Formal Model of Trustworthy Artificial Intelligence Based on Standardization

Eduard Manziuk[a], Olexander Barmak[a], Iurii Krak[b,c], Olexander Mazurets[a] and Tetiana Skrypnyk[a]

[a] *Khmelnytskyi National University, Institutska str., 11, Khmelnytskyi, 29016, Ukraine*
[b] *Taras Shevchenko National University of Kyiv, Volodymyrska str., 60, Kyiv, 01601, Ukraine*
[c] *Glushkov Cybernetics Institute, Academician Glushkov Avenue, 40, Kyiv, 03187, Ukraine*

Abstract

The widespread and rapid distribution and application of artificial intelligence (AI) systems requires the development of formalized approaches and the construction of basic principles for the functioning of domain areas of AI use. This need is embodied in the development of recommendations and standards to obtain maximum benefits from the use of AI and minimize possible risks. The regulatory framework is being built on a human-centric basis. Accordingly, the developed standards should form the basis for further activities aimed at the use of AI and be applicable at all stages of creating practical solutions. Therefore, an important stage is the formalization of requirements, principles and provisions of legal and ethical norms in the form of practical template approaches for practical application. With this method, models and ontology of standardized concept of AI credibility are developed within the research. This made it possible to identify the main concepts that allow forming a position of trust, are a meaningful part of the concept of trustworthy AI, determine the need for its existence and pose a threat to it. On the basis of ontology of the domain area, models were developed and further decomposition of structural substantive concepts was carried out. In the future, the characteristics of the concept of trustworthiness formation are defined.

Keywords

Human-centric AI, ethic AI, ontology, model, trustworthiness, standardization AI.

1. Introduction

The development of artificial intelligence systems today is undergoing significant growth and is finding implementation in a wide range of practical tasks in various spheres of human life. The problems solved by artificial intelligence (AI) are quite diverse, for example, such as object recognition, languages, classification, clustering, etc. Prospects for the practical application of AI in the areas of automatic car management, diagnosis of diseases in medicine, in the field of finance, safety are achievable and everywhere activities are carried out on their implementation with the help of subject information technology.

2. Related Works

Thus, the approaches used in AI and can be summarized to it are widely used in the field of healthcare [1, 2], robotics and automated systems [3], classification and detective systems [4,5],

IntelITSIS'2021: 2nd International Workshop on Intelligent Information Technologies and Systems of Information Security, March 24–26, 2021, Khmelnytskyi, Ukraine

EMAIL: eduard.em.km@gmail.com (E. Manziuk); alexander.barmak@gmail.com (O. Barmak); krak@univ.kiev.ua (I. Krak); exechong@gmail.com (O. Mazurets); tkskripnik1970@gmail.com (T. Skrypnyk)
ORCID: 0000-0002-7310-2126 (E. Manziuk); 0000-0003-0739-9678 (O. Barmak); 0000-0002-8043-0785 (I. Krak); 0000-0002-8900-0650 (O. Mazurets); 0000-0002-8531-5348 (T. Skrypnyk)

telecommunication systems [6,7,8], cyber-security [9]. The scope of application is rapidly expanding and constantly taking new forms [10,11] and is embodied in new approaches [12,13], specific areas of use [14,15].

At the same time, the level of development of artificial intelligence systems indicates that AI can make decisions that are unfair, biased, false, and are also unsuspected and misleading. This is not significant under the conditions of application in tasks that have no signs of increased responsibility, for example, image improvement, optimization of strategies in games, selection of offers, etc. However, if decision-making has vital signs for human life and activity, there is a question of trustworthy AI decisions. This is due to various areas of human life [16-19]. Today technologies are being developed with the aim of integrating humans into the process of obtaining machine solutions [20]. In the end, there is a problem of trust in problems and solutions of critical importance. The lack of resolution of this issue greatly reduces the potential of widespread use. Rapid development of AI applications requires actualization of efforts towards formality regarding law, ethics, control, security of AI application [21]. The goal is to ensure the development and implementation of AI systems through a control system. Lack of control affects trustworthiness and is therefore perceived as a risk that limits the adaptation and use of such systems. Assessing the advantages and prospects of AI technologies, AI regulation is carried out both at the national levels of countries and within the framework of international cooperation [22-24].

3. Domain area concepts

The construction of artificial intelligence (AI) systems should be based on certain principles according to which it is necessary to ensure the implementation of meeting the needs and according to which this system is built. It is necessary to determine the value principles within which the project is implemented. Thus, clear links between abstract principles of value priorities are necessary initial and basic positions according to which applied concrete solutions will be developed. Formalization of these norms and principles is implemented by developing standards according to the relevant norms. The main idea is that compliance with norms and standards is the main framework on which the design of AI systems is implemented. According to this approach, the principles laid down in the norms will be the basis on which AI systems are built. This is the basis for the human-centric development of AI. The widespread use of AI should be due to trust in processes, data, results, decision robust and security.

As a result of active cooperation and fruitful work, a number of requirements and regulations to AI are being developed. It is necessary to develop, implement and use AI systems that meet the ethical, legal, security norms of man and society. That is, there is such a form of AI as human-centric AI. The transition to the practical field of use forms the conditions and the need to create human-oriented information technologies. It is based on the provision of fundamental human rights based on collective, social and constitutional principles, in which individual freedom and respect for human dignity are both practically possible and meaningful, and does not suggest an overly individualistic understanding of man [21]. This approach is the basis and should be implemented at the stages of development, implementation, use and monitoring of AI systems.

The need to develop the basic fundamental foundations of AI systems led to a significant number of scientific publications and the work of expert groups in order to formally present them. So the EU Commission's High-Level Expert Group on AI (AI HLEG) European AI Alliance in "Ethics Guidelines for Trustworthy AI" [21] defined the concept of "trustworthy AI" as the basis for the widespread use of AI based on a human-centric approach. To ensure trustworthiness, it is necessary to implement seven key requirements, which are determined on the basis of critical discourses regarding the content of the concept. Thus, the increase in the social need for the benefits of AI use, due to the increase in scientific publications and wide discourse, is formalized by the results of the work of expert groups. Subsequently, implemented in the form of appropriate standards. For example, several of them: ISO/IEC NP TR 24027 (Information technology - Artificial Intelligence (AI) - Bias in AI systems and AI aided decision making), ISO/IEC AWI TR 24368 (Information technology - Artificial intelligence - Overview of ethical and societal concerns), ISO/IEC NP TR 24030 (Information technology - Artificial Intelligence (AI) - Use cases) and others. One of the basic is ISO/IEC TR 24028 [25] (Information technology - Artificial Intelligence (AI) - Overview of trustworthiness in Artificial Intelligence, which aims to analyze the factors that shape and influence AI systems.

The ISO/IEC TR 24028 standard describes an overview of AI trust issues. The standard aims to analyze the factors that influence the formation of trustworthy AI and the decisions it generates. The document briefly presents well-known approaches today that allow increasing confidence in technical systems and potential uses in AI systems. There are also possible vulnerabilities and approaches that reduce their impact on AI credibility. The structure of the standard and certainty of concepts make it possible to form ontology of the concept of trustworthiness, which we will present in a simplified form. The basis for the creation of ontology is the focus on determining the concepts, ensuring the implementation of which in practical technical means and information technology using AI, will ensure the parameter of trustworthy AI solutions.

In this direction, it is necessary to focus efforts on the development of models, methods of improvement, information technologies and other areas that will give a significant advantage in the practical application of systems. This will enable and formulate conditions for the use of AI in responsible and critical areas.

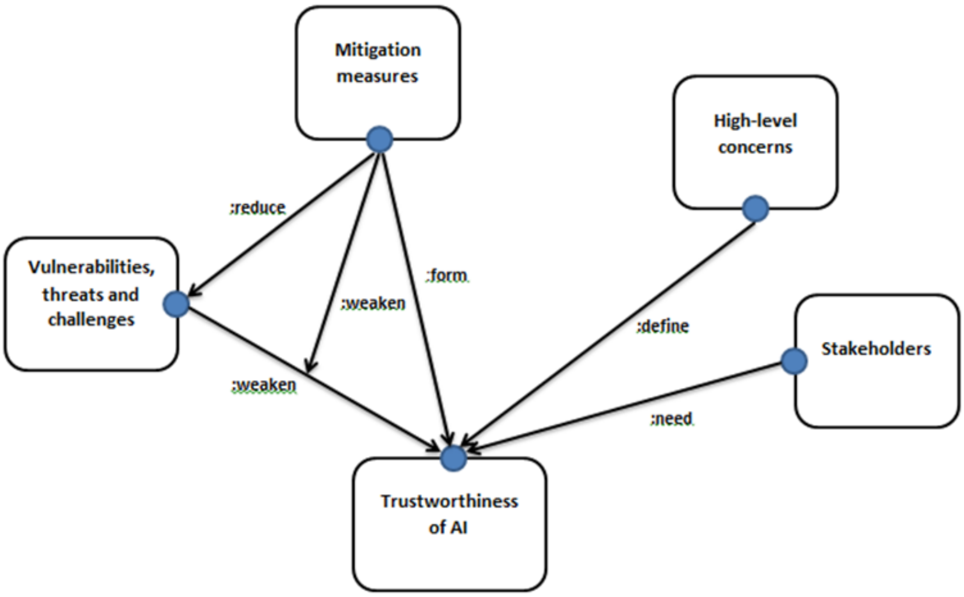

Figure 1: Ontology of trustworthy Artificial Intelligence standard ISO/IEC TR 24028 (simplified representation)

The central concept of the standard is trustworthy AI, which is defined as the ability of AI to meet the expectations of stakeholders, is verifiable and has the ability to implement such a check. The concept of threats, vulnerabilities and challenges is determined, the active influence of which on AI weakens the trust of stakeholders. This circumstance ultimately calls into question the need to use AI systems. The concept for preventing and reducing such an impact is determined, because such an impact is destructive and weakens the concept of trustworthiness and ultimately destroys it. The purpose of this concept is to reduce the magnitude of threats, vulnerabilities, challenges, and weaken their impact on trustworthiness. Thus, it is the concept whose structural elements allow forming trustworthiness. That is, the concept of "Mitigation measures" is a structural element for the formation of the concept "Trustworthiness of AI". The purpose of the "Mitigation measures" concept is to build trustworthiness, but it does not determine its content, that is, the basic concepts on which trustworthiness is based. To determine the content of trustworthiness, the standard establishes the concept of "High-level concerns". The relevant concept describes at the general level the main components of the content of trustworthiness. The presented attitudes of concepts that stakeholders rely on, expect from AI, hope to meet needs and invest in the content of trustworthy AI. At the same time, the purpose of this concept is a sprawling interpretation, understanding and representations that are formed on the expectations of stakeholders. The existence of the trustworthiness concept is conditioned and solely determined by the needs of stakeholders. It should be noted that one of the important concepts of the system is "Vulnerabilities, threats and challenges". It is this concept that causes, stimulates and forms the need of stakeholders for the need for trustworthiness. The importance of the existence of this concept is determined in terms of negative impact and difficulty in meeting their needs by stakeholders. In order for stakeholders to be able to use AI systems from a position of trust in them, that is, to meet the requirements of safety, control, sustainability of decisions and others (is an integral part of the "High-level concerns" concept), development, research and implementation at the level of practical information systems and technical solutions of the "Mitigation measures" concept components is necessary. The evolutionary development of this concept is the main factor in building trustworthiness and its constant increase in the direction of improving the satisfaction of stakeholder needs.

4. AI trustworthiness ontology model

The model is based on the standard defined by groups of characteristics that together determine the trust in decisions made using AI approaches. The Model "Information Technologies - Artificial Intelligence (AI)" (Overview of trustworthiness in Artificial Intelligence) is defined within the concepts of the relevant domain area. The standard defines high-level concepts that form the concept of trustworthiness. The set and concept of stakeholders as the main objects need in the concept of trustworthiness are also determined. Potential vulnerabilities in AI systems and the threats associated with them are identified in the concepts of vulnerabilities, threats, and challenges. To meet the needs of stakeholders in trustworthiness, the concept of mitigation measures is determined – possible controls, recommendations and guidelines that can reduce the impact of known AI vulnerabilities.

Model of the domain area of the standard ISO / IEC TR 24028 "Information Technologies - Artificial Intelligence (AI) - Overview of trustworthiness in Artificial Intelligence" based on ontology will be presented in the form.

$$SbDmTrw = \langle Con, Rel \rangle, \qquad (1)$$

where Con - a set of concepts defined within the standard;
Rel - a set of relationships between concepts.

The set of concepts (Fig. 1) is formed on the basis of standard information $Con = \{\{Vul, Thr, Chal\}, \{MtMs\}, \{HLConc\}, \{Sth\}, \{TrW\}\}$ and the set of relationships between them $Rel = \{"reduce", "weaken", "form", "define", "need"\}$.

Thus, the concept of "Mitigation measures" ($MtMs$) forms the trustworthiness of AI. Accordingly, the constituent elements of this concept form a set of characteristics and determine the functional relationship of trustworthiness $TrW = f(MtMs)$. Trustworthiness (TrW) according to the standard is a function of the main categories (phrase categories). The structure of "Mitigation measures" ($MtMs$) is as follows

$$MtMs = \{Ct, subCt, Ch\}, \qquad (2)$$

where Ch - characteristics, Ct - subcategories, $subCt$ - categories. The elements of the set are formed in a hierarchy $Ch \subseteq subCt \subseteq Ct$ and are functionally related to the attributes of trustworthiness $\langle Ct, subCt, Ch \rangle = f(\{atr_i\}_{i=1}^{n})$. Phrase categories are sometimes defined by a set of components and the form of representation is defined as follows $Ct \equiv \{subct_i\}_{i=1}^{n}$ or $Ct \equiv \{ch\}_{i=1}^{n}$. Each category is a set of characteristics, as combined by a standard according to semantic meaning, and forming semantic clusters. There are 10 categories, many of which in the general case can be represented $\{\cdot, \{\cdot\}\}$, due to the representation of some categories at the level of subcategories or characteristics. To form uniformity and convenience, we will present the category in the form of a category of the group that is defined $\{\{\cdot\}\}$. An example of representing categories from the standpoint of homogeneity: {{ Transparency },{ Reliability, Resilience, Robustness },…}. In this case, the homogeneity of the structure at the level of categories is formed. The model of the domain area of the standard is also built on this principle. Given the hierarchical structures of concept formation in the standard, the most convenient is to obtain a higher hierarchy of concepts, which can be represented as a homogeneous structure of phrase categories. The analysis showed that within the standard it is the level of subcategories. List of concept subcategories $MtMs$: Transparency, Explainability, Controllability, strategies for reducing Bias, Privacy, Reliability, Resilience, Robustness, mitigating system hardware Faults, functional Safety, Testing, Evaluation, Use, Applicability. There are many mitigation measures at the subcategory level $MtMs = \{mtms_i\}_{i=1}^{n}, n = 14$. There n - number of phrasal subcategories of the concept. The set of relations between the components of the concept $MtMs$ consists of the relation "*isPartOf*", ie. $Rel_{MtMs} = \{"isPartOf"\}$. Then the concept model $MtMs$ will take the form

$$M_{MtMs} = \begin{Bmatrix} Trs, Exp, Cont, Bsred, Prv, Rlb, Rsl, Rbs, Flmit, \\ Sffun, Tst, Evl, Us, Apl, "isPartOf" \end{Bmatrix}, \quad (3)$$

where Trs - transparency, Exp - explainability, $Cont$ - controllability, $Bsred$ - strategies for reducing bias, Prv - privacy, Rlb - reliability, Rsl - resilience, Rbs - robustness, $Flmit$ - mitigating system hardware faults, $Sffun$ - functional Safety, Tst - testing, Evl - evaluation, Us - use, Apl - applicability.

The proposed models for the use of standard information are the next step in formalizing the requirements for AI. This makes it possible to expand and simplify the use of standards in order to accelerate their implementation in AI systems.

5. Experiment, Results and Discussions

In order to verify the correctness of the chosen approach to determining the form-forming components of the trustworthiness concept, experimental studies of the degree of conformity of the proposed structure to the information domain area were carried out. Trustworthiness's concept is defined in the ethics domain of the human-centric approach. Therefore, for this purpose, we will use methods of direct conformity and covering the completeness of the use of proximity functions and interpretation of principles. It is based on research and data obtained in Jobin A. (2019) [27], which globally analyzes the requirements, technical standards of ethical guidelines AI. On the basis of content analysis, ethical principles and documents in which they are meaningful are defined. The value of the selected data is that it consists of documents that are the result of extensive discussion and consolidated opinion. That is, this is not the opinion of an individual scientist, researcher, and so on, but is data of a high level of generalization. The corps consists of documents private companies, governmental agencies respectively, private agencies, academic and research institutions, inter-governmental or supra-national organizations, non-profit organizations, professional associations/scientific societies, private sector alliances, research alliances, science foundations, federations of worker unions, political parties.

The analysis showed that the method of direct comparison coincides at the level of categories {Transparency}, {Privacy}, which is the relative importance of 16% and 11% in the amount of 27%. Using the method of partial complete coincidence and meaningful correspondence {Explainability} has full coincidence and is the content of {Transparency}, by the method of including codes of direct coincidence - {strategies for reducing Bias} is the code {Justice & fairness}, {functional Safety} is the code {Non -maleficence}, which is 15% and 14% in relative importance, respectively, and 56% in the amount of the previous principles. According to the method of full compliance, the partial content of {Controllability} contains the full meaning and has the phraseological equivalent {Freedom & autonomy}, which is 8% in relative importance and 64% in the sum with the previous principles. In terms of partial full match and full coverage of content {Reliability, Resilience, Robustness} in the subcategory Reliability completely coincides with partial {Trust}, and in the category {Testing, Evaluation} completely coincides with partial {Justice & fairness} with complete coincidence of interpretation of content. Accordingly, the relative importance of {Trust} is 6%, in sum with the previous principles 70%.

The categories {mitigating system hardware Faults} and {Use, Applicability} are represented at the level of proximity functions of content interpretation. This method also covers the {Responsibility} principle, which is 14% in relative importance, and 84% in sum with the previous principles . Principles such as Beneficence, Sustainability, Dignity, Solidarity, the proximity functions of content interpretation in the standard ISO / IEC TR 24028 are defined at the level according to which they do not have full substantive compliance and amount to 16%. The experiment showed the effectiveness of the proposed approach to determine the model of forming the content of the concept of trustworthiness based on the principles and guidelines for ethical AI. The standard presents the concepts of trustworthiness formation which are confirmed by the content analysis of the corps. All the concepts of the ontology of trustworthiness formation have been confirmed at the level of the main principles of ethics as the global landscape.

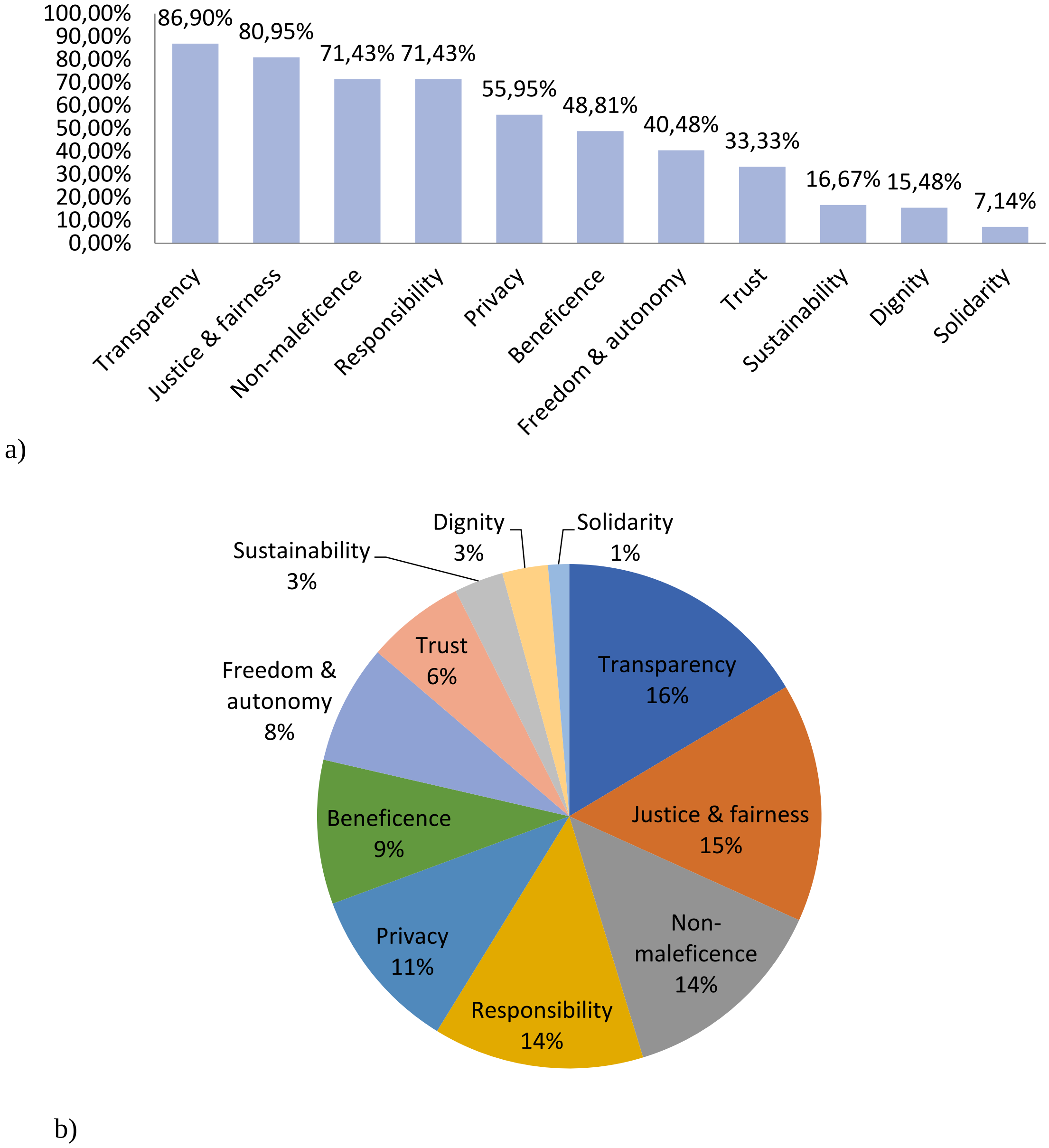

Figure 2: Based on data Jobin A. (2019) [26] distribution of AI ethical principles: a) the share of documents in which the principle is defined; b) the relative importance of the principle

This suggests that the standard is a relevant reflection of global trends. It should be noted that certain ethical principles identified in the present documents and regulations, which have a relative importance of 16% are not presented in the standard. This indicates that the ISO / IEC TR 24028 has a specialized purpose. The information content revealed by the content analysis at the level of identification of ethical principles is broader than the purpose of the ISO / IEC TR 24028 standard.

6. Conclusions

The rapid development of the AI information field is reviewed in the work of expert groups and international standards - International Organization for Standardization (ISO). Concepts, principles, approaches, scope, concepts of ethics and others are formalized. This is an important step in organizing, determining in the direction of practical use of exponential growth of the information field of the AI. The AI-based technical solutions market has gone beyond niche use and has acquired a value that has a significant impact on a person with prospects for rapid growth. Accordingly, this requires the development and implementation of technical AI standards. A long term is needed to develop standards competing with the rapid development of AI. Formalization of all stages, principles and approaches of AI requires in-depth development of scientific foundations. This, along with the acceleration of informatization of the whole society, is sometimes competitive. Formalization from this position is an important direction of systematization of domain field information. The development of model ontologies based on human-centric principles is an important step in the development of AI systems for important and critical applications.

It should be noted that the proposed model is developed within the standard and is limited by the standard itself. As the direction of trust in AI is at the stage of development and standardization, certain aspects may go beyond the considered standard and are accordingly not presented in the model. This can probably be considered a shortcoming, and it will be further refined with the evolution of trustworthiness AI. At the moment, the proposed model can be considered basic.

7. References

[1] S. M. Lee, J. B. Seo, J. Yun, Y. H. Cho, J. Vogel-Claussen, M. L. Schiebler & N. Kim, Deep learning applications in chest radiography and computed tomography. Journal of thoracic imaging, 34(2), 2019, pp. 75-85. doi: 10.1097/RTI.0000000000000387.

[2] A. S. Heinsfeld , A. R. Franco, R. C. Craddock, A. Buchweitz, & F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17, 2018, pp. 16-23. doi: 10.1016/j.nicl.2017.08.017.

[3] S. Grigorescu, B. Trasnea, T. Cocias, & G. Macesanu, A survey of deep learning techniques for autonomous driving. Journal of Field Robotics, 37(3), 2020, pp. 362-386. doi:10.1002/rob.21918

[4] A. Sahba, A. Das, P. Rad, & M. Jamshidi, Image graph production by dense captioning. In 2018 World Automation Congress (WAC), 2018, pp. 1-5. doi: 10.23919/WAC.2018.8430485.

[5] O. V. Barmak, Y. V. Krak, E. A. Manziuk, Characteristics for choice of models in the ansables classification. CEUR Workshop Proceeding 2139, 2018, pp. 171-179. https://doi.org/10.15407/pp2018.02.171.

[6] J. Boiko, I. Pyatin, O. Eromenko, & O. Barabash, Methodology for Assessing Synchronization Conditions in Telecommunication Devices. Advances in Science, Technology and Engineering Systems Journal, 5(2), 2020, pp. 320-327. doi: 10.25046/aj050242.

[7] B. Zhurakovskyi, J. Boiko, V. Druzhynin, I. Zeniv, & O. Eromenko, Increasing the efficiency of information transmission in communication channels. Indonesian Journal of Electrical Engineering and Computer Science, 19(3), 2020, pp. 1306-1315. http://doi.org/10.11591/ijeecs.v19.i3.pp1306-1315.

[8] O. Nedashkivskiy, Y. Havrylko, B. Zhurakovskyi, & J. Boiko, Mathematical Support for Automated Design Systems for Passive Optical Networks Based on the β-parametric Approximation Formula. International Journal of Advanced Trends in Computer Science and Engineering, 9(5), 2020, pp. 8207-8212. https://doi.org/10.30534/ijatcse/2020/186952020.

[9] J. Li, Cyber security meets artificial intelligence: a survey. Frontiers of Information Technology & Electronic Engineering, 19(12), 2018, pp. 1462-1474. https://doi.org/10.1631/FITEE.1800573.

[10] I. Krak, O. Barmak, E. Manziuk, Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology, Computational Intelligence, 2020, pp. 1-26. https://doi.org/10.1111/coin.12289.

[11] I. Krak, O. Barmak, E. Manziuk, & A. Kulias, Data Classification Based on the Features Reduction and Piecewise Linear Separation. In International Conference on Intelligent Computing & Optimization. Springer, Cham, 2019, pp. 282-289. https://doi.org/10.1007/978-3-030-33585-4_28.

[12] R. Damasevicius, J. Toldinas, A. Venckauskas, S. Grigaliunas, & N. Morkevicius, Technical Threat Intelligence Analytics: What and How to Visualize for Analytic Process. In 2020 24th International Conference Electronics, 2020, pp. 1-4. doi: 10.1109/IEEECONF49502.2020.9141613.

[13] I. Krak, O. Barmak, E. Manziuk, & H. Kudin, Approach to piecewise-linear classification in a multi-dimensional space of features based on plane visualization. In International Scientific Conference “Intellectual Systems of Decision Making and Problem of Computational Intelligence”, 2019, pp. 35-47. https://doi.org/10.1007/978-3-030-26474-1_3.

[14] Ö. N. Kenger, & E. A. Özceylan, Frequency-Based Approach For Multi-Class Data Classification Problem. In 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), 2020, pp. 1-4. doi: 10.1109/ICECCE49384.2020.9179335.

[15] J. Boiko, L. Karpova, O. Eromenko, & Y. Havrylko, Evaluation of phase-frequency instability when processing complex radar signals. International Journal of Electrical and Computer Engineering, 10(4), 2020, pp. 4226-4236. doi: 10.11591/ijece.v10i4.pp4226-4236.

[16] T. Neskorodieva, E. Fedorov, I. Izonin, Forecast Method for Audit Data Analysis by Modified Liquid State Machine. CEUR-WS vol. 2623, 2020. pp. 25-35.

[17] A. Nicheporuk, O. Savenko, A. Nicheporuk, Y. Nicheporuk, An android malware detection method based on CNN mixed-data model, CEUR Workshop Proceedings, Vol. 2732, 2020, pp. 198–213

[18] G. Markowsky, O. Savenko, S. Lysenko, A. Nicheporuk, The Technique for Metamorphic Viruses’ Detection Based on its Obfuscation Features Analysis, CEUR Workshop Proceedings, Vol. 2104, 2018, pp. 680-687.

[19] S. Lysenko, K. Bobrovnikova & O. Savenko, A botnet detection approach based on the clonal selection algorithm. In 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT). IEEE (2018) 424-428.

[20] A.V. Barmak, Y.V. Krak, E.A. Manziuk, & V.S. Kasianiuk, Information technology separating hyperplanes synthesis for linear classifiers. Journal of Automation and Information Sciences, 51(5), 2019, pp.54-64. doi: 10.1615/JAutomatInfScien.v51.i5.50.

[21] Ethics Guidelines for Trustworthy AI (AI HLEG, European AI Alliance, European Commission), 2019. URL: https://ec.europa.eu/futurium/en/node/6945#_ftnref50

[22] Council of Europe: AI Initiatives, 2020. URL: https://docs.google.com/spreadsheets/d/1mU2brATV_fgd5MRGfT2ASOFepAI1pivwhGm0VCT22_U/edit#gid=0

[23] EU Commission: Declaration of Cooperation on Artificial Intelligence, 2018. URL: https://ec.europa.eu/jrc/communities/sites/jrccties/files/2018aideclarationatdigitaldaydocxpdf.pdf

[24] EU Commission Lead DG (Responsible Unit) CNECT: Inception Impact Assessment: Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence, 2020. URL: https://ec.europa.eu/info/law/better-regulation.

[25] ISO/IEC TR 24028 Information technology. Artificial intelligence. Overview of trustworthiness in artificial intelligence (2020).

[26] A. Jobin, M. Ienca, & E. Vayena, The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 2019, pp. 389-399. https://doi.org/10.1038/s42256-019-0088-2.