Duality of artificial intelligence technologies in assessing cyber security risk

N I Romancheva
The Moscow State Technical University of Civil Aviation, Kronshtadtsky blvd 20, 125499, Russia
E-mail n.Romancheva@mstuca.aero

Abstract. Issues of estimation of information technologies influence on information security were and remain issues of special importance. During the pandemic 2020 over one million effects on structures including CII were registered, where all of them were related to application of modern technologies including those based on artificial intelligence. Extensive penetration of intellectual intelligence systems and of "smart" (end-to-end) technologies urges for revision of approaches to cyber security risk assessment. Application of the latest technologies in cyber security risk assessment creates new options on the one hand, and represents new threats from another. Under cyber security risk assessment we shall understand an integral index of information safety achievement in terms that cannot be easily predicted, such as complicity, uncertainty, inability to foresee, lack of stability. On that basis this work aims to analyze influence of artificial intelligent technologies on information security indexes in risk assessment in terms of VUCA-world. This paper proposes an approach for creating a model of duality assessment for such technologies in cyber security e risk assessment with games theory applied.

Keywords: Duality; Artificial intelligence; Risks; Cyber security; Games theory.

1. Introduction
General trends in information security support in the state and public security are defined by the Information Security Doctrine of the Russian Federation adopted by the Decree of the President of the RF No. 6468 [1] dated 05 December 2016. The Action Plan for "information security" for the "Digital Economics of the Russian Federation" program was approved as well. Legal and regulatory issues of the cyber security are governed by several regulatory documents of the FSTEC of the RF. The Development Strategy [2] contains definitions of various modern information technologies including those applying artificial intelligence (AI). For regulatory and technical governance of various aspects of AI technology application, in July 2019 Russia established the technical committee 164 "Artificial Intelligence" (TC 164) that provided a set of standards related to AI, two of which were approved. Existing theoretical, practical and legal-and-ethic norms of artificial intelligence application in corruption counterfeiting are extensively covered both in works of foreign authors [3, 4], and national studies [5, 6, 7]. In the works of V.S. Ovchinskiy and E. S. Larin the artificial intelligence is considered as a three-way technology that may be applied for civil, military and criminal purposes [8]. The work [9] notes that cyber risks layout is rapidly changing: the speed of innovations implementation is changing together with the speed of new cyber risks arising, response procedures established in regards to customary approaches to IT-solutions development get more complicated. Cisco report [10]
underlines that "active research for threats should consist in revealing of the tactics, methods, and procedures (TMP) used by attackers. This is the most valuable asset, as the attackers are hard to change them." In some cases simple actions for active search for threats are based on studies or reports on recently revealed threats.

The work [10] considers a "pyramid of pain" method by David Bianco that describes approaches to information protection that strongly complicate network attacks. Each of six levels of the pyramid represents a different approach: from the simplest to the most complex. By moving upwards the pyramid, an attacker will use more resources on the network attack, that will complicate and increase a probability of being detected.

According to the same document, currently the common practice represents inclusion to similar studies of compromise indexes (CI) that could be used by other parties involved. These data points usually consist of IP-addresses, URL-addresses, hashtags of files, and other CI constituting a threat.

The term "cyber security" had different definitions for years. We will turn our attention to the term provided by JSC Kaspersky Lab: "this is set of methods and practices for protection against attacker’s attacks on computers, servers, mobile devices, digital systems, networks, and data" [11].

The risk management idea is usually based on detection of trouble spots in network or systems, establishment of its importance, appointment of priorities and further adoption of respective measures for their elimination. It may include detection of sources of threats, and active research of threats may aid in obtaining information for risks assessment. With above all potential risks are considered (known and unknown).

Management of cyber risks assessment as a rule is performed based on monitoring of indexes and analyzing of their alteration dynamics. The work [12] offers an approach for cyber risks assessment based on a company's business processes analysis: assessment of effects on business and risk analysis, development of a response strategy for the risks, development of a strategy implementation plan.

The main problem of the modern cyber risks assessment consists in absence of complete formal description of parameters of the technologies applied, and their mutual effect on the final cyber risks assessment, application of methods and means that do not consider resistance to cyber accidents, and negative effects mitigation.

Analyzing existing and tested forms of risks assessment does not fully reflect the changed context in regards to society digitalization. Consideration of duality of the modern technologies applied, and of artificial intelligence in particular, in generating the information security system enhances and has a qualitative effect on the cyber risk assessment. Therefore, assessment of the duality effect of artificial intelligence technologies on cyber security risks assessment represents a critical task.

2. **Informal statement of the problem**

The purpose of this work consists in analysis of the dual impact of artificial intelligence technologies on the quality of cyber security risks assessment in modern conditions, and in generating a model of cyber risks assessment and management in VUCA-conditions. The research problem could be divided as follows: classification of the existing artificial intelligence (AI) technologies, and their application in the information security sphere; definition of AI technologies duality, and their effect on errors of cyber security risks assessment; ensuring set parameters of cyber security when selecting one or another artificial intelligence technology.

Therefore, the following issues shall be resolved to solve the set problem: to generate sets of existing cyber security risks assessment; to establish a type of reflection of sets of existing AI technologies on sets of risks assessment models corresponding to VUCA-security concept. A feature of this problem solution represents a necessity to define conditions and assessment of a probability of applied AI technology getting in an "abnormality" sphere. The term "abnormality" sphere will be applied to a zone of adverse effects of the applied technology on the cyber security indexes.
3. Methods

3.1. Classification of Existing Cyber security Risks Assessments

Understanding that information technologies carry threat to information security has resulted in coordination of criteria and establishment of the International Standard "Common Criteria for IT Security Evaluation, ISO/IEC 15408, by several states.

First of all, when using the artificial intelligence technologies for information security, focus is given on predicative analytics to assess risks of one or another threat arising, and a possibility to prevent an attack or other destructive actions.

Cyber security counterfeits three types of threats: a cyber crime is an action arranged by one or several attackers aimed to attack a system for breaking its work or receiving financial benefit; a cyber attack is actions aimed to information collection, mainly of politic type; cyber terrorism means actions aimed to destabilization of digital systems aimed to create fear or panics.

Modern information systems and computer networks consist of different components and represent a complex system. An issue of ability to preserve properties available by its components arises at an every stage of a complex system life cycle. The system property to preserve peculiar properties of its components is called compositionality [13], that is also referred to security properties.

Factors which negatively influence a complex system state, may include risks related to violation of protection of infrastructure, services, resources, information leaking by technical channels, not authorized and unintended impact on data and (or) other resources applied in the digital sphere. Thus, for example, intellectual intelligence application in various areas of industry, production, on the one hand, allows to adapt in new social and economic conditions, on the other hand, risk of qualitative replacement or sudden 5G-loss of large volume of information increases. Growth of deepfake technologies application, when AI is used to create a contact with replacement of faces and voices of famous people.

The risk management practice outlines several risk analyzing methods: statistics, expert assessment, analytics, method of analogue application, and risk accumulation consequences analyzing. Use of the statistic risk analysis method in utilization of AI technologies is based on historic data analysis, that in terms of uncertainty may deface risk assessment in general. The expert assessment method could be used at the first step of risk level definition until a well-organized registry of cyber risks related to application of one or another AI technology and consequences of their negative affect is available. The analogue application method for analyzing risk levels may be useful for threats registry formation, and is based on collection of information for previous incidents received from various sources. Supervised classification algorithms are widely spread recently in practice. However, complicity to control the tree size does not allow describing precisely complex areas of one class of objects.

3.2. The Risks Assessment Model with Consideration of AI Technologies Duality

AI technology represents a single process including such stages as information resource creation, ensuring this resource resistance to internal and external influence during transfer and processing. Following this consideration, we will deal with AI technology as with a complex system with multiple functional elements and relations between them, segregated from the environment in accordance with the required purpose for a certain moment of time with ability to analyze, self-educate and preserve resistance to destabilizing factors. Analysis and self-education properties allow the above technologies to show duality and, therefore, have both positive and negative effect on the information object that is a digital recourse. We will consider models, methods, and facilities for generation and utilization of an intellectual resource in the society as technology elements.

It is possible to outline technologies having a analysis mechanism, that unifies data receipt, storage and enhancement using processing of natural language to extract data from structured and non-structured data using AI, and multiple technologies based on comparison of visual information with some predefined classes based on an independent components analysis algorithm.
Problem setting is formulated as follows: in arbitrary time intervals, based on consequence of events and connections between them, to assess level of influence of the i-th parameter of the j-th AI technology belonging to the set of applied technologies on assessment of cyber security risks in accordance with the constructed model.

We will introduce a notion of "AI distrust index" that will mean a level of process destructurization leading to cyber threats implementation. Destructurization may be related to unintentional actions of the technologies (logics errors in the code, e.g., not detected during testing stages), and intentional actions (e.g. application of self-education algorithms for generation of new trouble points and threats). The set of acceptable risks of solutions has an uncertain origin and depends on VUCA-parameters: volatility, uncertainty, complexity, ambiguity. Impact of AI technologies parameters on the final target function may be both positive and negative: Therefore, the set problem may be considered as a problem of achieving minimum damage to information resources and other correlated infrastructure system elements with applied AI technologies, i.e. absence of inverted maximum system protection result (1):

\[
\min \left( \sum_{i=1}^{k} \left( \sum_{j=1}^{n} p_{ij} - p_{ij}^0 \right) \right) \rightarrow \min ,
\]

where \( p_{ij}^0 \) is minimum specified damage of j-th resource with applied i-th technology, \( p_{ij} \) is actual damage of j-th resource at application of i-th technology.

We admit that if each AI technology element does not fall within the "abnormality area", the result of this technology application has no impact on the cyber security risks assessment.

The risk will be estimated through the likelihood of a risk event multiplied by the assessment of its consequences for the system operation. To assess such consequences, it is offered to rank AI technologies by the value of cyber security damage and by the value of measures related to consequences removal. At the first stage the problem of abnormality zone definition for each AI technology is defined by an expert method with further automation of this process.

Risk identification supposes a staged analysis of all technology elements involved in cyber security provision starting from the most to the least accessible for a threat. The cyber security system assessment model may be represented as an integrated process in the form of combination of interrelated processes. Initial risk assessment, assessment of inverted risk probability and counter-measures efficiency, as well as analysis of residual initial and inverted risks should be grounds for this approach. For this we will introduce three initial sets of processes (of risk map formation, vulnerability map formation, formation of protection measures) into consideration, based on which a set of assessment areas is formed. The risk map includes both typical system risks and inverted influence risks of the used technologies. The vulnerability map is formed from a pool of fixed and declared vulnerabilities.

The threat level is proposed to calculate by vulnerability (VL), based on the criticality and likelihood of threat implementation through this vulnerability. The threat level shows how the impact of this threat on one or another process performed by the system is critical with account of its implementation probability. The single threat mode uses equation (2):

\[
VL = \frac{K \cdot P(VL)}{100} ,
\]

where K - criticality of threat implementation (specified quantitatively, i.e. in points assigned by experts, or, where applicable, as difference between BS failure in result of threat implementation, and probability of BS failure without threat implementation); \( P(V) \) is probability of threat implementation through such vulnerability related to AI technology application.

Vulnerability threat level values range from zero to one. Currently the existing vulnerabilities are classified in accordance with [14] as multifactorial. Accordingly, equation (2) may be presented as (3):
where $\mathcal{G}_i$ is weighted index of i-th vulnerability.

Risk $R$ related to application of AI technology is proposed to calculate by equation (4):

$$R_T = \frac{1}{a} \cdot (VL_f \cdot D),$$

where: $a$ is AI distrust index; $VL_f$ is total level of threats; $D$ is resource criticality (is specified qualitatively (be expert) or as quantitatively (as difference between failure probability without AI technology application, and failure probability with this technology applied). We describe the protection measures formation process as a process of action program selection in terms of uncertainty and suppose the focus on an "expected value" of results of such actions. For its assessment mathematic expectations usually used.

Output data of risks assessment models represent binary forecasts for non-inclusion of applied technology $T$ to the abnormality area. It allows to compare the simulation results with a selected working area tolerance limit (to ensure the specified parameters of cyber security). The risks assessment quality grows with extension of planning horizon This allows to reduce the time of risk management model formation in VUCA-conditions. Minimization of structural arrangement of technology elements between risk zones with account of VUCA-parameters and the abnormality zone (Figure 1) is used as a selection criterion.

![Figure 1. Structural arrangement definition](image)

When assessing cyber security risks in conditions of uncertainty, it is required to specify VUCA working area, exceeding borders thereof shall be considered as a risk of exceeding the specified parameter and entering the abnormality zone.

Application of AI technologies as an instrument of analytics to assess cyber security risks is aimed to improve such processes as management processes. This includes placement and achievement of goals of long-term and short-term forecasting, transformation and detailed management of risks. Application of a balanced approach developed with consideration of priorities for protection from known and new threats allows to establish correlations between security properties and indexes. In terms of massive transition to remote work, a number of information channels requiring monitoring increases; in addition, there are events in data streams that need to be tracked, i.e. management of risks and their prediction is required. Threats analysis aimed to detection of dangerous actions is represented as a factor of readiness to threats including ones created in VUCA world. To ensure resilience to such threats, it is required to create conditions for restoration after cyber accidents, as well as negative consequences mitigation.
Therefore, applied advance technologies for proactive detection of cyber security threats should be balanced between the working parameters zone and VUCA - parameters zone. Process parameters entry to the abnormality zone indicates absence of flexibility properties in the applied AI technology, namely, effective measures to reduce risks and protection/recovery methods.

Ignoring a mutual effect of AI-based defenses and the threats created by the same technologies may lead to rather heavy consequences, increasing an attack space. Examples include chatbots that behave inappropriately, fake applications with malicious content embedded in components.

As it was mentioned earlier, the considered AI technologies consist of single elements, respectively; parameters of various elements of the technologies usually are not correlated. By the values of mathematical expectation of random variables $m_{yi}$, variance $\sigma_{yi}$ corresponding to the period t, we can calculate the probability $P_i$ that the values of the determining parameter $\sigma_{yi}$ in the i-th section of random process $T_{vuca}(t)$ when using AI technology are within the risk zone tolerance and do not fall into the abnormality zone.

Assessment of risks of response to destructive processes created by AI technology elements includes such common stages as analysis, identification, description, measurement, and quantitative/ qualitative assessment.

Duality of AI technologies impact appears in, among others, in growth and in difficulty in understanding and assessment of risks of information loss or misrepresentation. For example, AI technologies applied in interception systems, on the one hand, increase coverage of channels, assist in general data review; on the other hand, cause data overloads and make their interpretation difficult in terms of reliability. New technologies allow to indicate important changes in processes, to pay attention to potential or currently existing threats in an events flow. From the other hand, all recent attacks on systems occur based on machine education algorithms.

The risks assessment model normally includes separate groups of risks known and mathematically described. Nevertheless, statistic analysis in general cannot follow changes in parameters of information flows, and their environment.

Algorithms of machine education applied for cyber security risks assessment separate and perform search of repeating patterns in information processes by separating a set of hidden interrelations between parameters, thus, forming a zone of abnormal behaviour and destructive characteristics.

In the abnormal behavior zone it is possible to indicate profiles of threats and vulnerabilities both traditional and related to AI application. Analysis of a complicated chain of events (information process phases) is performed for this purpose, and symptoms and processes, which lead to negative consequences, are identified.

To assess threats with consideration of duality of AI information technologies we will draw a table of risks analysis, perform risk calculation, and select a risk management method.

Increase in cyber attacks using AI technologies is primary related to application of machine education systems with an open code that allows attackers to avoid security systems. Classification of instruments applied for cyber security analysis, and their impact on changes in information resource properties is given in Table 1. This table uses the following designations: C - breach of confidentiality, I - breach of integrity, A - breach of accessibility. Changes of destructurization functions for process conditions depending on applied AI technologies are highlighted with color: green color - reduction, pink color - increase, orange color - ambiguous change, white color - change not detected.
Table 1. Classification of AI Technology Instruments and Their Impact on Information Properties

| Instrument                          | Result                                                                 |
|-------------------------------------|------------------------------------------------------------------------|
| In-depth Traffic Analysis (NTA)     | Availability of specific comprometation indexes                       |
| Intellectual algorithms (X-Spider, Max-Patrol, PT Application Inspector) | Vulnerability and errors in web-applications                          |
| In-depth technologic traffic analysis (PT, ISIM) | Detection of attacks related to hidden tunnel effect and extraction of data from APCS Legal actions (development of attack vector) Activity inside network |
| IoT                                 | Breach of IS policies                                                  |

To assess information technologies with AI application, the games theory [15] will be used. Games theory application allows to simulate a conflict situations process (technologies duality) and take a decision based on the analysis result on practicability to apply one or another information technology. A game in a normal form is a triplet consisting of a set of players and their strategies, as well as player's payoff function, which assigns a payoff of this player to each set of strategies (situations). We will consider a case when a number of players is two. Attributes of AI technology will be considered as players having positive and, respectively, negative impact on results of its application. We will suppose that all players are rational, each player considers options available in his possession, forms notions in regards to unknown parameters, has clear preference, and choses own actions resulting from a certain optimization process. Players strategy is, respectively, minimization and maximization of the probability of destructive event occurrence risk. Winning means profit (based on attack vector development). We will employ a dominating strategy of player 2. To check that player 2 strategy dominates over player 1 strategy, it is necessary to consider behavior of these strategies against a certain standard. Consequent removal of dominating strategies of one of the players should lead to the balance. Choice of a unique balance from several equal outcomes is performed based on established criteria of options selection based on destructive functions analysis.

Input data for modelling are fixed-dimensionality attribute vectors, which contain numerical values, with support of a special variable value representing absence of the cyber security risk. A destructive event will be interpreted as a reverse spread of an error arising in the process leading to reduction or seizure of target function performance in result of reverse influence on the AI technology system.

Vectors of attributes of analyzed parameters falling into the abnormality zone are collected dynamically from results of processing of the risk assessment model parameters by the analytical module with further consideration of possibilities for risks implementation and developing measures for their mitigation. In result the module receives a possibility to define direction of AI technology adaptation to alteration of one or another parameter affecting cyber security risks.

4. Conclusion

- The paper provides fundamental principles of the offered approach to consideration of AI technologies duality.
- The set of existing cyber security risks assessments has been considered.
- The type of the set of existing AI technologies reflection to the set of risks assessment models has been established.
The validity of the offered approach at the development stage is confirmed by expert assessments of specialists in the cyber security sphere provided in the public press that will be supported for its further testing.

References
[1] The general trends in information security support in state and public security are defined by Informational Safety Doctrine of the Russian Federation adopted by the Decree of the President of the RF No. 6468 [1] dated 05 December 2016
[2] Strategy for Information Society Development in the Russian Federation for 2017-2030 approved by the Decree of the President of the RF of 09.05.2017 No. 203 URL: http://kremlin.ru/acts/bank/41919 (reference date - 27.06.2020)
[3] Aiello M F 2018 Policing through Social Networking: Testing the Linkage between Digital and Physical Police Practices The Police Journal 91 1 pp 89-101
[4] Webb A 2019 The Big Nine. How the Tech Titans & Their Thinking Machines Could Warp Humanity. Published March 5th 2019 by PublicAffairs (first published 2019).
[5] Borovskaya E V and Davydova N A 2018 Artificial Intelligence Fundamentals Knowledge Laboratory p 127
[6] Larina E S and Ovchinskyi V S 2008 Crime of the Future Is Already Here Book World p 512
[7] Larina E S and Ovchinskyi V S 2018 Artificial Intelligence. Big Data. Crime Book World p 416
[8] Ovchinskyi V S 2019 Digital World Criminology: Text Book for Magistrate Norm Printing Press p 352
[9] Cyber Risks Assessment Based on Company's Business Goals / JSC Deloitte&Touche CIS [Digital resource] URL: https://www2.deloitte.com/content/dam/Deloitte/ru/Documents/risk/russian/Ocenka-kiberriskov-na-osnove-biznes-celej-organizacii.pdf (reference date 02.07 2020)
[10] CISCO publications on information safety 2019 [Digital resource], URL: https://www.cisco.com/c/dam/global/ru_ru/assets/pdfs/cybersecurity-series-2019-threat-hunting-ru.pdf (reference date 02.07 2020)
[11] What Is Cyberdefense? [Digital resource] URL: https://www.kaspersky.ru/resource-center/definitions/what-is-cyber-security (reference date 02.07 2020)
[12] Romancheva N I 2019 Conditions of Information Technologies Application in Digital Sphere Transactions of International Conference. Innovation, Information Technologies and Communication Technologies - Sochi pp 136-141
[13] Kosachev A S and Ponomarenko V N 2004 Analysis of Approaches to Verification of Safety and Mobility Functions Triumph p 101
[14] GOST R 56546-2015 National Standard of the Russian Federation Information Protection. Vulnerabilities in Information Systems Classification of Information System Vulnerabilities. Information Protection. Vulnerabilities in Information Systems. The Classification of Vulnerabilities in Information Systems OKC 35.020 Date of entry 2016-04-01.
[15] Smith M J 1982 Evolution and the Theory of Games Cambridge University Press p 10