	GPT-3.5 Turbo
2023	Execution	Code generation and DLL injection	GPT-3
2023	Initial Access - Reconnaissance	Collecting victim data to develop an attack email	GPT-3.5, GPT-4
2023	Initial Access - Execution - Defense Evasion	Crafting malicious scripts	GPT-3.5 Turbo, GPT-4, text-davinci-003
2018	Initial Access	Spear Phishing link	AWD-LSTM
2023	Defense Evasion	Code obfuscation, file format modification	GPT-3.5
2023	Initial Access - Credential Access	Password guessing using LLMs	GPT-2
2023	Initial Access - Reconnaissance	Impersonation for phishing aims	GPT-3.5 Turbo
2022	Initial Access	Generating content for misinformation	GPT-2

Fig. 3. AI attacks

Potential attackers place great hopes on LLMs (which are, accordingly, of great concern to the cybersecurity community) to automate attacks. Here are examples of LLMs being used in cyberattacks (as of early 2024) in killing chain mitigation

It should be noted that such lists will constantly grow. This process is absolutely natural. If we want to teach LLM to write code, then the idea that it could be malicious code or some kind of data encryptor arises almost automatically. If we demonstrate the capabilities of the same LLM to write selling marketing offers, then it is easy to guess that the text for phishing mailings will not be much different. And so on.

Automation (democratization - lowering the entry threshold and reducing costs) is a natural process. The same, accordingly, applies to protection: there is simply no other way out. Attacking robots must be met by the same robots-defenders

Solving captcha. A large number of works are devoted to such tasks. Ob- jectively, image recognition is one of the most traditional tasks for machine (deep learning). Examples of works. How it looks, we will analyze using the example of work. The work describes an attack on text captchas (text recognition in a picture). Examples of such tasks are shown in Figure 4. The length of the line in characters and the modifications made are indicated Fig.4.

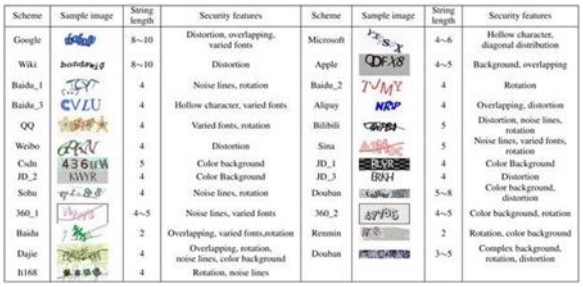

Fig. 4. AI attacks

Text captchas As shown in Figure 5, the attack consists of 3 steps. Fig.5. Attack pattern

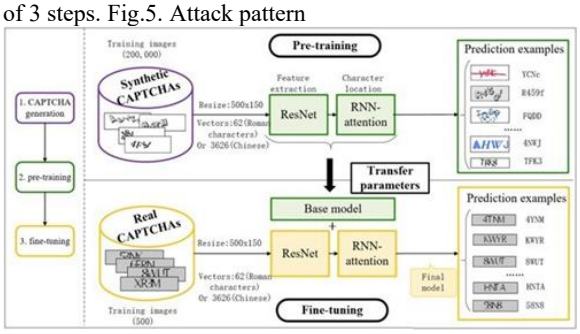

Fig. 5 AI attacks

Step 1: CAPTCHA Generation: This step uses image processing algorithms to generate CAPTCHAs unrelated to the target scheme to train our recognition network. In the attack under consideration, all pre- samples are generated completely randomly without any special design, which is easy to implement and significantly reduces the effort spent on collecting training samples.

Step 2: Pre-training: Once generated, the synthetic CAPTCHAs are fed directly into the recognition engine without any pre-processing to train the base model. After pre-training, we adopt the trained model as the base model for all subsequent schemes.

Step 3: Fine-tuning: Finally, for each scheme, 500 real samples were used to fine-tune the base model. This step was accomplished by re-training the base model using transfer learning to update the parameters to match the real features. Note that only domain-specific adaptation of transfer learning was used and the model remained consistent across the pre-training and re-training steps. Basic architectural decisions:

To reduce the cost associated with manual labeling, synthetic CAPTCHAs were generated as pre-training data for the pre-training. All training data for the baseline model is generated using simple image processing algorithms from the Pillow library.

As shown in Fig. 6, all pre-training samples are generated with black characters on a pure white background. Unlike the original CAPTCHAs, there are no security features in the generated CAPTCHAs: for example, there are no noise lines, distortions, overlays, etc. Instead, the samples were generated in the simplest way to reduce the generation cost, since this type of CAPTCHA is easy to implement and does not require much effort. The gener- ated CAPTCHAs are completely unrelated to the target CAPTCHAs (Fig. 4) and do not resemble any of the target schemes.

For the Latin character-based schemes, the text string length is set in the range from 4 to 10; The fonts are randomly selected from the font library, including both

https://doi.org/10.56143/3030-3893-2025-3-59-63 A bridge between science and innovation

regular and hollow styles; all images are the same size, and the text rotation angle is set from minus 45 to 45 degrees. For Chinese patterns, the line length was set to a range of 2 to 5. 500,000 images were generated to pre- train the base model. Fig. 6. Some examples of randomly generated CAPTCHAs for training the base model.

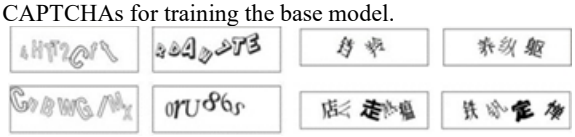

Figure 6: AI attacks

Like any other applied problem in machine learning, the main challenge is training data and feature engineering. All samples were of the same size 500 × 150.

2. In pre-training (base model), a combination of CNN and LSTM was used. In order to recognize the entire character string in one step, the combined model described in, consisting of a CNN and a Long Short-Term Memory (LSTM) model, was used as the recognition engine. The CNN is responsible for extracting the feature vector of the CAPTCHA image. For this, the ResNet v2-101 model was chosen, which is designed to address the degradation problem that occurs as the network depth increases. LSTM converts the feature vectors extracted by the CNN into a single text string; it can be thought of as a character-level language model. Decisions are made using the latest states in the memory cells.

In this experiment, the number of LSTM cells depends on the maximum string length of the target CAPTCHA.

3. In the last step (fine-tuning), transfer learning is used to fine-tune the parameters of the pre-trained modelwith several real CAPTCHAs. Transfer learning works as follows. In transfer learning, a domain D is denoted as D= X, P(X), where X is the feature space and P(X) is the marginal probability distribution. For a particular domain, the task can be defined as T = Y, f, where Y denotes the label space and f denotes the target predictor. In general, the complete transfer learning process includes one source domain (DS) and one target domain (DT), which correspond to one source task (TS) and one target task (TT), respectively. From the knowledge in DS and TS, transfer learning aims to improve the learning of the target predictor f in DT. In this CAPTCHA solver, f denotes the predictor in ResNet, and DS and DT are as follows:

As for the training data is a synthetic CAPTCHA, and yTi is the corresponding CAPTCHA label, a character string. Here have the same values as in real CAPTCHAs. Note that all labels remain the same in DS and DT (62 or 3626 characters), but the feature spaces are different because the features of the synthetic and real CAPTCHA have different details. For each Roman character-based scheme, 500 hand-labeled real samples were used. Considering that Chinese CAPTCHAs have a larger character set than Roman CAPTCHAs, 1000 real hand-labeled Chinese CAPTCHAs were used per Chinese scheme.

3. Conclusion

As a conclusion, we present the following 5 points that, according to the authors of, determine the future of offensive AI. They attribute this to generative AI and large language models (LLMs) trained to generate malicious content (e.g. FraudGPT). 1. Automated social engineering and phishing attacks LLMs like FraudGPT demonstrate the ability of generative AI to support convincing scenarios for pretexts that can mislead victims. One use case is for attackers to ask LLMs to write science fiction stories about how a successful social engineering or phishing strategy works, thus forcing the LLM itself to provide attack recommendations. Other use cases could be to request instructions for attacks in national languages, in which case security filters set to English may not work. 2. AI-generated malware and exploits. FraudGPT has proven its ability to generate malicious scripts and code tailored to a specific victim's network, endpoints, and broader IT environment. Novice attackers can quickly master the latest defenses by using AI-powered generative systems like FraudGPT to learn and then deploy attack scripts. This is why organizations must do everything they can to ensure cyber hygiene, including endpoint protection. AI-generated malware can bypass older cybersecurity systems that were not designed to detect and prevent this threat. 3. Automated asset discovery by cybercriminals.

Generative AI will reduce the time it takes to conduct manual research to find new vulnerabilities, find and collect compromised credentials, learn new hacking tools, and master the skills needed to launch sophisticated cybercrime campaigns. Attackers of all skill levels will use it to discover unprotected endpoints, attack unprotected threat surfaces, and launch attack campaigns based on information obtained through simple clues.

It is noted that along with identification, endpoints will be subject to more attacks. Self-healing endpoints are noted to be critical, especially in mixed IT and operational technology (OT) environments that rely on Internet of Things (IoT) sensors. A self-healing endpoint is a technology for automating the monitoring and diagnosis of performance and security issues across multiple network nodes or endpoints.

Traditional incident response often requires significant manual intervention to identify and remediate compro- mised systems. Self-healing endpoints, on the other hand, use AI and machine learning algorithms to automatically detect, isolate, and remediate security incidents without human intervention. These endpoints continuously moni- tor and analyze system behavior, enabling proactive threat detection and autonomous response, resulting in reduced response times and a lower chance of widespread compro- mise.

These endpoints can proactively detect anomalies and potential security threats by continuously monitoring their behavior and network communications. This proactive approach not only reduces the need for constant human intervention, but also helps detect and mitigate risks, strengthening the overall security posture.

4. AI-powered evasion is just getting started, and we haven't seen the real problems yet.Weaponized generative AI is still in its infancy, and FraudGPT is just the beginning. More sophisticated and lethal tools are emerging. They will use generative AI to evade endpoint detection and response systems, and create malware variants that can evade static signature detection.

5. Difficulty of Detection and Attribution. FraudGPT and future generative AI applications and tools will reduce the detection and attribution barrier to anonymity. Security teams will have a hard time attributing AI- enabled attacks to a specific threat group or campaign based on forensic artifacts or evidence. Greater anonymity and difficulty in detection will lead to longer dwell times and allow attackers

to perform long-term attacks that are typical of advanced persistent threat (APT) attacks on high-value targets. Weaponized generative AI will eventually make this possible for every attacker.

References

[1] Echocardiographic images via machine learning algo- rithmics // Analysis of world scientific views International Scientific Journal Vol 2 Issue 1 IF(Impact Factor)8.2 / 2023

[2] O'.I.Begimov, T.M.Bo'riboyev / General Theory About the Traditional Methods and Algorithms of Ma- chine Learning // AMERICAN Journal of Public Diplo- macy and International Studies Volume 02, Issue 04, 2024 ISSN (E):2993-2157.

[3] T.M.Bo'riboyev / Hetnet tizimi asosida avtonobillar- ing harakat trafigini boshqarish va tahlil qilish // Nejmet- tin, 03-06 Ekim 2023 tarihlerinde Erbakan Üniversitesi ve Alfraganus üniversitesi öncülüğünde düzenlenen "ipek Yol- unun Ötesinde kongreler dizisi: Bir Yol, Bir Kuşak: Göç, turizm ve ekonomi politik Kongresi (SIRCON 2023)" pro- gramına sertifika almak için katıldı. (Sayfa 320-324)

[4] NIST AI 100-2 E2023 Adversarial Machine Learn- ing: A Taxonomy and Terminology of Attacks and Mitiga- tions https://csrc.nist.gov/pubs/ai/100/2/e2023/final : 15.07.2024

[5] Perdisci, Roberto, et al. "Misleading worm signature generators using deliberate noise injection." 2006 IEEESymposium on Security and Privacy (SP'06). IEEE, 2006.

[6] 14 Risks and Dangers of Artificial Intelligence (AI) https://builtin.com/artificial-intelligence/risks-of-artificial- intelligence 15.08.2024.

[7] Song, Junzhe, and Dmitry Namiot. "A survey of the implementations of model inversion attacks." International Conference on Distributed Computer and Communication Networks. Cham: Springer Nature Switzerland, 2022.

[8] NIST SP 800-53 Rev. 5 Security and Privacy Controls for Information Systems and Organizations https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final 15.08.2024.

[9] OWASP https://owasp.org/ 15.08.2024.

Information about the author

Begimov Uktam Ibragimovich	Alfraganus University, PhD, associate professor of Faculty of Digital Technologies E-mail: uktam1985beg@mail.ru Tel.: +99894 669 63-18 https://orcid.org/0000-0002-6983-6709
Buriboev Tolibjon Mirali ugli	Alfraganus University, assistant teacher of Faculty of Digital Technologies E-mail: buriboevtolib@gamil.com Tel.: +99894 563 19-98 https://orcid.org/0009-0001-6700-4095

CONTEXT / MUNDARIJA