REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

DEVELOPMENT OF A FUZZY SEMANTIC APPROACH FOR EXPLAINABLE ARTIFICIAL INTELLIGENCE

Specialty:	3338.01 – "System analysis, control and information processing" (data processing)
Field of science:	Technical sciences

Applicant: **Pavel Igorevich Kosov**

Baku – 2026

The work was performed at the "Computer Engineering" department of the Azerbaijan State Oil and Industry University.

Scientific supervisors: Doctor of Technical Sciences, professor
Latafat Abbas Gardashova

Professor of Computer Science
Cecilia Zanni-Merk

Official opponents: Doctor of Technical Sciences, professor
Alakbar Ali Agha Aliyev

Doctor of Technical Sciences, professor
Ramin Rza Rzayev

Doctor of Technical Sciences, professor
Javanshir Firudin Mammadov

Dissertation council FD 2.48 of the Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Azerbaijan State Oil and Industry University

Chairman of the
Dissertation council: Corresponding member of ANAS,
Doctor of Technical Sciences, professor

Rafik Aziz Aliev

Scientific secretary of
the Dissertation council: Doctor of Technical Sciences,
associate professor

Akif Vali Alizadeh

Chairwoman of the
Scientific seminar: Doctor of Technical Sciences, professor

İMZANI TƏSDİQ EDİRƏM
ADNSU-nun Elmi katibi

Kamala Rafik Aliyeva

GENERAL CHARACTERISTICS OF THE WORK

Relevance of the topic. Today, in many fields of human activity, there is an increase in the complexity and performance of artificial intelligence (AI) models, the study of which is addressed by the field of machine learning and neural networks. Billions of parameters demonstrate the effectiveness of such systems, which has led to their widespread adoption across industries, from healthcare and law to manufacturing. At the same time, there is a growing need to create systems that comply with the principles of responsible AI, whose main task is the development and study of reliable, fair, and accountable AI systems.

The use of high-performance but opaque AI models operating as “black boxes” within the framework of responsible AI gives rise to many problems. The main output of such models is a prediction without the ability to explain it. However, a necessary condition for creating a responsible system is the ability to explain and justify the decisions obtained. The use of powerful AI models in itself is not a sufficient condition for building trusted systems; this creates a need for research into explainable AI (XAI) methods capable of overcoming the “black box” problem.

Existing XAI methods, in turn, face two fundamental limitations: algorithmic approaches lack semantic depth, providing explanations at the feature level without their meaningful interpretation, while classical ontological approaches are unable to adequately model the uncertainty of the real world. This makes it relevant to develop new methods that combine the semantic power of ontologies and the mathematical tools of fuzzy logic in order to create a new generation of XAI systems.

Object and subject of the research. The object of the research is the processes, models, and methods for generating explanations for decisions made by AI systems, namely “black-box” models. The research is grounded in issues related to the development and application of a methodology for building XAI based on the integration of semantic technologies and fuzzy logic. The subject of the research includes semantic properties represented in an ontology,

methods for formalizing fuzzy knowledge based on Fuzzy OWL2, as well as algorithms and architecture for generating explanations with degrees of confidence and for evaluating them.

Purpose and objectives of the work. The purpose of the work is a comprehensive analysis of existing XAI approaches, as well as the development and study of new approaches to building XAI based on the integration of semantic technologies and fuzzy logic in order to ensure its practical applicability for increasing the reliability, depth, and comprehensibility of explanations produced by "black-box" models. Achieving this goal involves the following main objectives:

- Analysis of existing XAI systems to identify key limitations, the elimination of which necessitates the development of new methods;
- Development of the concept of semantic properties in an ontology for the universal representation of expert knowledge;
- Development of the "fuzzy explainability" method for modeling imprecise knowledge and generating explanations with degrees of confidence;
- Creation of a flexible architecture of an explainable system for integrating machine learning models with a fuzzy ontology;
- Development of a procedure for processing fuzzy clustering results for representing data properties in an ontology;
- Conducting a series of computational experiments to test and evaluate the proposed approaches on heterogeneous datasets.

Research methods. The stated objectives were addressed through the application of fuzzy set theory, knowledge engineering methods, ontology and semantic network modeling, machine learning and deep learning, cluster analysis, computer modeling methods, mathematical computational experiments, as well as expert evaluation and comparative analysis methods.

Main provisions submitted for defense. The following main provisions and results of the dissertation are submitted for defense:

- The concept of "explanatory" properties, representing a new method of ontological formalization of expert knowledge that ensures applicability to different data types;

- The methodology of “fuzzy explainability,” which makes it possible to model the uncertainty and vagueness of real-world data and knowledge using Fuzzy OWL2 fuzzy ontologies and to generate explanations with quantitative degrees of confidence;
- A flexible method and architecture of an explainable system that make it possible to apply semantic-fuzzy explanation to any existing “black-box” models without compromising their predictive accuracy and interpretability;
- A comprehensive methodology for assessing the quality of generated fuzzy semantic explanations, including function-oriented, human-oriented, and hybrid approaches.

Scientific novelty. The main essence of the scientific novelty of this dissertation lies in the following:

- For the first time, abstract semantic “explanatory” properties have been created for the ontological representation of expert knowledge in explainable artificial intelligence;
- For the first time, explanations in an environment of imprecise knowledge have been developed using newly created “explanatory” properties for a fuzzy OWL2 ontology;
- For the first time, a method has been developed for explaining “black-box” results based on different data types represented as fuzzy knowledge;
- The results of fuzzy clustering of data properties have been processed for representation in an ontology;
- Development of computer simulation, analysis, and evaluation of the original results obtained on the basis of the proposed approaches.

Scientific and practical significance. The results obtained in the dissertation have both theoretical and practical significance. The use of a fuzzy approach in the work makes it possible to take uncertainty into account. The dissertation presents a comprehensive methodology for creating explainable artificial intelligence systems that makes it possible to synthesize explanation modules for high-performance “black-box” models. The advantage of this methodology

lies in its flexibility and universality; in its ability to generate semantically rich and intuitively understandable explanations; in its ability to adequately model uncertainty; as well as in increasing trust in artificial intelligence systems and improving the quality of joint human-machine decision-making. The presented methodology was tested on datasets from various subject areas (finance, medicine, materials science, cybersecurity, computer vision, and social analytics). For most components of the methodology, a software prototype was developed demonstrating the possibility of its practical implementation.

Approbation of the work. The main provisions of the work were presented at the following conferences and seminars:

- First Research Workshop of the Computer Science Research Department PhD candidates. UFAZ, 2024;
- Ümummilli Lider Heydər Əliyevin 101 illiyinə həsr olunmuş Respublika Elmi Konfransı. ADNSU, 2024;
- 28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES-2024). Seville, Spain, 2024;
- The 2024 International Conference on Decision Aid Sciences and Applications (DASA). Manama, Bahrain, 2024;
- X Международная Научная Конференция «Информационные Технологии Интеллектуальной Поддержки Принятия Решений» (ITIDS'2024). Уфа, Россия, 2024;
- Gənc Tədqiqatçıların və Doktorantların "Elm Günü"-nə həsr olunmuş Böyük Elmi Seminarı. ADNSU, 2025;
- Second Research Workshop of the Computer Science Research Department PhD candidates. UFAZ, 2025;
- Ümummilli Lider Heydər Əliyevin 102 illiyinə həsr olunmuş Respublika Elmi Konfransı. ADNSU, 2025.

Publications. 10 scientific works on the dissertation topic have been published (4 of them without co-authors), including 6 articles and 4 conference abstracts. 5 works are indexed in international databases. 1 article is included in the Scopus database (Q2 category), and 1

conference abstract is included in the proceedings of an international conference indexed in Web of Science and Scopus. Other international publications include: 1 article in a journal recommended by the Higher Attestation Commission of the Russian Federation (K2 category), 1 article in a journal recommended by the Higher Attestation Commission of the Republic of Uzbekistan, and 1 conference abstract in the proceedings of an international conference indexed in the RSCI database.

Name of the organization where the dissertation was performed. The work was performed at the "Computer Engineering" department of the Azerbaijan State Oil and Industry University.

Structure and volume of the work. The main body of the work is presented on 160 pages and consists of an introduction, 5 chapters, a conclusion, a list of references, and a list of abbreviations. It also contains 27 figures, 10 tables, and 150 items in the bibliography. Excluding images, tables, graphs, appendices, bibliography, and spaces in the text, the approximate volume of the work is as follows: Introduction – 8,000 characters, Chapter I – 25,000 characters, Chapter II – 45,000 characters, Chapter III – 40,000 characters, Chapter IV – 40,000 characters, Chapter V – 40,000 characters, and Conclusion – 2,500 characters. In total, the dissertation consists of approximately 200,000 characters.

CONTENT OF THE WORK

In the **introduction** was justified the relevance of the research topic, formulates the object, subject, purpose, and objectives of the dissertation, and defines the research methods, the provisions submitted for defense, the scientific novelty, and the scientific and practical significance of the results obtained. In addition, it provides information on the approbation of the work, publications on the research topic, as well as the structure and scope of the dissertation.

In the **first chapter** was analyzed the current state of methodologies in eXplainable Artificial Intelligence (XAI), substantiates the necessity and relevance of the research, and formulates the scientific problem of the dissertation.

The chapter shows that an e***xplanation*** within XAI should be considered as providing the user with information about the model's decision together with those factors and dependencies that make this decision understandable and verifiable.

Based on the above, the key limitations of existing approaches were identified: insufficient consideration of domain knowledge, weak semantic interpretation of features, and insufficient adaptation of explanations to the cognitive characteristics of the user.

Based on the analysis carried out in the ***chapter, the main scientific problem of the*** dissertation is the development and study of a new methodology for building XAI systems based on semantic technology and fuzzy logic, aimed at increasing the semantic richness, reliability, and interpretability of explanations generated for "black-box" machine learning (ML) models, as well as enabling work with fuzzy and incomplete knowledge**.**

In the **second chapter** was justified the use of semantic technologies and fuzzy logic as the formal basis of the proposed approach. Key concepts and approaches are described, and various aspects of fuzzy logic, fuzzy inference systems, and the Fuzzy OWL2 fuzzy ontology are analyzed.

The Semantic Web is considered as a set of standards of the World Wide Web Consortium (W3C) that ensures entity identification, knowledge representation, and querying. Its basic components include URIs for the unique identification of resources, RDF for representing knowledge in the "subject-predicate-object" triple form, SPARQL for querying RDF data, and ontologies as a means of specifying a vocabulary of concepts and their semantic relationships.

In computer science, ontology is used as a means of structured and formalized knowledge representation. It has a hierarchical organization (shown in Figure 1) and can be considered at the levels of upper, basic, and domain ontology[1].

[1] Navigli, R., Velardi, P., Gangemi, A. Ontology learning and its application to automated terminology translation // IEEE Intelligent Systems, – 2003. vol. 18, № 1, – p. 22-31.

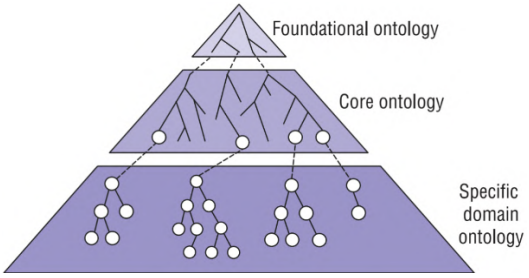

Figure 1. Three main levels of ontology in computer science[1]

Formally, an ontology is defined as a tuple[2] including a set of concepts, a set of properties, and a system of axioms fixing the relations and constraints of the domain by means of a formula $O = (C, \leq_C, R, \leq_R, A^O)$, where ontology O includes partially ordered $\leq_C$ concepts C, properties R with a partial order $\leq_R$, and axioms A^O.

The relevance of using ontologies in XAI is described, and it is established that an ontology makes it possible to connect machine learning results with formally defined concepts and relations of the domain, integrate symbolic knowledge with data, and generate explanations understandable to humans.

The mathematical foundation of modern ontologies is ***description logics (DL),*** which provide a strict yet interpretable description of a domain. An ontology built on the basis of description logics includes *TBox*, which contains axioms about concepts and their *rela*tionships (for example, inclusion $C1 \sqsubseteq C2$, where $C1$ is a sub concept of $C2$; equivalence $C1 \equiv C2$); *ABox*, which contains assertions about individuals, their membership in concepts (for example, $a: C$), and the relations between them through roles; and *RBox*, which contains axioms about roles and their characteristics (for example, hierarchy $R1 \sqsubseteq R2$ and transitivity).

For example, in DL, the ALUEC language is the basic AL language extended with union (U), full existential quantification (E),

[2] Maedche, A., Staab, S. Measuring Similarity between Ontologies // Proceedings of the 13 International Conference on Knowledge Engineering and Knowledge Management, – Siguenza: – 1-4 October, – 2002, – p. 251-263.

and negation of arbitrary concepts (C). The names of more complex logics are formed according to the same principle. For example, SHOIN is the logic S (ALC + transitivity), extended with role hierarchies (H), nominals (O), inverse roles (I), and number restrictions (N). This is discussed in more detail in Table 1[3].

Table 1
Types of some different DLs[3]

Sign	Meaning
ALC	Attributive language with complement (negation) of arbitrary concepts.
S	Transitivity axioms for roles.
E	Full existential quantification ($\exists R.C$).
U	Union of concepts ($\sqcup$).
H	Role hierarchy (sub-roles).
R	Additional axioms for roles (reflexivity, irreflexivity, and disjointness).
O	Nominals (concepts consisting of a single individual).
N	Cardinality restrictions (unqualified).
Q	Qualified cardinality restrictions.
I	Inverse roles ($R-$).
F	Functional properties for roles.

An essential feature of this logic is th***e Open World Assumption (OWA),*** under which the absence of information is regarded as incompleteness of knowledge rather than the falsity of statements. In this regard, OWL2[4] is considered as a basic means of ontological representation.

OWL 2 DL is the most expressive of the standards and is based on the SROIQ(D) logic. This logic extends SHOIN(D) with complex role axioms (R) and qualified cardinality restrictions (Q). Fuzzy

[3] Rudolph, S. Foundations of Description Logics // Lecture Notes in Computer Science, – 2011. vol. 6848, – p. 76-136.
[4] W3C. OWL 2 Web Ontology Language Document Overview (Second Edition): [Electronic resource] / World Wide Web Consortium. – 2012. URL: https://www.w3.org/TR/owl2-overview

OWL2[5] is a fuzzy extension for describing fuzzy concepts, roles, and axioms. Fuzzy OWL2 has the following properties[4]:

- *Fuzzy concepts* – membership of an individual in a fuzzy class by degree;
- *Fuzzy roles* – *m*embership relations between individuals or between individuals and data values;
- *Fuzzy data type*s – define fuzzy sets over standard data types;
- *Fuzzy modifiers* – apply linguistic modifiers;
- *Fuzzy axioms* – extend standard OWL2 axioms, allowing them to have a degree of truth:
  - *Fuzzy membership assertions*.
  - *Fuzzy role assertions*.
  - *Fuzzy inclusion axioms*.
  - *Other fuzzy axioms* – extensions for class equivalence, disjointness, etc.

The integration of Fuzzy OWL2 into XAI systems opens significant opportunities for creating more flexible, semantically rich, and reliable explanations[6,7].

In the **third chapter** was proposed a methodology for modeling new semantic properties aimed at increasing the explainability of XAI systems. As a key element, ***"explanatory" properties*** are proposed, representing semantic attributes formed not only on the basis of observable data characteristics, but also on the basis of logical inference, expert knowledge, users' mental models (MMs), and similarities between objects.

The general scheme for constructing an ontology-oriented XAI system is presented in Figure 2. It reflects the sequence of transition

---

[5] Bobillo, F., Straccia, U. An OWL Ontology for Fuzzy OWL 2 // Lecture Notes in Computer Science, – 2009. vol. 5722, – p. 151-160.

[6] Косов, П.И. Разработка Нечёткой Онтологии для Объяснимого Искусственного Интеллекта для Принятия Решений в Нечёткой Среде // – Tashkent: Chemical Technology, Control and Management, – 2025. № 2, –p. 19-26.

[7] Косов, П.И., Гардашова, Л.А. Повышение достоверности объяснимого искусственного интеллекта посредством нечеткой логики и онтологии // – Воронеж: Моделирование, Оптимизация и Информационные Технологии, – 2025. т. 13, № 2(49), – с. 1-11.

from expert knowledge and raw data to the formation of a semantically interpretable explanation.

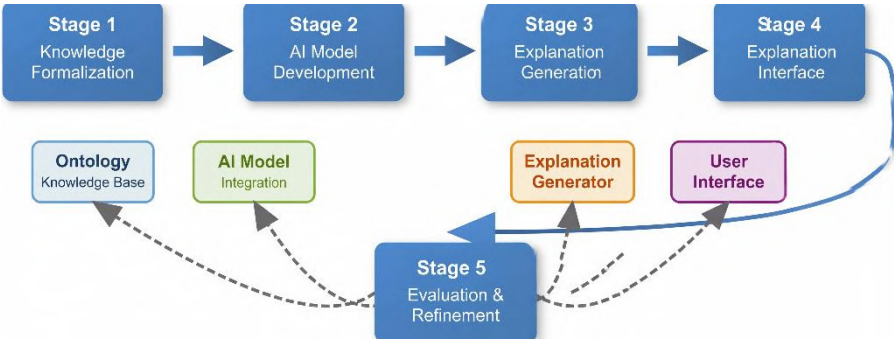

Figure 2. Scheme for creating XAI based on ontologies

It was determined that creating an effective XAI explanation is useless if it does not resonate with the user's cognitive structures, in which ***mental models (MMs)*** occupy a central place. In the proposed approach, mental models are considered a necessary component: the user model sets requirements for the form and content of the explanation, whereas the expert model serves as the source of the concepts, features, and relations included in the ontology.

Distributed MMs improve coordination and communication, make it possible to predict each other's actions, adapt more quickly to changes, and make more effective joint decisions[8]. Mathematically, this can be represented as a non-empty intersection of the sets representing the MMs of team members $M_A \cap M_B \cap M_C \neq \emptyset$ where M_A, M_B, M_C are the MMs of individual persons. If there is no such common intersection $M_A \cap M_B \cap M_C = \emptyset$ the team does not possess a fully shared model.

[8] Burtscher, M.J., Manser, T. Team mental models and their potential to improve teamwork and safety: A review and implications for future research in healthcare // Safety Science, – 2012. vol. 50, № 5, – p. 1344-1354.

On this basis, a procedure for selecting "explanatory" properties was proposed, including domain analysis, the involvement of expert knowledge, consideration of the user's level of training, analysis of the data and task context, as well as iterative adjustment of the set of properties based on the results of evaluating their interpretive usefulness.

The proposed concept *of "explanatory" properties*[9,10] represents semantic attributes that go beyond the development of semantically enriched XAI. These properties are semantic attributes that go beyond a simple description of observable characteristics and are built on the basis of:

- Logical inferences regarding data and their interrelations;
- Expert knowledge in a specific domain;
- Users' MMs, taking into account their level of expertise and way of perceiving information;
- Identified similarities and analogies between objects or data instances.

The main goal of "explanatory" properties is to provide a deeper and more intuitive justification of decisions made by ML models by taking into account not only explicit but also implicit, inferred, or contextual aspects o*f different dat*a types.

Key characteristics and advantages of "explanatory" properties:

- *Universality and flexibil*ity. The possibility of building uniform explanatory mechanisms for heterogeneous data;
- *Grounding in expert knowledge and MMs.* They are formed with consideration of how experts and users (with different levels of training) conceptualize data and interpret information,

[9] Kosov, P. Advancing XAI: new properties to broaden semantic-based explanations of black-box learning models / P. Kosov, N. El Kadhi, C. Zanni- Merk, L. Gardashova // Procedia Computer Science, – 2024. vol. 246, – p. 2292-2301.

[10] Kosov, P., El Kadhi, N., Zanni-Merk, C., Gardashova, L. Semantic-Based XAI: Leveraging Ontology Properties to Enhance Explainability // Proceedings of the 2024 International Conference on Decision Aid Sciences and Applications, – Manama: – 11-12 December, – 2024, – p. 1-5.

which increases the relevance and comprehensibility of the generated explanations;

- *Contextuality and subjectivity*. More personalized and adapted explanations are created rather than striving for a universal but potentially less relevant solution;
- *A basis for deep and substantive explanations. They mak*e it possible to form fuller, more accurate, and more informative explanations of AI model decisions.

"Explanatory" properties were defined as a set $P_{\text{exp}} \subseteq R$, where is a R set of relations and each property $p \in P_{\text{exp}}$ can contribute to the explanation of a specific case d_i with a fuzzy degree of membership. $\mu_p(d_i) \in [0,1]$. These properties serve as semantic building blocks for constructing explanations within the proposed framework. Such properties must satisfy several criteria:

- Accessibility for both experts and non-specialists;
- The possibility of representing ambiguous or uncertain data;
- Compatibility with ontological inference under OWA;
- Simplicity and interpretability.

A key aspect of ontological representation is the clear definition of Domains and Ranges for each "explanatory" property. For example, for the property hasWeatherType, the domain may be the class Clothes, and the range may be the class WeatherType. The use of both positive and negative property restrictions (for example, Sandal that Not hasWeatherType some Cold) makes it possible to create more precise and detailed semantic descriptions in the ontology.

The property selection procedure is a methodological approach based on principles and stages for identifying the most relevant semantic attributes in data in order to generate meaningful explanations:

- In-depth domain analysis and the involvement of expert knowledge;
- User orientation and consideration of MMs;
- Analysis of the data and task context;
- Targeted focus on improving explainability;
- Iterativeness and adaptability of the process;

– Consideration of the capabilities of ontological modeling;
– Recognition of subjectivity and contextuality.

Further in the chapter, the foundation of ***fuzzy explainability*** through distributed MMs is laid. Our study led to the conclusion that MMs contribute to the representation of fuzzy explainability and aim to generate explanations that:

– *Correspond to cognitive representatio*ns: by using linguistic variables (“high,” “low,” “probably”) and degrees of confidence, fuzzy explanations can be mapped directly onto fuzzy concepts in the user’s mental model.
– *Reflect real uncertainty:* instead of hiding or ignoring the uncertainty inherent in the data or in the model’s operation, fuzzy explanations make it explicit, enabling the user to form a more adequate mental model of the situation.
– *Provide gradations:* they show the degree of influence of a particular factor, allowing the user to build a more nuanced and detailed mental model of causal relationships.

The same chapter also considers approaches to evaluating the quality of explanations. It is substantiated that explainability analysis should rely on a combination of quantitative and qualitative metrics: the former characterizes the consistency, distinguishability, and stability of explanations, while the latter characterize their comprehensibility, depth, learnability, and convenience for the user.

In the **fourth chapter** was developed a formal model of fuzzy explainability and proposes the architecture of a system implementing this approach. Fuzzy explainability is defined as the ability of an XAI system to generate explanations that connect the model’s decision with semantic concepts of the domain while simultaneously reflecting the degree of confidence in these connections.

The corresponding architecture is presented in Figure 3. Its input components are observations, mental models, and “explanatory” properties; its central element is a fuzzy explanation module; and its output is an interpretable decision supplemented with new fuzzy knowledge.

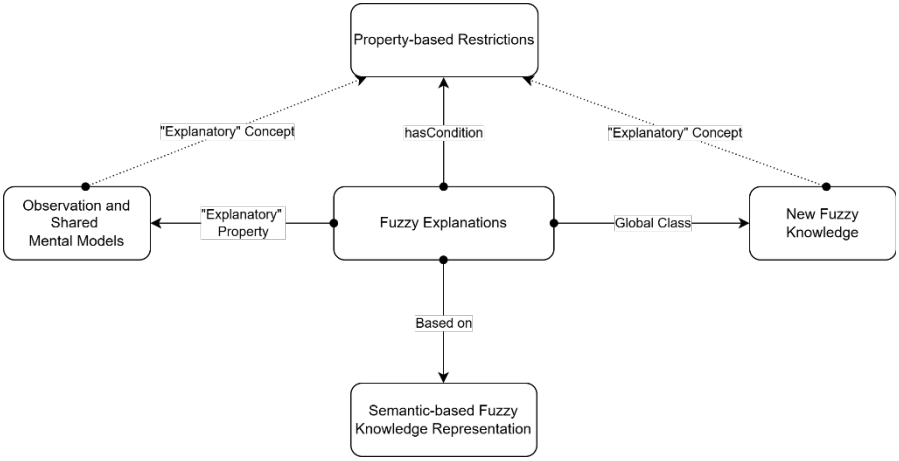

Figure 3. System design defining the proposed fuzzy explanations based on an ontology

For example, credit risk may be based on various explainable properties obtained from obser*vations and reflecting a distributed* MM. In our semantically oriented *fuzzy knowledge represent*ation, a credit risk ontology can define concepts such as ShortDuration, SmallCredit, LittleSaving, etc., and relate them to the concept of credit risk. Property-based restrictions (for example, has*Duration*, *hasCreditAmount*, *hasS*avingAccount, etc.) refine the explanation by means of fuzzy conditions. New fuzzy know*ledge is extra*cted from the dataset, defining t*he global class. Th*e system *output forms a* fuzzy explanation in w*hich the borrower* has GoodRisk (0.94), supported by the properties ShortDuration (0.67), SmallCredit (0.55), and LittleSaving (0.54).

The need to fuzzify “explanatory” properties is caused by the fuzziness of the initial data, the uncertainty of expert knowledge, and the need for a graded representation of the factors influencing the model’s output. Unlike the binary approach, such a formalization makes it possible to describe situations in which an object has a class or property only to a certain degree.

The goal of fuzzy explainability is to generate explanations that not only connect model predictions with semantic concepts, but

also reflect the degree of confidence in these links and the uncertainty inherent in them. This is especially important in environments with fuzzy knowledge, where binary answers are inadequate.

Formally, ***the fuzzy explainability framewo***rk was defined as, $FE = \left(M, O_f, P_{\text{exp}}, K_f\right)$ where M represents the common distributed MM, O_f denotes the fuzzy ontology, P_{exp} is the set of explanatory properties, and K_f corresponds to fuzzy knowledge extracted from the dataset. MMs contribute to our representation of fuzzy explainability by corresponding to the cognitive representations of user and expert knowledge, reflecting the real uncertainty of knowledge, and providing gradations and factor influences.

A fuzzy explanation for a decision or event y is expressed as $E_f(y) = \{\left(x_i, \mu(x_i)\right) \mid x_i \in X, \mu(x_i) \in [0,1]\}$, where $X = x_1, x_2, \dots, x_n$ is the set of explanatory factors and $\mu(x_i)$ represents the degree of contribution of factor x_i to explanation y.

For the practical implementation of the proposed approach, an architecture was formed based on the use of Fuzzy OWL2, fuzzy "explanatory" properties, fuzzy axioms for explanations, mechanisms for inferring degrees of confidence, and procedures for presenting explanations to the user[11].

As a fuzzy logical inference mechanism, fuzzyDL is considered, supporting fuzzy description logics and the processing of ontologies in the Fuzzy OWL2 format. The fuzzyDL algorithm is based on a combination of a tableau algorithm and the solution of an optimization problem. Table 2 shows some rules of fuzzy ALC with an empty ABox[12].

---

[11] Косов, П.И. Объяснимый Искусственный Интеллект: Исследование Нового Метода для Улучшения Объяснений на Основе Онтологий // Ümummilli Lider Heydər Əliyevin anadan olmasının 101-ci ildönümünə həsr olunmuş Doktorantların və Gənc Tədqiqatçıların Respublika Elmi Konfransının Materialları, – Bakı: – 6-7 may, – 2024, – s. 1-5.

[12] Bobillo, F., Straccia, U. Generalizing type-2 fuzzy ontologies and type-2 fuzzy description logics // International Journal of Approximate Reasoning, – 2017. vol. 87, – p. 40-66.

Table 2
Logical inference of fuzzy ALC with an empty ABox

Rule	Precondition	Action
$(\bot)$	$\bot \in \mathcal{L}(v)$	$C = C \cup \{x_{v:\bot} = 0\}$
$(\top)$	$\top \in \mathcal{L}(v)$	$C = C \cup \{x_{v:\top} = 1\}$
$(\neg)$	$\neg A \in \mathcal{L}(v)$	$C = C \cup \{x_{v:\neg C} = \ominus x_{v:C}\}$
$(\sqcap)$	$C_1 \sqcap C_2 \in \mathcal{L}(v)$	$\mathcal{L}(v) = \mathcal{L}(v) \cup \{C_1, C_2\}$ $C = C \cup \{x_{v:C} \otimes x_{v:D} = x_{v:C_1 \sqcap C_2}\}$
$(\sqcup)$	$C_1 \sqcup C_2 \in \mathcal{L}(v)$	$\mathcal{L}(v) = \mathcal{L}(v) \cup \{C_1, C_2\}$ $C = C \cup \{x_{v:C} \oplus x_{v:D} = x_{v:C_1 \sqcup C_2}\}$
$(\exists)$	$\exists R.C \in \mathcal{L}(v)$	create a new node w $\mathcal{L}(\langle v, w \rangle) = \mathcal{L}(\langle v, w \rangle) \cup \{R\}$, and $\mathcal{L}(w) = \mathcal{L}(w) \cup \{C\}$, and $C = C \cup \{x_{\langle v,w \rangle:R} \otimes x_{w:C} = z, z \geq x_{v:\exists R.C}\}$
$(\forall)$	$\forall R.C \in \mathcal{L}(v)$ $R \in \mathcal{L}(\langle v, w \rangle)$	$\mathcal{L}(w) = \mathcal{L}(w) \cup \{C\}$ $C = C \cup \{x_{v:\forall R.C} \geq z, z = x_{(v,w):R} \Rightarrow x_{w:C}\}$

A key architectural principle of the software implementation is the "model-per-property" paradigm. Instead of a single monolithic model, an ensemble is used that includes a global classifier and a set of separate classifiers, each responsible for assessing one of the "explanatory" properties. This makes it possible to directly connect the results of the models with the ontological representation of knowledge. Figure 4 presents the proposed new architecture. Such a system design not only increases modularity but also makes the decision-making process more transparent. The results of any individual model constitute a separate building block for the final explanation.

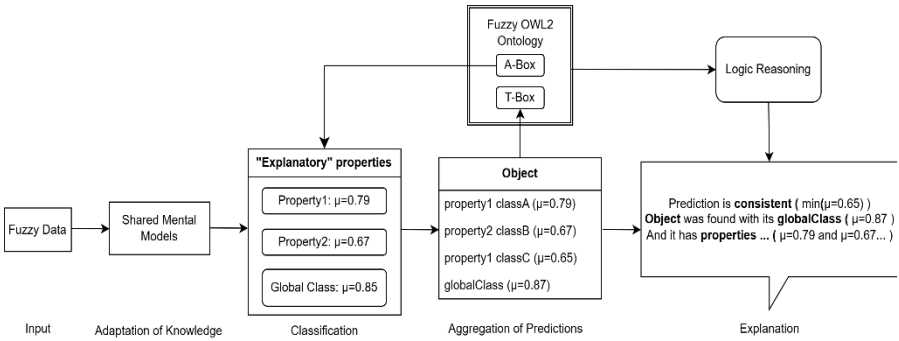

Figure 4. Proposed architecture for a new XAI system

The proposed operating scheme of the new system. The software complex operates according to a clearly defined principle that can be broken down into the following steps:

Step 1. Data preprocessing: the system receives data about an object as input (for example, a row from a table or an image). The data go through a preparation stage: for tabular data this may include encoding categorical features, normalization, or discretization of numerical values to bring them into the interval formats described in the ontology (for example, “Low,” “Medium,” and “High” age). Integration of distributed mental models also takes place.

Step 2. Parallel classification: the prepared feature vector is simultaneously fed to the global classifier and to all N property classifiers. Each of them returns its own prediction (the target class and the values for each of the “explanatory” properties).

*Step 3. Ontology population: t*he obtained predictions are used to programmatically create an instance in the ontology’s ABox. For example, for a new patient, an individual of the Patient class is created, and through object properties it is assigned links to the corresponding individuals of the property classes. The process proceeds in conjunction with distributed mental models.

*Step 4. Representation of fuzzy kn*owledge: at this stage, fuzziness is integrated. Degrees of membership obtained from ML models or computed by a fuzzy clustering algorithm are used to annotate the created assertions in the ontology.

*Step 5. Logical inference and consistency checking: a*fter the ontology is populated with data about a specific individual, logical inference is launched. It performs a key function, namely, checking whether the new assertions contradict the axioms and constraints already existing in the ontology.

*Step 6. Explanation generat*ion: at the final stage, the system aggregates the results. The final class, all "explanatory" pro*perties, and thei*r *degree*s of confidence by means of fuzzy values are extracted from the ontology. This information is formatted into a structured form that can then be passed to any user interface for visualization.

Thus, the developed software complex is a flexible and scalable platform. It effectively solves the problem of building explainable systems capable of working with incomplete and fuzzy knowledge. Such a platform makes it possible to explain classification results obtained using a "black-box" model, as well as clustering tasks. This can be regarded as an important step toward creating more reliable and trustworthy XAI systems.

In the **fifth chapter** was presented the results of the experimental approbation, verification, and evaluation of the proposed fuzzy-semantic approach.

The methodology was verified on 6 datasets related to different domains: Fashion MNIST (computer vision), Glass Identification (materials science), German Credit Risk (credit analytics), Heart Failure Clinical Records (prediction in medicine), Network Traffic Data for Malicious Activity Detection (cybersecurity in computer networks), and Student Performance (student achievement). This choice made it possible to assess the applicability of the proposed approach to heterogeneous data types and different interpretation scenarios.

The experimental approbation was carried out in three stages. At the first stage, the basic concept of "explanatory" properties was verified in the environment of a classical OWL2 ontology. At the second stage, a hybrid scheme was investigated in which a crisp ontology was supplemented with a quantitative representation of uncertainty. At the third stage, a full transition to Fuzzy OWL2 was carried out, which made it possible to directly interpret model outputs

as degrees of membership for fuzzy axioms. Comparison of the results of the three stages showed that the fully fuzzy variant provides the greatest expressiveness, flexibility, and meaningfulness of the generated explanations.

At ***the first verification stage***, the ***basic concept*** of "explanatory" properties was tested. For this purpose, an OWL2 ontology and a standard logical inference mechanism were used. Figure 5 shows an example of the structure of the corresponding ontology obtained during the experiment[9].

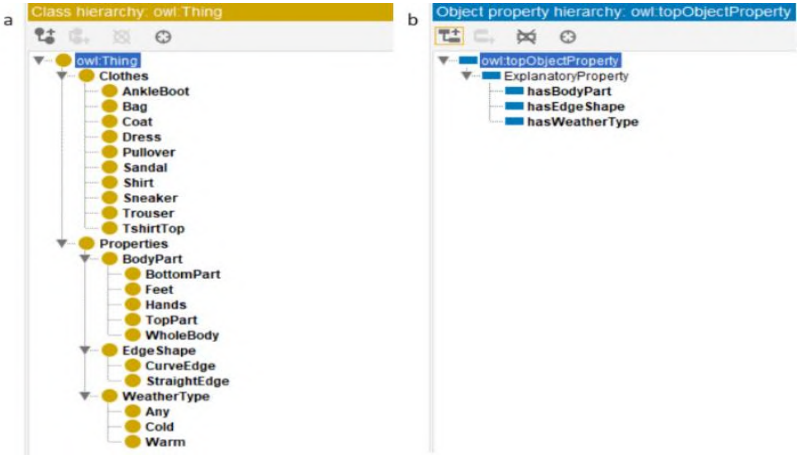

Figure 5. (a) Class hierarchies and (b) object property hierarchies for the Fashion MNIST data

The "black-box" model performed classification of the dataset. As a result, the fundamental possibility of representing "explanatory" properties by means of a classical OWL2 ontology and standard logical inference for explaining the operation of the classifier was confirmed.

However, the explanation was overloaded with a large amount of information, and property values were taken into account both from the side of class membership and from the side of non-membership. This system did not have fuzzy values, which led to such overload. An example of the explanation is given below:

*"**Image_0** belongs to the class TshirtTop and is **consistent** in the ontology. Does **not** have WholeBody in BodyPart, has TopPart in BodyPart, does **not** have BottomPart in BodyPart, does no**t** have Feet in BodyPart, does **not** have Hands in BodyPart, does **not** have Cold in WeatherType, has Warm in WeatherType, **does not** have Any in WeatherType**,** does **not** have StraightEdge in EdgeShape, has CurveEdge in EdgeShape."*

Another confirmation of the flexibility of the properties is the exp[13]eriment on the "Glass Identification" dataset, where it was necessary to classify the type of glass based on its chemical composition. An ontology was created in which the "explanatory" properties were the chemical elements themselves.

At ***the second s***tage, a hybrid va***riant of the approach was*** implemented, in which a crisp OWL2 ontology was supplemented with a quantitative representation of uncertainty. To store the degree of membership, a special data property was used, acting as a container for fuzzy characteristics within a formally crisp structure. An example of this type of ontology is shown in Figure 6.

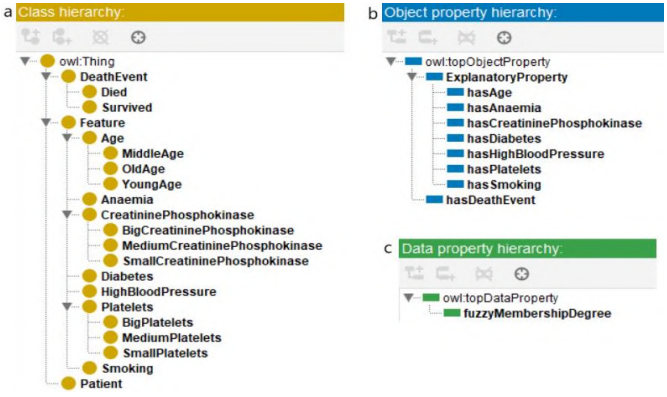

Figure 6. Class hierarchies (a), object property hierarchies (b), and data property hierarchies (c) for the heart failure dataset

[13] Qardaşova, L.A., İbrahimova, S.R., Kosov, P.İ. Şüşənin Kimyəvi Tərkibinə Əsaslanan İdentifikasiyasında İzah Oluna Bilən Süni İntellektin Tətbiqi // – Bakı: Elmi Məcmuə (Milli Aviasiya Akademiyası), – 2025. c. 26, № 4, – s. 91-99.

To support logical inference, Semantic Web Rule Language (SWRL) rules are also used. Full explainability is achieved because the rules include all existing "explanatory" properties in the ontology (Figure 7).

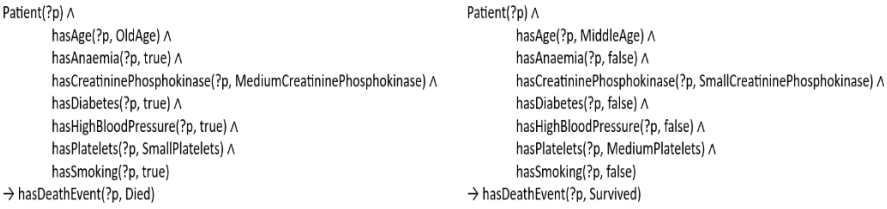

Figure 7. Possible SWRL rules for logical inference in the ontology for the heart failure dataset[7]

It has been proven that introducing a quantitative characteristic of uncertainty into a crisp ontological structure makes it possible to increase the flexibility of representing explanatory factors, although the model itself still does not provide full fuzzy semantics. An example explanation is given below:

*"**Entry_0** for the Patient class belongs to the Survived class**.** It is **consistent** in the ontology. The explanatory properties are as follows: MiddleAge for hasAge, false for hasAnaemia, SmallCreatininePhosphokinase for hasCreatininePhosphokinase, false for hasDiabetes, false for hasHighBloodPressure, MediumPlatelets for hasPlatelets, false for hasSmoking. The target class is survived for hasDeathEvent and has a fuzzy membership degree of 0.8."*

Thanks to the fuzzy membership function, it was possible to obtain less information in the explanation while preserving the main essence. However, the absence of fuzzy membership for the "explanatory" properties leaves the problem of representing imprecise knowledge unresolved and makes the explanations insufficiently clear.

The final stage corresponded to a full transition to ***fuzzy explainability*** based on Fuzzy OWL2. The Student Performance

dataset, containing features that naturally admit linguistic and fuzzy interpretation, was used as the experimental basis. In this configuration, quantitative attributes were represented through linguistic variables and fuzzy terms, while model outputs were directly interpreted as degrees of membership for fuzzy axioms[10].

The software configuration of the system underwent key changes:

- *The fuzzy ontology (Fuzzy OWL2) ma*de it possible to introduce the following fundamental fuzzy constructions:
  - *Fuzzy axioms: ea*ch assertion in the ABox now has a certain degree of confidence that directly corresponds to the output value of the ML model;
  - *Linguistic variables: qua*ntitative attributes (for example, "study time") were represented as linguistic variables with fuzzy terms ("insufficient," "moderate," "significant"). The process of defining these terms and their membership functions (for example, trapezoidal or triangular) was based on expert knowledge in pedagogy, which made it possible to embed human understanding of the domain into the model.
- *Architecture***:** the "model-per-property" principle was preserved, but now the output results of the models (for example, probability 0.85 for the hasStudyTime property) were directly and organically interpreted as degrees of membership for fuzzy axioms, creating integration between ML inference and the semantic representation of knowledge.

It was established that the use of Fuzzy OWL2 provides the most natural and complete representation of degrees of membership, thanks to which explanations become more informative, graded, and interpretable.

An example of a fuzzy axiom for an "explanatory" property is shown in Figure 8.

```
<owl:Axiom>
    <fuzzyLabel>
        <fuzzyOwl2 fuzzyType="axiom">
            <Degree value="0.79" />
        </fuzzyOwl2>
    </fuzzyLabel>
    <owl:annotatedSource rdf:resource="&fuzzy_student;Student_0"/>
    <owl:annotatedTarget rdf:resource="&fuzzy_student;HighGrades"/>
    <owl:annotatedProperty rdf:resource="&fuzzy_student;hasGrades"/>
</owl:Axiom>
```

Figure 8. Example of an “explanatory” property with a Fuzzy OWL2 fuzzy axiom using student data

The results showed that in the fully fuzzy variant the system forms more informative and flexible explanations in which each “explanatory” property is accompanied by a measure of confidence. This makes it possible not only to indicate which factors influenced the decision, but also to reflect the degree of their participation in forming the final conclusion12. An example of the system output is:

“*The student (Student_0) was classified as GoodAcademicPerformance (with confidence 0.88). This decision is based on the following factors: HighStudyTime (0.85), HighGrades (0.79), GoodFamRel (0.67), LittleAbsences (0.62), and MediumGoOut (0.58).*”

This example demonstrates the key advantages of our approach, since the system not only identifies the relevant characteristics that influenced the positive decision, but also quantitatively evaluates the contribution of each factor.

The chapter noted that although the classification accuracy in our system did not surpass existing solutions, the main goal of the study was precisely to improve the interpretability of predictions. In this respect, fuzzy ontological structures showed a significant advantage over classical approaches.

Comparative analysis of the developed fuzzy and crisp approaches. To clearly demonstrate the difference in explanations, a comparative analysis of the two approaches was carried out on the German Credit Risk dataset based on “explanatory” properties.

T*he crisp appro*ach uses a standard OWL2 ontology. For each “explanatory” property (concept), a separate ML model was trained.

The model outputs (for example, the predicted class LittleSaving) were transformed into absolute, binary assertions that were added to the ABox (Figure 9).

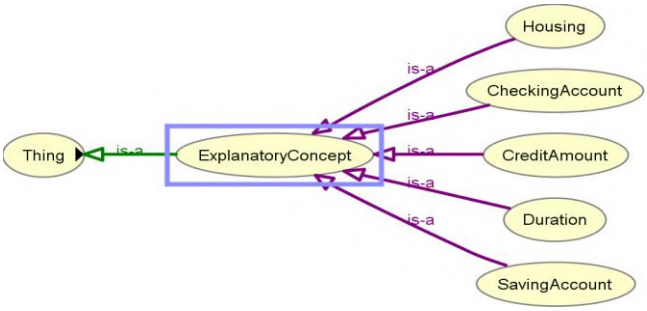

Figure 9. “Explanatory” concept for the credit risk dataset

T*he fuzzy approa*ch employed a Fuzzy OWL2 fuzzy ontology. This made it possible to represent concepts such as SmallCredit or LongDuration not as discrete categories, but as linguistic variables with continuous degrees of membership. The process of forming an explanation consisted in aggregating these fuzzy assertions and their degrees of confidence.

To support logical inference, 15 SWRL rules were developed (9 for GoodRisk and 6 for BadRisk), linking combinations of properties with the final decision. For example: “Loan*Taker(?x)* ^ *hasCheckingAccount(?x, NAChecking)* ^ *hasSavingAccount(?x, LittleSaving) -> hasRisk(?x, Bad)*”.

Next, an evaluation and ***comparative analysis of the developed method with other approaches was carried out***. The main goal was to verify the effectiveness and reliability of the proposed explanation methods.

To ensure comparability of the results, the same initial conditions were fixed: the same sample element, the same model prediction, and the same property space. This eliminated the influence of changing the observed object, the set of factors, and the final forecast, and shifted the comparison to the differences between the explanation-building mechanisms themselves.

The quantitative criteria used were Euclidean distance, the Jaccard coefficient, a weighted fuzzy average, Bayesian probability of explanation consistency, and also a weighted-sum model based on Z-numbers. LIME, SHAP, counterfactual explanations, explanations based on fuzzy rules, and explanations based on fuzzy decision trees were compared with one another.

Nevertheless, for a reliable assessment, a group of experts and users was also surveyed in order to understand whether all information was reflected in the obtained explanations. An assessment of comprehensibility, satisfaction, trust, effectiveness, and usefulness was carried out.

The obtained results confirm the expediency of moving from a crisp semantic model to a fuzzy-ontological architecture as a direction for the development of XAI systems. The proposed approach provides deeper, context-rich, and cognitively compatible explanations, and the verification proves that the integration of ontologies and fuzzy logic creates a formal basis for increasing trust in machine learning results and the ability to work with fuzzy knowledge. A working version of the research code for the proposed system and instructions for use were posted on the GitHub online resource[14].

In the **conclusion** was presented the scientific and practical findings obtained during the course of work.

MAIN RESULTS OF THE WORK

The main results fully correspond to the stated objectives and are as follows:

1. An analysis of modern approaches to building explainable AI systems was carried out, as a result of which the limitations of existing algorithmic and ontological explanation methods were identified, associated respectively with insufficient semantic

[14] Kosov, P. Fuzzy Ontology Explanatory Properties Research Code: [Electronic resource] / GitHub. – 2025. URL: https://github.com/pavelkosov99/Fuzzy-Ontology-Explanatory-Properties-Research-Code

depth and the inability to adequately account for the uncertainty of real-world data.

2. The concept of explanatory properties was developed and formalized as a means of ontological representation of expert knowledge, ensuring the interpretation of data mining results with due regard to logical dependencies, the functional purpose of objects, and the characteristics of user perception.
3. The expediency of using the concept of the user's mental models in building explanations was substantiated, and it was shown that aligning an explanation with the user's cognitive representations increases its comprehensibility and practical value for different categories of users.
4. A methodology of fuzzy explainability based on the integration of semantic technologies and fuzzy logic methods for representing and processing imprecise knowledge in explainable artificial intelligence tasks was proposed and theoretically substantiated.
5. A method was developed for representing explanatory properties and ontological axioms using Fuzzy OWL2 fuzzy ontologies, ensuring the formation of explanations with a quantitative assessment of the degree of confidence.
6. A procedure was developed for processing fuzzy clustering results, making it possible to automatically form assertions about object properties in a fuzzy ontology on the basis of degrees of membership obtained from data.
7. A mechanism was developed for generating fuzzy semantic explanations based on the apparatus of fuzzy description logics, ensuring the obtaining of logically justified and interpretable explanations of the results of "black-box" models without modifying the models themselves.
8. The proposed methodology was experimentally tested on six datasets from different subject domains. The obtained results confirmed the universality, scalability, and practical effectiveness of the developed approach.

The proposed methodology lays the scientific and practical foundation for a new generation of XAI systems. It was proven that a

system integrating a semantic network and fuzzy logic is capable of adequately reflecting the real uncertainty of real-world data. It also makes it possible to create not only more transparent, but also more reliable and trustworthy artificial intelligence systems.

The main results of the dissertation were published in the following scientific works:

1. Gardashova, L., Kosov, P. A Review of the Solutions for Autonomous Exposure of Intrusions and Malicious Activities in Automated Networks in the Environment of Big Datasets // – Bakı: Azərbaycan Ali Texniki Məktəblərinin Xəbərləri, – 2023. c. 30, № 07, – s. 338-350.
2. Косов, П.И. Объяснимый искусственный интеллект: исследование нового метода для улучшения объяснений на основе онтологий // Ümummilli Lider Heydər Əliyevin anadan olmasının 101-ci ildönümünə həsr olunmuş Doktorantların və Gənc Tədqiqatçıların Respublika Elmi Konfransının Materialları, – Bakı: – 6-7 may, – 2024, – s. 1-5.
3. Kosov, P. Advancing XAI: new properties to broaden semantic-based explanations of black-box learning models / P. Kosov, N. El Kadhi, C. Zanni-Merk, L. Gardashova // Procedia Computer Science, – 2024. vol. 246, – p. 2292-2301.
4. Косов, П.И. Краткий разбор логики объяснимого искусственного интеллекта и её нечёткости с применением онтологий // Труды X Международной Научной Конференции «Информационные технологии интеллектуальной поддержки принятия решений», – Уфа: – 12-14 ноября, – 2024, – с. 23-28.
5. Qardaşova, L.A., İbrahimova, S.R., Kosov, P.İ. Şüşənin kimyəvi tərkibinə əsaslanan identifikasiyasında izah oluna bilən süni intellektin tətbiqi // – Bakı: Elmi Məcmuə (Milli Aviasiya Akademiyası), – 2025. c. 26, № 4, – s. 91-99.
6. Kosov, P., El Kadhi, N., Zanni-Merk, C., Gardashova, L. Semantic-Based XAI: Leveraging Ontology Properties to Enhance Explainability // Proceedings of the 2024

International Conference on Decision Aid Sciences and Applications, – Manama: – 11-12 December, – 2024, – p. 1-5.
7. Qardaşova, L.A., Kosov, P.İ. İzah oluna bilən süni intellekt və onun sənayedə tətbiqlərinin icmalı // – Bakı: Azərbaycan Ali Texniki Məktəblərinin Xəbərləri, – 2025. c. 49, № 02, – s. 601-621.
8. Косов, П.И. Анализ объяснимости вторжений и вредоносной деятельности в компьютерных сетях посредством нечёткой онтологии // Ümummilli Lider Heydər Əliyevin anadan olmasının 102-ci ildönümünə həsr olunmuş Doktorantların və Gənc Tədqiqatçıların Respublika Elmi Konfransının Materialları, – Bakı: – 7-8 may, – 2025, – s. 1-4.
9. Косов, П.И., Гардашова, Л.А. Повышение достоверности объяснимого искусственного интеллекта посредством нечеткой логики и онтологии // – Воронеж: Моделирование, Оптимизация и Информационные Технологии, – 2025. т. 13, № 2(49), – с. 1-11.
10. Косов, П.И. Разработка нечёткой онтологии для объяснимого искусственного интеллекта для принятия решений в нечёткой среде // – Tashkent: Chemical Technology, Control and Management, – 2025. vol. 2025, № 2, – p. 19-26.

The author's personal contribution to works published in co-authorship:

[1], [3], [5], [6], [7], [9] – theoretical research, software implementation, verification, and preparation of the studies in article format.

The defense will be held on 05 May 2026 at 14:00 at the meeting of the Dissertation council FD 2.48 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at Azerbaijan State Oil and Industry University.

Address: AZ 1010, Baku, Azadlig avenue, 34.

Dissertation is accessible at the Azerbaijan State Oil and Industry University Library.

Electronic version of the abstract is available on the official website of the Azerbaijan State Oil and Industry University.

Abstract was sent to the required addresses on 02 Aprel 2026.

Signed for print: 12.03.2026

Paper format: A5

Volume: 37 000

Number of hard copies: 20