ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 004.82.023

doi: 10.32620/reks.2022.1.04

Sergiy DOTSENKO[1], Vyacheslav KHARCHENKO[2], Olga MOROZOVA[2], Andrzej RUCINSKI[3], Svitlana DOTSENKO[4]

[1] ***Ukrainian State University of Railway Transport, Ukraine***
[2] ***National Aerospace University "Kharkiv Aviation Institute", Ukraine***
[3] ***University of New Hampshire, USA***
[4] ***H. S. Skovoroda Kharkiv National Pedagogical University, Ukraine***

HEURISTIC SELF-ORGANIZATION OF KNOWLEDGE REPRESENTATION AND DEVELOPMENT: ANALYSIS IN THE CONTEXT OF EXPLAINABLE ARTIFICIAL INTELLIGENCE

From the analysis of the main theoretical provisions of heuristic self-organization systems and logical models, it follows that according to O. G. Ivakhnenko's systems of heuristic self-organization, the first task is to determine the factors content "that determine the essence of different images". These are the images that characterize the objects of a particular subject area. After determining the composition and content of these images, the next problem is solved, namely, the problem of "generating the new successful heuristic", which in content is a solution that leads to increased accuracy. Note that we are talking about improving the accuracy of solving the problem of data processing. It follows from the above mentioned that heuristic self-organization systems are data processing systems. This allows the multiplicity of heuristics. Heuristics in content correspond to the logical rules applied in heuristic self-organization systems. The main provisions of the heuristic self-organization system theory were developed by O. G. Ivakhnenko in the eighties of the last century, but they remain unnoticed to this day. At this time, the task is to explain why the neural network makes such a decision and not another. Based on this, the concept of "explainability of artificial intelligence" was introduced for artificial intelligence. It is the content of heuristics that forms the structure of the neural network in the form of logical rules and determines the logic of the decision made. It is established that the derivation rule, which is the basis for constructing artificial neural networks, is an abductive rule, which, unfortunately, does not meet the fourth heuristic and does not meet the definition of intelligence: intelligence is the ability to measure things. Unfortunately, none of the neural networks can measure things. From the analysis of the basic rules content of inference, it follows that the dialectical method of inference is general (generating) for the basic logical methods of inference. The difference lies in the composition and content of the middle member of the triangular relationship, namely, in the form of the element combination of the relationship: the transition from one concept to another. The explainability of artificial intelligence refers to the laws of the structure and activity of artificial neural networks. But modern theories of artificial neural networks ignore the existence of logical rules (heuristics), which were established by O. G. Ivakhnenko. After all, only knowing the rules based on which problems are solved, it is possible to check the correctness of the decision, but not by searching for such rules. The three hypotheses about the explainability of artificial intelligence and the theory of machine identification can be further defined as statements or theorems and strictly proved.

Keywords: *heuristics; self-organization; knowledge; logical inference; explainable artificial intelligence.*

1. Introduction

1.1. Motivation of research

Artificial intelligence (AI) and intelligent systems have become a powerful trend in the development of cybernetics and information technologies. The technological aspects have become prevalent over the last decade; technologies and products that implement methods of creating the artificial (quasi- and sometimes pseudo-artificial) intelligence are filling the digital world, becoming more accessible and pervasive. Leading industry giants Google, Microsoft, Amazon and other companies have created a powerful market and provide a new type of service "AI as a Service" (Artificial Intelligence as a Service, AIaaS) [1]. Such services are used by developers and integrators of different systems, choosing the appropriate services and products for different applications. Therefore, the emergence and improvement of AIaaS actually forms the chain of artificial intelligence tools (AIT) consumption as the so-called Commercial of the Shelf (COTS) [2].

Thus, the level of people's comfort live, efficiency, competitiveness and security of industry and society as a whole are becoming increasingly dependent on the AIT. In the future, this dependence may become somewhat critical. As a result, the direction of research and development, which is called Explainable AI (XAI) and Trustworthy AI (TAI), i.e. explained (or understandable) and credible (or truthful) AI is developed in parallel in recent years [3]. The first normative documents of the European Union were developed, in particular, for the aviation and aerospace industries [4], the National Institute of Standards and Technologies of the USA NIST [5]. The works of scientists from the Institute of Cybernetics of the National Academy of Sciences of Ukraine are devoted to the problems of artificial intelligence, which relate to its use in high-performance computing, knowledge representation, intellectual modeling.

To successfully implement of artificial intelligence, it is need to understand whether AIT is a white box in the full sense? It is also need to answer the few questions:

1) first, to what extent the methods of AIT creating and in general the methodological foundations of intelligent systems are perfect in terms of dialectics and theory of knowledge and ensuring the explanation of AIT;

2) second, to what extent their operational base meets the challenges, which related to the need to ensuring the explanation and trustworthiness.

To do this, it is logical to analyze of knowledge representation models, relevant rules and heuristics, methods of AIT creating, such as neural networks, expert systems, etc.

1.2. State of the Art

1.2.1. Logical and heuristic models

According to [6] models of knowledge representation can be divided into logical and heuristic. Logical models are based on the concept of formal theory. In this case, formal theories are implemented on the basis of deductive and inductive methods of logical inferences (conclusions). The deductive type of inference includes methods of calculating the predicates and concrete systems of products. The inductive type of logical inferences includes methods of relationship logic (pseudo-physical logic). It follows that there are basically two types of knowledge models, namely, logical and heuristic models.

In [7] it is also noted that in logical models of knowledge representation, the relationships that exist between individual units of knowledge are expressed only by those few means that are represented by the syntactic rules of the formal system used. The formal system is defined by four of such type [8]:

$$M=\langle T,P,A,B\rangle, \qquad (1)$$

where T is the set of basic elements of different nature; P is the set of syntactic rules by means of which syntactically correct sets are formed from the elements T, in the set of which some subset A is distinguished, the elements of which are called axioms; B is the set of inference rules, applying which to the elements A, it can be getting the new syntactically correct sets, to which it can be again apply the rules of B.

Heuristic models are characterized by the presence of a diverse set of tools that convey the concrete features of a particular problem area. In [6] it is also noted that it is for this reason that heuristic models are superior to logical ones, both in terms of their ability to represent the problem area and in terms of the inference rules used effectiveness. Heuristic models used in expert systems, respectively, include network, frame [8] and production models [9].

The question arises, the inference rules in logical and heuristic models are the same or they different?

1.2.2. Heuristic self-organization systems

The research of heuristic self-organization methods was performed in [10, 11]. In these works, the main methods of heuristic self-organization for the processes of finding the new information, which are implemented in AI systems [10], as well as in the processes of semantic thinking and semantic activity in natural intellectual systems [11] were investigated.

Systems (programs) of heuristic self-organization are defined as systems in which there are generators of random combinations (hypotheses) and several series of useful information selection. The complexity of combinations from line to line increases, and the accuracy of the solution increases until the optimal algorithm for processing the information is obtained, which correspond to the minimum of the appropriate selection criterion [12].

Thus, the system of heuristic self-organization provides a solution to the problem of finding the "optimal in complexity algorithm for information processing" based on the selection laws. In addition, self-organization refers to information processing activities. However, the composition and content of the processed information are unknown. Instead, there are known problems that need to be solved using this information, namely: pattern recognition and choosing the right hypothesis in decision making [13].

The main elements of such a system are:

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

– generators of random combinations (hypotheses);

– series of useful information selection.

Regarding the content of the "heuristic self-organization" concept, in [12] it is proposed to compare the system of heuristic self-organization with a multi-layered "pie", because: in it heuristic self-selections of useful information are interspersed with mathematical data processing several times according to the scheme "heuristics – processing – heuristics – processing – heuristics, etc".

Heuristic self-selection of useful information must be identified with generators of random combinations (hypotheses), heuristics. Unfortunately, an unambiguous definition of the heuristic generator model has not been proposed. However, it should be pay attention to the following circumstance. Exploring the components of the thinking process, namely, the creative component of thinking in the form of a heuristic generator, in [4] notes that factor analysis allows you to find "factors" – the values that determine the nature of different images. By using a machine to get the factors and knowing their dimensions, a person can easily guess the nature of the new features that need to be introduced into recognition to make it more accurate and faster. The most pleasant (and most important) part of the process remains for a person is generating a new successful heuristic, and factor analysis helps to easily and simply find new very effective signs.

That is, the task is to minimize person participation in the generation of heuristics through the generation by a computer random sets of signs (factors).

It should be noted that the computer selects features from a given set – a set given by person [12]. However, it is not necessary to rely on the computer to identify and analyze the factors. A person can also solve this problem and not only indicate the composition and content of signs (factors), "which determine the essence of different images" (factors), but also to establish a new form of relationship between them.

On the basis of the provisions formulated in relation to the systems of heuristic self-organization, in [10] the content of the following heuristics of O. G. Ivakhnenko was determined:

– the first is the choice of elementary algorithms;

– the second is the choice of evaluation criteria and algorithms for their change;

– the third is the choice of integrated actions and schemes for their implementation.

To form these heuristics, a method that initiated the third direction in cybernetics, which is to model the laws of evolution and selection that are observed in nature, was used [12].

In [10, 11] the definition of the content of the fourth and fifth heuristics was formulated: the fourth is dialectical self-organization for the concepts of "general" "single" [10]; the fifth is architectures of logical models of semantic thinking and semantic activity are formed using two pairs of factors created for each model generated on the basis of the fourth heuristic, which in meaning correspond to pairs of process and resource factors. In this case, the elements of each pair are related by causal relationships and correspond to the architecture of the Cartesian coordinate system [11]. Their construction is based on the concept of "factor", the law of dialectically opposite concepts unity, as well as the central pattern of integrative activity of the brain (for the fifth heuristic) [12]. These heuristics are the basis for the formation of four-factor models of knowledge [11].

To implement the fourth heuristic requires the ability to divide the factors according to their content into "general" and "concrete". It is enough to understand that the formation of general concepts is the result of mental activity, and the formation of concrete concepts is associated with the reflection in the mind of the characteristics and properties of concrete objects or actions in the physical environment.

For both the third and fifth forms of heuristics, the question arises as how to make a choice of integral actions and schemes for their implementation, that is, how to integrate factors into a single heuristic. For the third heuristic, these factors are unique to a particular subject area. In contrast to the fifth heuristic, the composition and content of factors are clearly defined – these are process and resource factors, pairs of which are the basis for building logical models [11].

On the other hand, along with the architecture of logical models based on the fifth heuristic, discussed above, there are other models of factor knowledge representation. Most of them include eight factors [14]. In addition, there are other four-factor models of knowledge representation that correspond to the fifth heuristic [15].

1.2.3. Analysis of the artificial intelligence tools characteristics

The main characteristics of modern AI systems are considered in [16, 17]. Among them, the most important and discussed in recent times are their explainability and trustworthiness. Trustworthiness of AI systems is defined by the following terms of [16-18]:

– the confidence in the AI system: consumer confidence, and if necessary, the organizations, that responsible for regulating the creation and application of AI systems, and other stakeholders, that the system is able to perform the tasks assigned to it with the required quality.

– the trusted AI system: AI system in which the consumer and, if necessary, the organizations, that responsible for regulating the creation and application of AI systems, show confidence.

The principles of explained AI are defined in the US standard of the National Institute of Standards and Technology [5], namely: explanation, meaningful, explanation accuracy, knowledge limits.

Let's emphasize that these and other characteristics of AI, which are extremely important for its further use, especially in critical systems, are in no way related to the model base of AI and require further development and research at the methodological level.

1.3. The purpose and tasks of research

From the above review of literature sources, it follows that the following tasks need to be solved:

– analysis of the main theoretical provisions of heuristic self-organization systems and logical models;

– comparison of the inference rules content used in logical models based on formal theory and in logical models based on the fourth heuristic.

– research of explainability of AI in the context of logical and heuristic models' analysis.

The purpose of this work is to solve these problems and search for universal logical rules, and hence the common properties of different logical models of knowledge (Section 2), considering the rules of inference (Section 3). In addition, it is necessary to determine their compliance with the requirements to ensure the explainability and trustworthiness of the AIT (Section 4) and to summarize the results of research in the context of technological trends in the development of intelligent systems (Section 5).

2. Analysis of the fundamentals of heuristic self-organization systems and logical models

2.1. Analysis of the theoretical provisions of heuristic self-organization systems O. G. Ivakhnenko

Heuristic self-organization systems are evolving as cybernetic systems. The third direction of <cybernetics> is to model the laws of evolution and selection, which are observed in nature. At the same time, systems (programs) of heuristic self-organization are defined as systems in which there are generators of random combinations (hypotheses) and several series of selection of useful information.

The complexity of combinations from line to line increases, and the accuracy of the solution increases until the optimal algorithm for processing information is obtained, which meets the minimum of the appropriate selection criterion [12].

The question arises what is the generator of combinations (hypotheses, heuristics) and what is the composition of the respective series of selection, which are the part of the heuristic self-organization system?

Heuristic self-selection of useful information must be identified with generators of random combinations (hypotheses), heuristics. Unfortunately, an unambiguous definition of the heuristic generator model has not been proposed.

O. G. Ivakhnenko reveals the meaning of the "heuristics" concept in contrast to the determinists, who define it as an unreasonable decision that leads to practically sufficient results, and as a decision that primarily leads to increased accuracy [12]. As examples of concrete heuristics identified the following heuristics:

– the choice of the initial set of signs;

– the choice of criteria for useful information selection;

– organization of perceptron structure, which allows to repeatedly strengthen the integral effect of heuristic criteria on the flow of information.

It should be noted that the heuristic self-organization system is like a multi-layered "pie": in it heuristic self-selection of useful information is interspersed with mathematical data processing several times according to the scheme "heuristics" – "processing" – "heuristics" – "processing" – "heuristics", etc.

It is interesting to note the essence of the processes that are implemented in the perceptron [12]. It is noted that the perceptron as a system, that performs integral influences and the selection of useful information, with increasing from a number to a number of its complexity, is still not understood. This conclusion may seem somewhat conservative in view of the powerful development of neural networks, but important things are fundamental things – the explainability of AI remains, even becoming an increasingly important problem of AI systems.

The content of integrated influence in technical cybernetics is defined as one that does not use information about the state of each element of a complex system separately, but is selected by the total result of the action on many elements. Threshold self-selections are the simplest form of integral influence [19]. The concrete realization of integral influences is realized in the form of threshold self-selections in the corresponding elements of artificial neurons models [20].

From the above it follows that the integral effect is on the outputs of all neurons in the form of threshold self-selection. In this case, the input signals in the neuron are summed with the corresponding weights.

Exploring the components of the thinking process, namely, the creative component of thinking in the form of a heuristic generator, O. G. Ivakhnenko notes that

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

factor analysis allows you to find "factors" – the quantities that determine the essence of different images.

By using a computer to know the factors and knowing their dimensions, one can easily guess the nature of the new features that need to be introduced into recognition to make it more accurate and faster. The most pleasant (and most important) part of the process remains for a person is generating a new successful heuristic, and factor analysis helps to easily and simply find new very effective signs [12].

If it is understanding the content of heuristics as solutions that lead to increased accuracy, then the content of logical rules that are implemented in the system of heuristic self-organization in the form of a perceptron is as follows:

– heuristic self-selection of useful information;
– integral action (summation of signals in the neuron);
– threshold self-selection (mathematical data processing).

Thus, in the systems of heuristic self-organization, the first task is to determine the content of factors 'that determine the essence of different images". It is about images that characterize the objects of a particular subject area.

After determining the composition and content of these images, the next problem is solved, namely, the problem of "generating a new successful heuristic", which in content is a solution that leads to increased accuracy. It should be noted that it is talking about improving the accuracy of the data processing problem.

It follows from the above that heuristic self-organization systems are data processing systems. The plurality of heuristics is allowed. Examples of such heuristics are given above (see subsection 1.2).

2.2. Some conclusions about the explainability of AI

It follows from the above that heuristics in content correspond to the logical rules applied in the systems of heuristic self-organization. On the other hand, according to [21], the operation of removing rules from neural networks is introduced. It is noted that artificial neural networks (ANN) are well-known parallel computational models that demonstrate excellent behavior in solving complex problems of AI.

However, many researchers refuse to use them because they are like "black box". This is especially true for deep learning networks. This means that determining why a neural network makes such a decision is a difficult task [21].

Based on this, for AI introduced the concept of explainable or intelligible AI [3, 18]. The operation of removing the logical rules used in the neural network is important. Table 1 shows the content of logical rules used in the relevant models of neural networks, as well as algorithms that ensure the implementation of the operation of their extraction [20].

From the analysis of this table content, it could be conclude that the following logical rules are applied in neural networks:

– Binary rule;
– Decision tree;
– Binary Decision tree;
– IF-THEN;
– M-of-N;
– M-of-N spilit;
– Hyperplane rule.

Table 1

The Content of Logical Rules

Algorithm	Network Type	Algorithm Type	Extracted Rule Type
DIFACON- miner	Standard MLP	Decompositional	IF-THEN
CRED	Standard MLP	Decompositional	Decision tree
FERNN	Standard MLP	Decompositional	M-of-N. IF-THEN
KT	Standard MLP	Decompositional	IF-THEN
Tsukimoto's Algorithm	Standard MLP and RNN	Decompositional	IF-THEN
TREPAN	Standard MLP	Pedagogical	M-of-N spilit decision tree
HYPINV	Standard MLP	Pedagogical	Hyperplane rule
BIO-RE	Standard MLP	Pedagogical	Binary nil
KDRuleEX	Standard MLP	Pedagogical	Decision tree
RxREN	Standard MLP	Pedagogical	IF-THEN
ANN-DT	Standard MLP	Pedagogical	Binary Decision tree
RX	Standard MLP	Eclectic	IF-THEN
Kahramanliand Allahverdi's Aljorithm	Standard MLP	Eclectic	IF-THEN
DeepRED	DNN	Decompositional	IF-THEN

As it can be seen, for artificial neural networks it is talking about logical rules, which are elements of inference rules, but it is not about applying inference rules, which are required attribute of logic models. It is this circumstance that determines the need to develop the concept of AI explainability.

From the above it is also valid to conclude that the views of O. G. Ivakhnenko to methodology for constructing artificial neural networks, unfortunately, is not taken into account in full, namely, the need to factorize the search space and is not taken into account; formation of heuristics in the form of inference rules.

2.3. Analysis of the main theoretical principles of logical models

The logical model based on the formal system (1), which is described in paragraph 1.1, begins to be developed with the formation of a basic elements set T. It is noted that these elements have a different nature. Their content is determined by a concrete subject area.

The next stage is the formation of syntactic rules P by which the elements of T form syntactically correct sets. In fact, these expressions characterize certain elements properties of the subject area and their relationships. It is clear that the set of these rules is unique to the particular elements set of the relevant subject area.

After that, the axioms A, which are the reference expressions for this subject area, are selected from the set of syntactic rules. It is clear that these reference expressions are in some way related. These connections are disclosed using inference rules.

The question arises, what actually determines these rules? According to national standards of Ukraine [22] the rules of inference are defined as follows:

– inference is the process of obtaining the new knowledge on the basis of previously known;

– logical inference is a sequence of reasoning that leads from a fixed set of preconditions for the conclusion using axioms and inference rules;

– deductive inference is a logical inference based on the use of the deduction principle – from general to partial;

– inductive inference is a logical inference based on the use of the induction principle – from partial to general.

From the definition meaning of the "logical inference" concept, it follows that for the inductive and deductive rules of inference it is necessary to form a fixed set of preconditions. These preconditions are concepts that in G. Hegel's dialectical logic are defined as "general" and "concrete" concepts.

This raises two questions:

– what axiom(s) is used in these rules?

– what exactly is the content of the inference rule (direction of transition)?

Before answering these questions, it should be pay attention to the following logical rule of inference, namely: abductive inference is a plausible inference from partial (concrete) to partial (concrete). In this case, the content of plausible inference is determined as follows: plausible inference is a method of inference, in which each step is accompanied by the calculation of the reliability assessment of the obtained conclusion [23].

Let's compare the content of this logical rule of inference with the content of the "heuristic self-organization" concept according to [12], where the system of heuristic self-organization is compared with a "multilayer pie", where heuristic self-selection of useful information is several times interspersed with mathematical data processing according to the scheme "heuristics – processing – heuristics – processing…".

In this system of heuristic self-selection, the logical rule of inference in the form of abductive inference is realized according to which each step is accompanied by calculation of reliability estimation of the received conclusion in the form of transition from partial (concrete) to partial (concrete) concepts. After all, the data that are distributed in the neural network are always concrete (partial) in contrast to the general concepts that are part of the deductive and inductive rules of inference. The content of abductive inference corresponds to the content of the third heuristics of O. G. Ivakhnenko – the choice of integral actions and the scheme of their implementation.

Let's move on to answer the questions:

– what axiom is used in the rules of inductive and deductive inference?

– what exactly is the content of the inference rule (direction of transition)?

3. Analysis and comparing the inference rules used in logical models based on formal theory and fourth heuristics

3.1. Analysis of the inference rules

Based on the representation of G. Hegel's logical method in the form of "general" → "special" → "concrete", it is proposed to represent this relationship as a mathematical relationship [23]:

$$\text{“general”} \gtrdot \text{“concrete”}. \qquad (2)$$

In this expression, the sign $\gtrdot$ is a sign of the usual relational relation of the unity of dialectically opposite concepts and corresponds to the concept of "special" in

Radioelectronic and Computer Systems, 2022, no. 1(101)
ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

the triple dialectical relation. The work [23] defines the meaning of the sign that denotes this relationship: “sign of dialectically opposite concepts unity, “⋗” is a sign of the usual relational relationship, by which the concept of a single object or single factor of activity is combined with the concept of class (set) of such objects or “general” factors of its activity. In this case, the concept of the objects class, or “general” factors of its activities, it is understanding the knowledge of individual objects, which in some way formed by mental semantic activity as general concepts”. The sign provides a formal mathematical notation of a binary relational dialectical relation, missing in the algebra of relations.

Regarding the belonging of heuristics to the subject of mathematics O. G. Ivakhnenko [24] notes that self-organization should be associated with heuristics – assumptions about the appropriateness of one or another action. Heuristics are the decisions related to the consumer's desires for the results of solving the problem, with factors of his motivation. They do not belong to the subject or to the competence of mathematics, and therefore, no improvement of the mathematical apparatus can replace them or compare them with them in action. Therefore, the accuracy of heuristic methods was incomparably higher than the accuracy of the most advanced and general mathematical methods that use concrete (deterministic) approaches.

However, for a certain fourth heuristic (unity of dialectically opposite concepts “general” and “singular”), a formal mathematical description of the relationship between the factors that reveal the essence of the researched object of equation (2) is possible. That is, the thesis that heuristics belong neither to the subject nor to the competence of mathematics, and therefore, no improvement of the mathematical apparatus can replace them, or be compared with them in action can be refuted at least for the principle of heuristic dialectical self-organization of the intellectual system in its “existence”.

It follows from the above that deductive and inductive inference are practical demonstrations of the dialectical unity law and concepts interdependence, on the basis of which the fourth heuristic is formed. In addition, it could be concluded that the main inference rule in logical models of knowledge based on formal theory is actually the fourth heuristic, which is based on philosophical foundation in the form of the basic law of philosophy.

Unfortunately, the fourth heuristic does not answer the question, how to apply these inference rules? After all, these rules define the direction of transition between concepts. It is clear that any reality is governed by the passage of time. That is, there is a certain logic of transition between certain physical states. This logic of transition in time is demonstrated in the transitions between concepts for relation (2). Based on this, it is proposed to denote this form of transition between the concepts for relation (2) by the following mathematical relations in the form of ordinary relational operators of primacy:

“general” • ⋗ “concrete”, (3)

“abstract concrete” • ⋖ “general”. (4)

The dot denotes a concept that is primary in the implementation of the logical inference rule by moving between the concepts of the dialectical relationship “general” ⋗ “singular”. The operators •⋗ and •⋖ are additional to the dialectical unity operator and provide a mathematical record of the operation “sequence in time” of the dialectical relation realization (2).

It should also be noted that these relations are representations of the well-known logical methods of G. Hegel (3) and Karl Marx (4). This form corresponds to the well-known triangular dialectical relation: “general” ⋗ “special” ⋗ “concrete”. The study of this relation was performed in [12], where it is shown that the content of the middle term of this triangular dialectical relation determines the form of combination (form of motion) for parts of the studied whole.

3.2. Comparison with known inference rules. Definitions of concepts

Let's compare these inference rules with the known inference rules according to [10]. Interestingly, this approach allows a formal representation in terms of set theory of basic logical methods (rules) of inference, namely:

deductive conclusion: “general” • ⋗ “concrete”,
from general to concrete; (5)

inductive inference: “concrete” • ⋖ “general”,
from concrete to general. (6)

Note that the abductive inference can be represented by a combined sign of unity:

“concrete” ⋖⋗ “concrete”, (7)

that is, from concrete to concrete.

In this form of inference, it is not defined which of the concrete concepts is primary. Abductive inference is not a consequence of relational operators of primacy. If it will be applying this rule to the perceptron model, in which, according to O. G. Ivakhnenko, specific operations (actions) are defined by the following concepts:

– heuristic self-selection of useful information;
– integral action;

– threshold self-selection (mathematical data processing),

then mathematically the perceptron model using abductive inference can be represented as follows:

heuristic self-selection of information $\bullet \lhd\!\rhd$
integral action $\bullet \lhd\!\rhd$ threshold self-selection. (8)

Based on the introduced relational operators, it is possible to present mathematical notation of other rules and methods of logical inferences defined in [10]:

logical inference: "set of preconditions"
$\bullet \rhd$ "axioms, inference rules"
$\bullet \rhd$ "conclusion", (9)

monotonous inference: "set of preconditions"
$\bullet \rhd$ "axioms and inference rules"
$\bullet \rhd$ "conclusion, the truth of the conclusion does not decrease", (10)

non-monotonic inference: "set of preconditions"
$\bullet \rhd$ "axioms and inference rules"
$\bullet \rhd$ "conclusion, the truth of the conclusion decreases", (11)

plausible inference: "set of preconditions"
$\bullet \rhd$ "axioms and inference rules"
$\bullet \rhd$ "conclusion, the truth of the conclusion determined by the assessment calculation of its reliability", (12)

inference by analogy (variant of plausible inference): "set of preconditions"
$\bullet \rhd$ "analogies between subject area structures" $\bullet \rhd$ "conclusion". (13)

The considered logical rules of inference (8) – (13) are based on the fourth heuristic. In this case, the formation of axioms sets and inference rules in (8) – (10), as well as analogies between the structures of the subject area in (13).

The following inference rules of axioms and inference rules are not formed:

fuzzy inference: "set of preconditions"
$\bullet \rhd$ "conclusion, statement take the meaning of "truth", "lie", as well as intermediate values", (14)

probabilistic inference: "set of preconditions"
$\bullet \rhd$ "conclusion, each statement" (15)

direct inference (strategy): "output preconditions"
$\bullet \rhd$ "target conclusion", (16)

reverse inference (strategy): "given conclusion"
$\bullet \lhd$ "output preconditions". (17)

The inference rules (14) – (17) are also based on the fourth heuristic.

Thus, all the rules of logical inference except abductive inference are based on the law of concepts dialectical unity. No less important is the role of the second basic law of philosophy, namely: the law of primacy. In general, it is defined as the law of primacy determination for the concepts of "matter" and "consciousness". Materialists recognize matter as primary, and idealists, on the contrary, recognize consciousness as primary. Important for this study is the principle on the need to establish a sequence in the time of concepts application for their dialectical unity in form (2).

Therefore, it is possible to form the following definitions of "dialectics" and "logic". Dialectics is the principle of the opposite states unity for objects of living and non-living nature, as well as concepts for the processes of thinking and semantic activity.

Logic is the principle of the transition in time between the states of parts in the organized whole for objects of living and non-living nature, as well as between concepts for the thinking processes and semantic activity. It should be noted that the dialectical relation in form (2), which is the content of the fourth heuristic, has an additional interesting property: the dialectical unity of the concepts is defined as a measure [14]. A linear measure is formed from the specified pair. Let's analyze the definitions of "thinking", "measure" and "intelligence"" by also adding the concept of "dimension":

Definition 1. A measure is the representation of a thing in the form of the dialectical unity of the concepts "general (qualitative definition) $\rhd$ single (quantitative definition)", namely: the general concept of the thing $\rhd$ the concrete concept.

For example, the well-known G. Hegelian "fruit" $\rhd$ "cherry" is an example of measuring the particular thing in thinking through the dialectical unity of quantitative (cherry) and qualitative (fruit).

Definition 2. Thinking is the ability to present things in measure.

Definition 3. A dimension is a process of intellectual activity that results in the formation of two dialectically related concepts about a thing or its properties.

Definition 4. Intelligence is the ability to realize the process of measuring the things.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Therefore, intelligent systems, both natural and artificial, must be able to "measure" things, as well as their properties. It also follows that the concepts that define knowledge about the subject area of intelligent systems in the knowledge base should be presented in measure. All considered methods of inference for models of knowledge representation form a single linear measure (2): "general" $\rhd$ "single". The practical application of this relation is linear measures in the form of relations (3) and (4). They are generating for relations (5), (6), (9) – (17).

Using the described models and rules, a knowledge management model was developed for the information security management system based on the standards ISO / IEC 15408, ISO / IEC 18045 and multi-share control structure [25, 26].

4. Explainability of artificial intelligence in the context of logical and heuristic models analysis

4.1. Principles of explainability of artificial intelligence and the reasons for its limitations

The National Institute of Standards and Technology (NIST) published in August the first draft of the Explainability of Artificial Intelligence (XAI) principles list. The provisions of the document focus on the status of intelligibility of artificial intelligence and define four principles underlying in the base of intelligible artificial intelligence [5].

1. **Explanation**. AI systems must provide the reasons and circumstances on the basis of which certain decisions were made. The principle of explanation obliges the AI system to provide explanations in the form of "evidence or justification of each result". This principle does not impose any additional requirements on the quality of the explanation, but only requires that the AI system be able to provide an explanation. The standards of such explanations are governed by other principles.

2. **Meaningful.** Artificial intelligence systems that can be explained should provide explanations that are understandable to individual users. The principle of meaningful establishes that the recipient of the explanation must be able to understand the explanation. The document emphasizes that this principle is not intended for universal application. Explanations should be tailored to the audience at both the group and individual levels.

3. **Explanation accuracy**. The explanation should accurately reflect the essence of the processes implemented by the artificial intelligence system to generate results. The principle of explanation accuracy corresponds to the principle of meaningful for regulating the quality of explanations, assuming the accuracy of explanations, but not the accuracy of decisions.

4. **Knowledge limits**. The system works only in the conditions for which it was designed or when the system achieves proper reliability in its results. The principle of knowledge limits requires that the system noted any cases for which it was designed. The purpose of this principle is to prevent misleading the explanations or conclusions from the system.

It follows from the above that the problem of explainability and trustworthiness of artificial intelligence systems arises from ignorance of the actual laws of its formation and operation. The question arises, what exactly causes the opacity and incomprehensibility of artificial intelligence? Table 2 provides answers to the questions for artificial and natural intelligence.

These four principles show that artificial intelligence solutions must have the necessary transparency to inspire confidence in their functioning and conclusions of the system. Artificial intelligence, which can be explained or transparent artificial intelligence, is a system in which people can easily understand the actions of artificial intelligence. The concept of intelligent artificial intelligence can strengthen trust in technology, as companies will have to explain how and why their artificial intelligence systems make certain decisions [24].

At the same time, artificial intelligence systems are presented in two versions: on the basis of heuristic self-organization systems according to O. G. Ivakhnenko and the theory of artificial neural networks. In our opinion, the following problems cause low explainability of artificial intelligence:

– inconsistency of the artificial neuron model with the actual processes occurring in the natural neuron;

– the pattern of the human cerebral cortex activity does not correspond to the algorithms implemented in the artificial neural network;

– the content of tasks assigned to artificial intelligence systems does not correspond to the content of tasks solved on the basis of natural intelligence.

From the content of these differences it is clear that the level of artificial intelligence explanation depends on the level of explanation and compliance of artificial intelligence with natural intelligence. Table 3 analyzes the compliance of the studied logical and heuristic models with the principles that characterize the explainable artificial intelligence (XAI).

Therefore, the explainability of artificial intelligence relates to the laws of structure and activity of artificial neural networks. But modern theories of artificial neural networks partially ignore the existence of logical

Table 2

The properties of systems that use artificial and natural intelligence

Types of models	Neuron model	Regularities of activity	Problems that are solved with the use of intelligence
Artificial neural network	Summation	There are no clearly defined logical rules. The operation of finding logical rules in existing models of neural networks is being implemented (see Table 1)	Solving data mining problems. Pattern recognition, decision making support.
Heuristic self-organization systems for O. G. Ivakhnenko [12]	Summation	Heuristics: – heuristic self-selection of useful information; – integral action (summation of signals in the neuron); – threshold self-selection (mathematical data processing).	Solving data mining problems. Pattern recognition, decision making support, management
Natural intelligence according to P. K. Anokhin [13]	Simultaneous convergence of four forms of signals on each of the neurons	Central pattern of integrative activity of the brain: simultaneous convergence of four forms of signals on neuron complexes the fifth heuristic	Forming a project of the future result of activity and ensuring its implementation as part of a functional system as an organized whole
Logic models in expert systems [7, 8]		Formal theory (1), inference rules, logical rules	Gaining new knowledge based on axioms and inference rules

Table 3

Analysis of the compliance level of the studied logical and heuristic models with the principles that characterize the explainable artificial intelligence (XAI)

Types of models	The principles of the explainable artificial intelligence (XAI)				Ability to ensure compliance with the principles
	Explanation	Meaningful	Explanation accuracy	Knowledge limits	
Artificial neural network	Low	Low	Low	Low	Additional analysis of the applied logical rules is needed
Heuristic self-organization systems for O. G. Ivakhnenko [12]	High	High. Pre-factorization accurately determines the purpose of the system	High	High	Provided by a preliminary definition of the heuristics content
Natural intelligence according to P. K. Anokhin [13]	High	High. Designed to solve the fundamental problem of the existence and operation of an intelligent system	High	Sufficient	Provided by the implementation of the fourth and fifth heuristics
Logic models in expert systems [7, 8]	High	High. Designed to solve the fundamental problem of a living system existence. Extraction of new knowledge	High	High. Clearly defined logical rules	Provided by the implementation of the inference rules in the form of the fourth heuristic

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

(heuristics), the content of which is established by O. G. Ivakhnenko.

Thus, the artificial intelligence explainability relates to the laws of structure and activity of artificial neural networks. But modern theories of artificial neural networks partially ignore the existence of logical rules (heuristics), the content of which is established by O. G. Ivakhnenko. After all, only knowing the rules on the basis of which problems are solved, it is possible to verify the correctness of the decision, and not by searching for such rules according to [24].

The problem of AI explainability should be considered as a cybernetic problem. Two cybernetic principles must be considered: controllability and observability. Explainability includes observability, i.e. is the last part of it. It is impossible to provide explainability without observation. In turn, the explainability is part of controllability, as it is impossible to control the structure of AI without explainability.

4.2. Theory of machine identification and determination of conditions for AI explainability

It is important, in view of the problem of artificial intelligence explainability, to compare it with the problems solved in the theory of diagnosis and identification of digital finite state machines [27].

In fact, when analyzing the artificial intelligence tools explainability as a digital system, there is an analogy in the formulation of problems, namely as:

– determining the presence and search for an identifying sequence in the machine (the task of determining the state in which the machine is);

– search for the sequence of the machine installation in a given state;

– determination of the output sequences of the automaton for an arbitrary state and input sequences.

These tasks have additional interpretations and are significantly complicated in conditions where machine failures are assumed. In this case, the identification problems are further described by the model of the automaton defects and are formulated as follows:

– determining the presence and search for an identifying sequence in the machine (the task of determining the state in which the machine is) for a given set of defects;

– determining the presence of the search for the sequence of the machine installation in a given state for a given defect;

– determining the output sequences of the automaton for an arbitrary state and the input sequences for a given set of defects;

– (additional task): determining the control and diagnostic sequences of the machine for a given set of defects.

Let`s formulate three hypotheses about the artificial intelligence explainability and the theory of automata identification, which can be further defined as statements or theorems and strictly proved.

Hypothesis 1. The artificial intelligence system can be represented by a finite state machine with memory and described by sets of input X and output Z signals (input and output alphabets), set of states Y, time variable t, initial state $Y(t_0)$, conversion functions Λ: $\{X(t), Y(t)\} \rightarrow Y(t+1)$ and outputs Δ: $\{X(t), Y(t)\} \rightarrow Z(t)$.

Hypothesis 2. The problem of determining the artificial intelligence system explainability can be reduced to a number of identifying the finite state machine with memory problems.

Hypothesis 3. Necessary and sufficient conditions for the artificial intelligence system explainability are determined by the presence conditions of the corresponding identifying sequences of the finite state machine with memory.

Another interesting intersection of the artificial intelligence systems explainability problem and the theory of identification and diagnosis of machines is related to models of fully and partially defined and completely and partially correct digital machines [28]. This work defines the concept of a specified fully defined and partially correct automaton as having either fully predictable behavior, including known input sets or sequences on which the automaton operates incorrectly (does not correspond to the specified input-output transformations), or has unpredictable behavior on certain sets.

Different classes of automata in these works are described by sets of relevant metrics, and, in our opinion, their use can be useful for the practical assessment of explainability and trustworthiness.

5. Conclusions

From the above analysis of the main theoretical provisions of heuristic self-organization systems and logical models, it follows that according to O. G. Ivakhnenko in systems of heuristic self-organization, the first task is to determine the content of factors “that determine the essence of different images”. These are images that characterize the objects of a particular subject area. After determining the composition and content of these images, the next problem is solved, namely, the problem of “generating a new successful heuristic”, which in content is a solution that leads to increased accuracy. It should be noted that it is talking about improving the accuracy of solving the problem of data processing, and heuristic self-organization systems are data processing systems. In this case, the multiplicity of existence of heuristics is allowed. It also follows from the above that heuristics in content correspond to the logical rules applied in heuristic self-organization systems.

The main provisions of the heuristic self-organization systems theory were developed by O. G. Ivakhnenko in the eighties of the last century, but they remain unnoticed to this day. At this time, the problem is to explain why the neural network makes such a decision and not another. Based on this, the concept of "explainable artificial intelligence" was introduced for artificial intelligence. It is clear that it is the content of heuristics that form the structure of the neural network in the form of logical rules, and determines the logic of the decision made.

It is also important that the inference rule, which is the basis for the construction of artificial neural networks, is an abductive rule, unfortunately, does not correspond to the fourth heuristic, as well as the actual definition of intelligence: intelligence is the ability to measure things. Unfortunately, none of the neural networks can measure things.

It is also clear why artificial intelligence, which is equivalent to natural intelligence, has not yet been created. For example, pattern recognition implements an abductive inference in the form of "concrete" "concrete", from concrete to concrete. The concrete existing image "single" is compared with the physically existing also "single" image. This method of inference does not involve the formation of the image "measure". He correlates the resulting image with the prototype.

Analysis of the basic inference rules content allows to conclude that the dialectical method of inference is general (generating) for the basic logical methods of inference. The difference lies in the composition and content of the middle member of the triangular relationship, namely, in the form of relationship elements combination: the transition from one concept to another. It is the ambiguity in the definition of the concepts combining forms of "reference" and "conclusion" generates the considered forms of inference according to [24]. All of them are characterized by the concept – "transition". The proposed dialectical method implements a form of combination "unity", which generates "measure". This method corresponds to the law of dialectical unity in the form of the of mutual penetration of opposites law [29]. That's right, mutual penetration and mutual conditionality of opposites are important. No struggle, no denial, the combination itself, the unity of opposites. In our case, it is their mutual combination. By the way, most importantly, the tool of such combination is intelligence.

At the same time, the fifth heuristic is based on the formation of a double measure of process and resource factors. It corresponds to the central pattern of integrative activity of the brain. The concept of "factor" is used to form the fifth heuristic. This concept is decisive in the formation of heuristic self-organization systems. It is the formation of the composition and content of factors that characterize the relevant subject area, and determines the level of compliance of knowledge models with knowledge about the existence and activities of the subject area objects. After all, only relations that are formed between objects of living and non-living nature, as well as between concepts in the processes of semantic thinking, can be known. Explainability of artificial intelligence refers to the laws of structure and activity of artificial neural networks. But modern theories of artificial neural networks ignore the existence of logical rules (heuristics), the content of which is established by O. G. Ivakhnenko. After all, only knowing the rules on the basis of which problems are solved, it is possible to check the correctness of the decision, and not by searching for such rules. According to [24], the search for appropriate logical rules in the structures of existing neural networks. According to O. G. Ivakhnenko's primary goal is to form heuristics, which form these logical rules. If heuristics are formed in an explicit form, then there is no point in looking for logical rules.

The three hypotheses about the explainability of artificial intelligence and the theory of machine identification can be further defined as statements or theorems and strictly proved. A separate problem is the formal definition and assessment of the trustworthiness of artificial intelligence systems.

Further consideration should be given to the possibility of constructing the model of an artificial neuron and an artificial neural network based on the central pattern of integrative brain activity, as well as the content of logical models based on the fifth heuristic and the possibility of their implementation in artificial intelligence models.

Contributions of authors: conceptualization and methodology – **Sergiy Dotsenko**, **V. Kharchenko**; investigation – **Sergiy Dotsenko** (sections 1-5), **V. Kharchenko** (sections 1, 3-5), **O. Morozova** (sections 2, 3), **A. Rucinski** (sections 4, 5), **Dotsenko** (section 2); writing original draft – **O. Morozova**, **V. Kharchenko**; review and editing – **Svitlana Dotsenko**; supervision – **Sergiy Dotsenko**, **V. Kharchenko**. All authors have read and agreed to the published version of the manuscript.

Acknowledgements. This work was supported by the ECHO project which has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement no 830943. The authors very appreciated to scientific society of consortium and in particular the staff of Department of Computer Systems, Networks and Cybersecurity of Na-

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

tional aerospace university KhAI for invaluable inspiration and creative analysis during the paper preparation.

References (GOST 7.1:2006)

1. Elger, Peter. AI as a Service. Serverless machine learning with AWS [Text] / Peter Elger, Eoin Shanaghy. – Manning, NY, USA, 2020. – 282 p.

2. Artificial Intelligence as a Service [Text] / S. Lins, K. D. Pandl, H. Teigeler et al. // Business Information Systems Engineering. – 2021. – Vol. 63. – P. 441-456. DOI: 10.1007/s12599-021-00708-w.

3. Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities [Text] / Christian Meske, Enrico Bunde, Johannes Schneider, Martin Gersch // Information Systems Management. – 2022. – Vol. 39, No. 1. – P. 53-63. DOI: 10.1080/10580530.2020.1849465.

4. A Survey on Artificial Intelligence (AI) and eXplainable AI [Text] / A. Degas, M. R. Islam, C. Hurter, S. Barua et al. // Air Traffic Management: Current Trends and Development with Future Research Trajectory. Applied Sciences. – 2022. – Vol. 12, No. 1295. – P. 1-47. DOI: 10.3390/app12031295.

5. Four Principles of Explainable Artificial Intelligence [Text] / P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski, Mark A. Przybocki. – Draft NISTIR 8312, 2020. – 24 p. DOI: 10.6028/NIST.IR.8312-draft.

6. Попов, Э. В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ [Текст] / Э. В. Попов. – М. : Наука, 1987. – 288 с.

7. Субботін, С. О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень: навчальний посібник [Текст] / С. О. Субботін. – Запоріжжя : ЗНТУ. 2008. – 341 с.

8. Brown, Carol. Introduction to artificial intelligence and expert systems. Artificial Intelligence/Expert Systems [Text] / Carol Brown, Daniel O'Leary. – Section of the American Accounting Association, 1995. – 14 p.

9. Chris, Naylor. Build Your Own Expert System [Text] / Chris Naylor. – Sigma Technical Press, 1983. – 249 p.

10. Доценко, С. І. Інтелектуальні системи: принципи евристичної самоорганізації [Текст] / С. І. Доценко // Радіоелектронні і комп'ютерні системи. – 2020. – № 1(93). – С. 4-16. DOI: 10.32620/reks.2020.1.01.

11. Доценко, С. І. Інтелектуальні системи: принципи евристичної самоорганізації процесів смислового мислення та смислової діяльності [Текст] / С. І. Доценко // Радіоелектронні і комп'ютерні системи. – 2020. – № 2(94). – С. 4-22. DOI: 10.32620/reks.2020.2.01.

12. Ивахненко, А. Г. Принятие решений на основе самоорганизации [Текст] / А. Г. Ивахненко, Ю. П. Зайченко, В. Д. Димитров. – М. : Советское радио, 1976. – 280 с.

13. Анохин, П. К. Принципиальные вопросы общей теории функциональных систем [Текст] / П. К. Анохин. – М. : Медицина, 1975. – 448 с.

14. Штейнберг, В. Э. Теория и практика дидактической многомерной технологии [Текст] / В. Э. Штейнберг. – М. : Народное образование, 2015. – 350 с.

15. Krogerus, Mikael, The Decision Book: 50 Models for Strategic Thinking [Text] / Mikael Krogerus, Roman Tschäppeler. – W. W. Norton & Company, 2012. – 176 p.

16. Luger, George F. Artificial intelligence : structures and strategies for complex problem solving [Text] / George F. Luger. – Pearson Education, 2011. – 784 p.

17. Thiebes, S. Trustworthy artificial intelligence [Text] / S. Thiebes, S. Lins, A. Sunyaev // Electron Markets, – 2021. – Vol. 31. – P. 447-464. DOI: 10.1007/s12525-020-00441-4.

18. ГОСТ Р 59276-2020 Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения [Текст]. – Введения з 2021-03-01. – 12 с.

19. Ивахненко, А. Г. Системы эвристической самоорганизации в технической кибернетике [Текст] / А. Г. Ивахненко. – Київ : «Техніка», 1971. – 372 с.

20. Haykin, Simon. Neural Networks and Learning Machines [Text] / Simon Haykin. – Pearson. 2009. – 936 p.

21. Four Principles of Explainable Artificial Intelligence [Text] / P. Jonathon Phillips, Carina A. Hahn, Peter C. Fontana, David A. Broniatowski, Mark A. Przybocki // Internal Report 8312. – National Institute of Standards and Technology Interagency, 2020. – 24 p. DOI: 10.6028/NIST.IR.8312-draft.

22. ДСТУ 2481-94. Системи оброблення інформації. Інтелектуальні інформаційні технології. Терміни та визначення: Системы обработки информации. Интеллектуальные информационные технологии [Текст]. – Чинний від 1995-01-01. – К. : Держстандарт України, 1994. – 72 с.

23. Доценко, С. І. Теоретичні основи створення інтелектуальних систем комп'ютерної підтримки рішень при управлінні енергозбереженням організацій [Текст] : дис. ... д-ра техн. наук : 05.13.06 / Харківський національний технічний університет сільського господарства імені Петра Василенка / С. І. Доценко. – Харків, 2017. – 369 с.

24. Ивахненко, А. Г. Самообучающиеся системы распознавания и автоматического управления [Текст] / А. Г. Ивахненко. – Киев : Техніка, 1969. – 392 с.

25. Knowledge Management Model Based Approach to Profiling of Requirements: Case for Information Technologies Security Standards [Text] / S. Dotsenko, O. Illiashenko, I. Budnichenko, V. Kharchenko // Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data. – 2021. – Vol. 84. – P. 255–277. DOI: 10.1007/978-3-030-65722-2_16.

26. Embedding an Integrated Security Management System into Industry 4.0 Enterprise Management: Cybernetic Approach [Text] / S. Dotsenko, O. Illiashenko, S. Kamenskyi, V. Kharchenko // Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data, – 2021. – Vol. 84. – P. 279–296. DOI: 10.1007/978-3-030-65722-2_17.

27. Lee, David. Principles and methods of testing finite state machines – a survey [Electronic resource] / David Lee Mihalis Yannakakis. – Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.380.3405&rep=rep1&type=pdf. (accessed 27.01.2022).

28. Yakimets, N. 7 Reliable FPGA-Based Systems out of Unreliable Automata: Multi-Version Design Using Genetic Algorithms [Text] / N. Yakimets, V. Kharchenko // Design of Digital Systems and Devices. Lecture Notes in Electrical Engineering. – Springer, Berlin, Heidelberg, 2011. – Vol. 79. – P. 165-192. DOI: 10.1007/978-3-642-17545-9_7.

29. Engels, Friedrich. Dialectics of Nature [Text] / Friedrich Engels. – Lawrence & Wishart Ltd, 1977. – 496 p.

References (BSI)

1. Peter, Elger, Eoin, Shanaghy. *AI as a Service. Serverless machine learning with AWS*. Manning, NY, USA, 2020. 282 p.

2. Lins, S., Pandl, K. D., Teigeler, H. et al. Artificial Intelligence as a Service. *Business Information Systems Engineering*, 2021, vol. 63, pp. 441-456. DOI: 10.1007/s12599-021-00708-w.

3. Meske, Christian., Bunde, Enrico. et al. Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management*, 2022, vol. 39, no. 1, pp. 53-63. DOI: 10.1080/10580530.2020.1849465.

4. Degas, A., Islam, M.R., Hurter, C.; Barua, S. et al. A Survey on Artificial Intelligence (AI) and eXplainable AI. *Air Traffic Management: Current Trends and Development with Future Research Trajectory. Applied Sciences*. 2022, vol. 12, no. 1295. pp. 1-47. DOI: 10.3390/app12031295.

5. Phillips, P. Jonathon., Hahn, Carina A., Fontana, Peter C. et al. *Four Principles of Explainable Artificial Intelligence*. Draft NISTIR 8312, 2020. 24 p. DOI: 10.6028/NIST.IR.8312-draft.

6. Popov, E. V. *Ekspertnyye sistemy: Reshe-niye neformalizovannykh zadach v dialoge s EVM* [Expert systems: Resolving non-formalized problems in dialogue with computers]. Moscow, Nauka Publ., 1987, 288 p.

7. Subbotin, S. O. *Podannya y obrobka znan' u systemakh shtuchnoho intelektu ta pidtrymky pryynyattya rishen': Navchal'nyy posibnyk* [Representation and processing of knowledge in artificial intelligence systems and decision support: A textbook]. Zaporozhye, ZNTU Publ., 2008. 341 p.

8. Brown, Carol, O'Leary, Daniel. *Introduction to artificial intelligence and expert systems. Artificial Intelligence/Expert Systems.* Section of the American Accounting Association, 1995. 14 p.

9. Naylor, Chris. *Build Your Own Expert System*. Sigma Technical Press, 1983. 249 p.

10. Dotsenko, S. Intelektualʹni systemy: pryntsypy evrystychnoyi samoorhanizatsiyi [Intellectual systems: a principle of heuristic self-organization]. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2020, no. 1(93), pp. 4-16. DOI: 10.32620/reks.2020.1.01.

11. Dotsenko, S. Intelektualʹni systemy: pryntsypy evrystychnoyi samoorhanizatsiyi protsesiv smyslovoho myslennya ta smyslovoyi diyalʹnosti [Intellectual systems: principles of heuristic self-organization of the processes of sense thinking and sense activity]. *Radioelektronni i komp'uterni sistemi – Radioelectronic and computer systems*, 2020, no. 2(94), pp. 4-22. DOI: 10.32620/reks.2020.2.01.

12. Ivakhnenko A. G., Zaichenko, Yu. P., Dimitrov, V. D. *Prinyatie reshenii na osnove samoorganizatsii* [Decision making on the basis of self-organization]. Moscow, Sovetskoe radio Publ., 1976. 280 p.

13. Anokhin, P. K. *Printsipial'nyye voprosy obshchey teorii funktsional'nykh sistem. V kn. Ocherki po fiziologii funktsional'nykh sistem* [Fundamental questions of the general theory of functional systems. In book. Essays on the physiology of functional systems]. Moscow, Medicine Publ., 1975. 448 p.

14. Steinberg, V. E. *Teoriya i praktika didakticheskoy mnogomernoy tekhnologii* [Theory and practice of didactic multidimensional technology]. Moscow, Public education Publ., 2015. 350 p.

15. Krogerus, Mikael., Tschäppeler, Roman. *The Decision Book: 50 Models for Strategic Thinking*. W. W. Norton & Company Publ., 2012. 176 p.

16. Luger, George F. *Artificial intelligence : structures and strategies for complex problem solving*. Pearson Education, 2011. 784 p.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

17. Thiebes, S., Lins, S. & Sunyaev, A. Trustworthy artificial intelligence. *Electron Markets*, 2021, vol. 31, pp. 447-464. DOI: 10.1007/s12525-020-00441-4.

18. GOST R 59276-2020 *Sistemy iskusstvennogo intellekta Sposoby obespecheniya doveriya. Obshchiye polozheniya* [Artificial intelligence systems. Methods for ensuring trust. General]. Introduction date 2021-03-01. 12 p.

19. Ivakhnenko, A. G. *Sistemy evristicheskoy samoorganizatsii v tekhnicheskoy kibernetike* [Systems of heuristic self-organization in technical cybernetics], Kiev, Tehnika Publ., 1971. 372 p.

20. Haykin, Simon. *Neural Networks and Learning Machines*, Pearson, 2009. 936 p.

21. Phillips, P. Jonathon., Hahn, Carina A., Fontana, Peter C., Broniatowski, David A., Przybocki, Mark A.. *Four Principles of Explainable Artificial Intelligence National Institute of Standards and Technology Interagency*. Internal Report 8312, 2020. 24 p. DOI: 10.6028/NIST.IR.8312-draft.

22. DSTU 2481-94. *Systemy obroblennya informatsiyi. Intelektual'ni informatsiyni tekhnolohiyi. Terminy ta vyznachennya: Systemy obrabotky ynformatsyy. Yntellektual'nye ynformatsyonnye tekhnolohyy* [Information processing systems. Intelligent information technologies. Terms and definitions: Information processing systems. Intelligent Information Technologies]. Regional edition 1995-01-01, Kyiv, Derzhstandart of Ukraine Publ., 1994. 72 p.

23. Dotsenko, S. I. *Teoretychni osnovy stvorennya intelektual'nykh system komp''yuternoyi pidtrymky rishen' pry upravlinni enerhozberezhennyam orhanizatsiy. Diss. dokt. tekhn. nauk* [Theoretical Foundations for Creating Intelligent Computer Support Systems for Managing Energy Saving Organizations Dr. eng. sci. diss.]. Kharkov, Kharkivs'kyy natsional'nyy tekhnichnyy universytet sil's'koho hospodarstva imeni Petra Vasylenka Publ., 2017. 369 p.

24. Ivakhnenko, A. G. *Samoobuchayushchiyesya sistemy raspoznavaniya i avtomaticheskogo upravleniya* [Self-learning recognition and automatic control systems]. Kiev, Technique Publ., 1969. 392 p.

25. Dotsenko, S., Illiashenko, O., Budnichenko, I., Kharchenko, V. Knowledge Management Model Based Approach to Profiling of Requirements: Case for Information Technologies Security Standards. *Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data*, 2021, vol. 84, pp. 255–277. DOI: 10.1007/978-3-030-65722-2_16.

26. Dotsenko, S., Illiashenko, O., Kamenskyi, S., Kharchenko, V. Embedding an Integrated Security Management System into Industry 4.0 Enterprise Management: Cybernetic Approach. *Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data*, vol 84, 2021, pp. 279–296. DOI: 10.1007/978-3-030-65722-2_17.

27. Lee, David., Yannakakis, Mihalis. *Principles and methods of testing finite state machines – a survey*. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.380.3405&rep=rep1&type=pdf. (accessed 27.01.2022).

28. Yakimets, N., Kharchenko, V. 7 Reliable FPGA-Based Systems out of Unreliable Automata: Multi-Version Design Using Genetic Algorithms. *Design of Digital Systems and Devices. Lecture Notes in Electrical Engineering*. Springer, Berlin, Heidelberg, 2011, vol. 79, pp. 165-192. DOI: 10.1007/978-3-642-17545-9_7.

29. Engels, Friedrich. *Dialectics of Nature*. Lawrence & Wishart Ltd, 1977. 496 p.

Надійшла до редакції 12.11.2021, розглянута на редколегії 16.02.2022

ЕВРИСТИЧНА САМООРГАНІЗАЦІЯ ПРЕДСТАВЛЕННЯ ТА ФОРМУВАННЯ ЗНАНЬ: АНАЛІЗ В КОНТЕКСТІ ПОЯСНЮВАНОГО ШТУЧНОГО ІНТЕЛЕКТУ

С. І. Доценко, В. С. Харченко, О. І. Морозова, А. Русинські, С. О. Доценко

З аналізу основних теоретичних положень систем евристичної самоорганізації (СЕС) та логічних моделей слідує, що згідно О. Г. Івахненку в СЕС першою є задача визначення змісту факторів «які визначають сутність різних образів». Мова йде про образи, які характеризують об’єкти певної предметної області. Після визначення складу та змісту цих образів вирішується наступна задача, а саме, задача «генерування нової вдалої евристики» яка за змістом є рішенням, що призводить до підвищення точності. Слід зауважити, що мова йде про підвищення точності вирішення задачі оброблення даних. З наведеного слідує, що СЕС є системами обробки даних. При цьому допускається множинність існування евристик. Евристики за змістом відповідають логічним правилам, які застосовуються у СЕС. Основні положення теорії СЕС були розроблені О. Г. Івахненком ще у восьмидесятих роках минулого століття але вони залишаються поза увагою до цього часу. На цей час ставиться задача пояснення, чому нейронна мережа приймає саме таке рішення а не інше. Виходячи з цього, для штучного інтелекту введено поняття «пояснюваний штучний інтелект». Зрозуміло, що саме зміст евристик, які формують структуру нейронної мережі у формі логічних правил, і визначає логіку рішення, яке приймається. Встановлено, що правилом виведення, яке покладено у основу побудови штучних нейронних мереж, є абдуктивне правило, яке, на жаль, не відповідає четвертій евристиці, а також не відповідає визначенню інтелекту: інтелект – це здібність реалізувати процес виміру речі. На жаль, жодна з нейронних мереж не здатна вимірювати речі. З аналізу змісту основних правил виведення слідує, що діалектичний метод виведення, є загальним (породжувальним) для основних логічних методів виводу. Різниця полягає у складі та змісті середнього члену триарного відношення, а саме, у формі поєднання елементів відношення: переході від одного поняття до іншого. Пояснюваність штучного інтелекту стосується закономірностей побудови та функціонування штучних нейронних мереж. Однак сучасні теорії штучних нейронних мереж ігнорують існування логічних правил (евристик), зміст яких встановлено О. Г. Івахненком. Адже, тільки знаючи правила, на основі яких вирішуються задачі, можливо перевірити правильність отриманого рішення, а не шляхом пошуку таких правил. Сформовано три гіпотези щодо пояснюваності штучного інтелекту і теорії ідентифікації автоматів, які можуть бути визначені як твердження і строго доведені.

Ключові слова: евристика; самоорганізація; знання; логічне виведення; пояснюваний штучний інтелект.

ЭВРИСТИЧЕСКАЯ САМООРГАНИЗАЦИЯ ПРЕДСТАВЛЕНИЯ И ФОРМИРОВАНИЯ ЗНАНИЙ: АНАЛИЗ В КОНТЕКСТЕ ОБЪЯСНИМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

С. И. Доценко, В. С. Харченко, О. И. Морозова, А. Русински, С. А. Доценко

Из анализа основных теоретических положений систем эвристической самоорганизации (СЭС) и логических моделей следует, что согласно А. Г. Ивахненко в СЭС первой задачей является определение содержания факторов, «определяющих сущность различных образов». Речь идет об образах, характеризующих объекты определенной предметной области. После определения состава и содержания этих образов решается следующая задача, а именно задача «генерирования новой удачной эвристики», которая по содержанию является решением, приводящим к повышению точности. Следует заметить, что речь идет о повышении точности решения задачи обработки данных. Из приведенного следует, что СЭС являются системами обработки данных. При этом допускается множественность существования эвристик. Эвристики по содержанию соответствуют логическим правилам, применяемым в СЭС. Основные положения теории систем эвристической самоорганизации были разработаны А. Г. Ивахненко еще в восьмидесятых годах прошлого столетия, но они остаются без внимания до сих пор. В настоящее время ставится задача объяснения, почему нейронная сеть принимает именно такое решение, а не другое. Исходя из этого, для искусственного интеллекта введено понятие «объяснимый искусственный интеллект». Понятно, что именно содержание эвристик, формирующих структуру нейронной сети в форме логических правил, и определяет логику принимаемого решения. Установлено, что правилом вывода, положенным в основу построения искусственных нейронных сетей, является абдуктивное правило, которое, к сожалению, не соответствует четвертой эвристике, а также не соответствует определению интеллекта: интеллект – это способность реализовать процесс измерения вещи. К сожалению, ни одна из нейронных сетей не способна измерять вещи. Из анализа содержания основных правил вывода следует, что диалектический метод вывода является общим (порождающим) для основных логических методов вывода. Разница состоит в составе и содержании среднего члена триарного отношения, а именно, в форме сочетания элементов отношения: переходе от одного понятия к другому. Объяснимость искусственного интеллекта касается закономерностей построения и деятельности искусственных нейронных сетей. Однако современные теории искусственных нейронных сетей игнорируют существование логических правил (эвристик), содержание которых установлено А. Г. Ивахненко. Ведь, только зная правила, на основе которых решаются задачи, возможно проверить правильность полученного решения, а не путем поиска таких правил. Сформированы три гипотезы по поводу объяснимости искусственного интеллекта и теории идентификации автоматов, которые могут быть определены как утверждения и строго доказаны.

Ключевые слова: эвристика; самоорганизация; знание; логический вывод; объяснимый искусственный интеллект.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Доценко Серій Ілліч – д-р техн. наук, доц., доц. каф. спеціалізованих комп’ютерних систем, Український державний університет залізничного транспорту, Харків, Україна.

Харченко Вячеслав Сергійович – д-р техн. наук, проф., зав. каф. комп’ютерних систем, мереж і кібербезпеки, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Морозова Ольга Ігорівна – д-р техн. наук, доц., проф. каф. комп’ютерних систем, мереж і кібербезпеки, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Русинські Анжей – PhD, проф. каф. електричної інженерії, Університет Нью-Хемршир, Сполучені Штати Америки.

Доценко Світлана Олексіївна – д-р пед. наук, доц., зав. каф. інформаційних технологій, Харківський національний педагогічний університет, Харків, Україна.

Sergiy Dotsenko – Doctor of Technical Sciences, Associate Professor, Associate Professor at the Department of specialized computer systems, Ukrainian State University of Railway Transport, Kharkiv, Ukraine,
e-mail: sirius_3k3@ukr.net, ORCID: 0000-0003-3021-4192.

Vyacheslav Kharchenko – Doctor of Technical Science, Professor, Head of the Department of Computer Systems, Networks and Cybersecurity, National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine,
e-mail: v.kharchenko@csn.khai.edu, ORCID: 0000-0001-5352-077X.

Olga Morozova – Doctor of Technical Science, Associate Professor, Professor of the Department of Computer Systems, Networks and Cybersecurity, National Aerospace University “Kharkiv Aviation Institute”, Kharkiv, Ukraine,
e-mail: o.morozova@csn.khai.edu, ORCID: 0000-0001-7706-3155.

Andrzej Rucinski – PhD, Professor at the Department of Electrical Engineering, University of New Hampshire, New Hampshire, USA,
e-mail: Andrzej.Rucinski@unh.edu, ORCID: 0000-0002-0988-7336.

Svitlana Dotsenko – Doctor of Pedagogy Sciences, Professor at the Information Technology Department, H. S. Skovoroda Kharkiv National Pedagogical University, Kharkiv, Ukraine,
e-mail: dozenkosveta@gmail.com, ORCID: 0000-0002-4501-9130.