АЗЕРБАЙДЖАНСКАЯ РЕСПУБЛИКА

На правах рукописи

РАЗРАБОТКА НЕЧЁТКОГО СЕМАНТИЧЕСКОГО ПОДХОДА ДЛЯ ОБЪЯСНИМОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Специальность: 3338.01 – “Системный анализ, управление и обработка информации” (обработка данных)

Отрасль науки: Технические науки

Соискатель: **Косов Павел Игоревич**

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
доктора философии

Баку – 2026

Диссертационная работа выполнена на кафедре «Компьютерной инженерии» Азербайджанского Государственного Университета Нефти и Промышленности.

Научные руководители: Доктор технических наук, профессор
Гардашова Латафат Аббас гызы

Профессор по компьютерным наукам
Сесилия Зинни-Мерк

Официальные оппоненты: Доктор технических наук, профессор
Алиев Алекпер Али Ага оглы

Доктор технических наук, профессор
Рзаев Рамин Рза оглы

Доктор технических наук, профессор
Маммедов Джаваншир Фирудин оглы

Диссертационный совет FD 2.48 Высшей Аттестационной Комиссии при Президенте Азербайджанской Республики, действующий на базе Азербайджанского Государственного Университета Нефти и Промышленности

Председатель диссертационного совета: Член-корреспондент НАНА,
Доктор технических наук, профессор

Рафик Азиз оглы Алиев

Учёный секретарь диссертационного совета: Доктор технических наук, доцент

Акиф Вали оглы Ализаде

Председатель научного семинара: Доктор технических наук, профессор

Камаля Рафик гызы Алиева

İMZANI TƏSDİQ EDİRƏM
ADNSU-nun Elmi katibi

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. На сегодняшний день, во многих отраслях человеческой деятельности наблюдается рост сложности и производительности моделей искусственного интеллекта (ИИ), изучением которых занимается область машинного обучения и нейронные сети. Миллиарды параметров демонстрируют эффективность систем, что привело к их широкому внедрению в индустрии – от здравоохранения и юриспруденции до промышленности. Одновременно с этим, возрастает потребность в создании систем, соответствующих принципам ответственного ИИ, основной задачей которого является разработка и исследование надёжных, справедливых и подотчётных ИИ-систем.

Применение высокопроизводительных, но непрозрачных моделей ИИ, функционирующих как «чёрные ящики», в рамках концепции ответственного ИИ порождает многие проблемы. Основным результатом работы таких моделей является прогноз без возможности объяснять. Тогда как необходимым условием для создания ответственной системы является способность к объяснению и обоснованию полученных решений. Применение мощных моделей ИИ само по себе не является достаточным условием для создания доверенных систем; что влечёт потребность в проведении исследований по вопросам разработки методов объяснимого ИИ (XAI), способных преодолеть проблему «чёрного ящика».

Существующие методы XAI, в свою очередь, сталкиваются с двумя фундаментальными ограничениями: алгоритмические подходы лишены семантической глубины, предоставляя объяснения на уровне признаков без их смысловой интерпретации, а классические онтологические подходы не способны адекватно моделировать неопределённость реального мира. Это обуславливает актуальность разработки новых методов, объединяющих семантическую силу онтологий и математических средств нечёткой логики для создания нового поколения систем XAI.

Объект и предмет исследования. Объектом исследования выступают процессы, модели и методы генерации объяснений для решений, принимаемых системами ИИ, а именно модели «чёрного ящика». Основой исследования являются вопросы разработки и применения методологии построения XAI на основе интеграции семантических технологий и нечёткой логики. В предмет исследования входят семантические свойства, представленные в онтологии, методы формализации нечётких знаний на базе Fuzzy OWL2, а также алгоритмы и архитектура для генерации объяснений со степенями уверенности и их оценки.

Цель и задачи работы. Целью работы является комплексный анализ существующих подходов XAI, а также разработка и исследование новых подходов к построению XAI, основанной на интеграции семантических технологий и нечёткой логики, с целью обеспечения её практической применимости для повышения достоверности, глубины и понятности объяснений моделей «чёрного ящика». Достижение указанной цели предполагает выполнение следующих основных задач:

- Анализ существующих систем XAI для выявления ключевых ограничений, устранение которых обуславливает разработку новых методов;
- Разработка концепции семантических свойств в онтологии для универсального представления экспертных знаний;
- Разработка метода “нечёткой объяснимости” для моделирования неточных знаний и генерации объяснений со степенями уверенности;
- Создание гибкой архитектуры объяснимой системы для интеграции моделей машинного обучения с нечёткой онтологией;
- Разработка процедуры обработки результатов нечёткой кластеризации для представления свойств данных в онтологии;
- Проведение серии вычислительных экспериментов для апробации и оценки предложенных подходов на разнородных наборах данных.

Методы исследования. Поставленные задачи решались посредством применения теории нечётких множеств, методов инженерии знаний, моделирования онтологий и семантических сетей, машинного и глубокого обучения, кластерного анализа, методов компьютерного моделирования, математических вычислительных экспериментов, а также методов экспертной оценки и сравнительного анализа.

Основные положения, выносимые на защиту. На защиту выносятся следующие основные положения и результаты диссертационной работы:

- Концепция «объяснительных» свойств, представляющая собой новый способ онтологической формализации экспертных знаний, который обеспечивает применимость для различных типов данных;
- Методология «нечёткой объяснимости», позволяющая моделировать неопределённость и расплывчатость реальных данных и знаний с помощью нечётких онтологий Fuzzy OWL2 и генерировать объяснения с количественными степенями уверенности;
- Гибкий метод и архитектура объяснимой системы, позволяющие применять семантико-нечёткое объяснение к любым существующим моделям “чёрного ящика” без компромисса между их предиктивной точностью и интерпретируемостью;
- Комплексная методология оценки качества генерируемых нечётких семантических объяснений, включающая функционально-ориентированные, человеко-ориентированные и гибридные подходы.

Научная новизна. Основная суть научной новизны данной диссертационной работы заключается в следующем:

- Впервые созданы абстрактные семантические «объяснительные» свойства для онтологического представления экспертных знаний в объяснимом искусственном интеллекте;
- Впервые разработаны объяснения в среде неточных знаний с использованием новых созданных

«объяснительных» свойств для нечёткой онтологии OWL2;
- Впервые разработан метод для объяснения результатов «чёрного ящика» на основе разных типов данных, представленных в виде нечётких знаний;
- Обработаны результаты нечёткой кластеризации свойств данных для представления в онтологии;
- Разработка компьютерной симуляции, анализ и оценка полученных оригинальных результатов на основе предложенных подходов.

Научно-практическая значимость. Результаты, полученные в диссертационной работе, обладают как теоретической, так и практической значимостью. Использование в работе нечёткого подхода даёт возможность учитывать неопределённость. В диссертационной работе представлена комплексная методология по созданию систем объяснимого искусственного интеллекта, позволяющая осуществлять синтез модулей объяснения для высокопроизводительных моделей типа «чёрный ящик». Преимущество данной методики заключается в её гибкости и универсальности; в способности генерировать семантически богатые и интуитивно понятные объяснения; в возможности адекватно моделировать неопределённость; а также в повышении доверия к системам искусственного интеллекта и повышении качества совместного принятия решений человеком и машиной. Представленная методика была апробирована на наборах данных, относящихся к различным предметным областям (финансы, медицина, материаловедение, кибербезопасность, компьютерное зрение и социальная аналитика). Для большинства компонентов методики был разработан программный прототип, демонстрирующий возможность её практической реализации.

Апробация работы. Основные положения работы докладывались на следующих конференциях и семинарах:
- 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.

Публикации. По теме диссертации опубликовано 10 научных работ (4 без соавторов), где из них 6 статей и 4 тезиса. 5 работ индексируются в международных базах. 1 статья входит в базу данных Scopus (категория Q2), 1 тезис – в материалах международной конференции индексируемой в Web of Science и Scopus. В том числе другие международные публикации включают: 1 статью в журнале рекомендуемом ВАК Российской Федерации (категория К2), 1 статью в журнале рекомендуемом ВАК Республики Узбекистан и 1 тезис в материалах международной конференции, индексируемой в базе РИНЦ.

Название организации, в которой выполнена диссертационная работа. Диссертационная работа выполнена на кафедре «Компьютерной инженерии» Азербайджанского Государственного Университета Нефти и Промышленности.

Структура и объем работы. Основная часть работы представлена на 160 страницах и состоит из введения, 5-и глав, заключения, списка использованной литературы и списка сокращений. Также содержит 27 рисунков, 10 таблиц и 150 элементов в списке литературы. Без учёта изображений, таблиц,

графиков, приложений, библиографии и пробелов в тексте объём работы приблизительно является следующим: Введение – 8 000 знаков, Глава I – 25 000 знаков, Глава II – 45 000 знаков, Глава III – 40 000 знаков, Глава IV – 40 000 знаков, Глава V – 40 000 знаков и Заключение – 2 500 знаков. Всего общий объём диссертация примерно составляет 200 000 знаков.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы исследования, сформулированы объект, предмет, цель и задачи диссертационной работы, определены методы исследования, положения, выносимые на защиту, научная новизна и научно-практическая значимость полученных результатов. Кроме того, приведены сведения об апробации работы, публикациях по теме исследования, а также о структуре и объёме диссертации.

В **первой главе** выполнен анализ современного состояния методологий Объяснимого Искусственного Интеллекта (англ. eXplainable Artificial Intelligence, XAI), обоснована необходимость и актуальность исследования. Также поставлена научная задача диссертационной работы.

В главе представлено, что ***объяснение*** в рамках XAI следует рассматривать как предоставление пользователю информации о решении модели вместе с теми факторами и зависимостями, которые делают это решение понятным и проверяемым.

Основываясь на вышесказанное, были выявлены ключевые ограничения существующих подходов: недостаточный учёт доменных знаний, слабая семантическая интерпретация признаков и недостаточная адаптация объяснений к когнитивным особенностям пользователя.

По результатам проведённого анализа в главе, ***основная научная задача диссертационной работы*** является – разработка и исследование новой методологии построения систем XAI, основанной на семантической технологии и нечёткой логики, с целью обеспечить повышение семантической насыщенности,

достоверности и интерпретируемости объяснений, формируемых для моделей машинного обучения (МО) “чёрного ящика”, а также возможность работать с нечёткими и неполными знаниями.

Во **второй главе** обосновано использование семантических технологий и нечёткой логики в качестве формальной основы предлагаемого подхода. Описаны ключевые понятия и подходы, и также проанализированы различные аспекты нечёткой логики, нечёткой системы логического вывода и нечёткой онтологии Fuzzy OWL2.

Семантическая паутина рассматривается как совокупность стандартов Консорциума Всемирной паутины (англ. World Wide Web Consortium, W3C), обеспечивающих идентификацию сущностей, представление знаний и выполнение запросов к ним. В качестве базовых компонентов используются URI для уникальной идентификации ресурсов, RDF для представления знаний в триплетной форме «субъект-предикат-объект», SPARQL для запросов к RDF-данным и онтологии как средство задания словаря понятий и их семантических связей.

Онтология в информатике используется как средство структурированного и формализованного представления знаний. Она имеет иерархическую организацию (показанной рисунке 1) и может быть рассмотрена на уровнях верхней, базовой и предметной онтологии[1].

Рисунок 1. Три основных уровня онтологии в информатике[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.

Формально онтология задаётся кортежем, включающим множество концептов, множество свойств и систему аксиом, фиксирующих отношения и ограничения предметной области[2] при помощи формулы $O = (C, \leq_C, R, \leq_R, A^O)$, где онтология O включает частичное упорядоченные $\leq_C$ концепты C, свойства R с частичным порядком $\leq_R$, и аксиомы A^O.

Описана ***релевантность использования онтологий в XAI*** и установлено, что онтология позволяет связать результаты машинного обучения с формально определёнными концептами и отношениями предметной области, интегрировать символические знания с данными и формировать объяснения, понятные человеку.

Математической основой современных онтологий служат ***дескриптивные логики (англ. Descriptive Logic, DL)***, обеспечивающие строгое и при этом интерпретируемое описание предметной области. Онтология, построенная на базе дескриптивных логик, включает *TBox* – содержит аксиомы о концептах и их взаимосвязях (например, включение $C1 \sqsubseteq C2$, где $C1$ – подконцепт $C2$; эквивалентность $C1 \equiv C2$); *ABox* – содержит утверждения об индивидах, их принадлежности к концептам (например, $a: C$) и связях между ними через роли; *RBox* – содержит аксиомы о ролях и их характеристиках (например, иерархия $R1 \sqsubseteq R2$, транзитивность).

Например, в DL, язык ALUEC – это базовый язык AL, расширенный операциями объединения (U), полной экзистенциальной квантификации (E) и отрицания произвольных концептов (C). Названия более сложных логик формируются по этому же принципу. Например, SHOIN – это логика S (ALC + транзитивность), расширенная иерархией ролей (H), номиналами (O), обратными ролями (I) и количественными ограничениями (N). Более подробно рассмотрено в таблице 1[3].

---

[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.

[3] Rudolph, S. Foundations of Description Logics // Lecture Notes in Computer Science, – 2011. vol. 6848, – p. 76-136.

Таблица 1
Виды некоторых различных DL[3]

Обозначение	Значение
ALC	Атрибутивный язык с дополнением (отрицанием) произвольных концептов.
S	Аксиомы транзитивности для ролей.
E	Полная экзистенциальная квантификация ($\exists R.C$).
U	Объединение концептов ($\sqcup$).
H	Иерархия ролей (суб-роли).
R	Дополнительные аксиомы для ролей (рефлексивность, иррефлексивность, и дизъюнктность).
O	Номиналы (концепты, состоящие из одного индивида).
N	Количественные ограничения на кардинальность (неквалифицированные).
Q	Квалифицированные количественные ограничения на кардинальность.
I	Обратные роли ($R-$).
F	Функциональные свойства для ролей.

Существенной особенностью данных логик является ***предположение об открытом мире (англ. Open World Assumption, OWA)***[3], при котором отсутствие информации рассматривается как неполнота знаний, а не как ложность утверждений. В этой связи язык OWL2[4] рассматривается как базовое средство представления онтологического.

OWL 2 DL это самый выразительный из стандартов, основан на логике SROIQ(D). Эта логика расширяет SHOIN(D) сложными аксиомами для ролей (R) и квалифицированными количественными ограничениями (Q). Fuzzy OWL2[5] является нечётким расширением для описания нечётких концептов, ролей и аксиом. Fuzzy OWL2 имеет следующие свойства[4]:

[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

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

– *Нечёткие концепты* – принадлежность индивида к нечёткому классу посредством степени;
– *Нечёткие роли* – принадлежность между индивидами или индивидами и значениями данных;
– *Нечёткие типы данных* – определяют нечёткие множества над стандартными типами данных;
– *Нечёткие модификаторы* – применяют лингвистические модификаторы;
– *Нечёткие аксиомы* – расширяют стандартные аксиомы OWL2, позволяя им иметь степень истинности:
  – *Нечёткие утверждения о принадлежности*.
  – *Нечёткие утверждения о ролях*.
  – *Нечёткие аксиомы включения*.
  – *Другие нечёткие аксиомы* – расширения для эквивалентности классов, дизъюнктности, и т.д.

Интеграция Fuzzy OWL2 в системы XAI открывает значительные возможности для создания более гибких, семантически насыщенных и достоверных объяснений[6,7].

В **третьей главе** предложена методика моделирования новых семантических свойств, направленных на повышение объяснимости XAI-систем. В качестве ключевого элемента предложены ***«объяснительные» свойства***, представляющие собой семантические атрибуты, формируемые не только на основе наблюдаемых характеристик данных, но и на основе логического вывода, экспертных знаний, пользовательских ментальных моделей (ММ) и сходств между объектами.

Общая схема построения онтологически ориентированной XAI-системы представлена на рисунке 2. Она отражает последовательность перехода от экспертных знаний и исходных

---

[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.

данных к формированию семантически интерпретируемого объяснения.

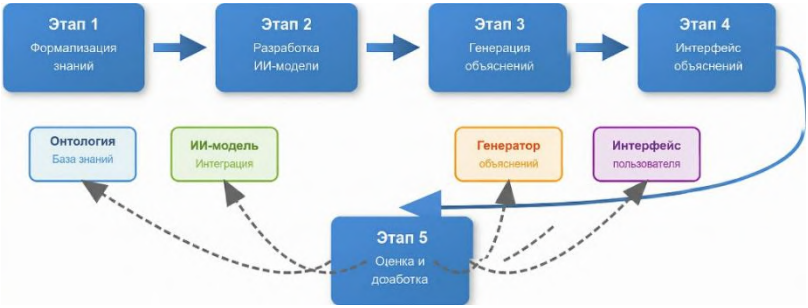

Рисунок 2. Схема создания XAI на основе онтологий

Было определено, что создание эффективного XAI объяснение бесполезно, если оно не находит отклика в когнитивных структурах пользователя, где основное место занимают ***ментальные модели (ММ)***. В предлагаемом подходе ментальные модели рассматриваются как необходимый компонент: модель пользователя задаёт требования к форме и содержанию объяснения, тогда как модель эксперта выступает источником концептов, признаков и связей, включаемых в онтологию.

Распределённые ММ улучшают координацию, коммуникацию, позволяют предсказывать действия друг друга, быстрее адаптироваться к изменениям и принимать более эффективные совместные решения[8]. Математически это можно представить как непустое пересечение множеств, представляющих ММ членов команды: $M_A \cap M_B \cap M_C \neq \emptyset$ где M_A, M_B, M_C являются ММ отдельных личностей. Если такого общего пересечения нет как в $M_A \cap M_B \cap M_C = \emptyset$, команда не обладает полностью разделяемой моделью.

[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.

На этой основе предложена процедура выбора «объяснительных» свойств, включающая анализ предметной области, привлечение экспертных знаний, учёт уровня подготовки пользователя, анализ данных и контекста задачи, а также итеративную корректировку набора свойств по результатам оценки их интерпретационной полезности.

Предложенная концепция *“объяснительных” свойств*[9,10] представляет собой семантические атрибуты, которые выходят за рамки развития, семантически обогащённого XAI. Эти свойства представляют собой семантические атрибуты, которые выходят за рамки простого описания наблюдаемых характеристик и строятся на основе:

- Логических выводов относительно данных и их взаимосвязей;
- Экспертных знаний в конкретной предметной области;
- ММ пользователей, учитывающих их уровень экспертизы и способ восприятия информации;
- Выявленных сходств и аналогий между объектами или экземплярами данных.

Основная цель «объяснительных» свойств – предоставить более глубокое и интуитивно понятное обоснование решений, принимаемых моделями МО, путём учёта не только явных, но и неявных, выведенных или контекстуальных аспектов *различных типов* данных.

Ключевые характеристики и преимущества “объяснительных” свойств:

- *Универсальность и гибкость*. Возможность для построения единообразных объяснительных механизмов для гетерогенных данных;

[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.

– *Обоснованность экспертными знаниями и ММ*. Формируются с учётом того, как эксперты и пользователи (с различным уровнем подготовки) концептуализируют данные и интерпретируют информацию, что повышает релевантность и понятность генерируемых объяснений;
– *Контекстуальность и субъективность*. Создаются более персонализированные и адаптированные объяснения, а не стремяться к универсальному, но потенциально менее релевантному решению;
– *Основа для глубоких и содержательных объяснений*. Позволяют формировать более полные, точные и содержательные объяснения решений моделей ИИ.

«Объяснительные» свойства были определены как множество $P_{\text{exp}} \subseteq R$, где R является множеством связей и каждое свойство $p \in P_{\text{exp}}$ может способствовать объяснению конкретного случая d_i с нечёткой степенью принадлежности $\mu_p(d_i) \in [0,1]$. Эти свойства служат семантическими блоками для построения объяснений в рамках предлагаемой структуры. Такие свойства должны удовлетворять нескольким критериям:

– Доступность как для экспертов, так и для неспециалистов;
– Возможность представления неоднозначных или неопределённых данных;
– Совместимость с онтологическим выводом в OWA;
– Простота и интерпретируемость.

Ключевым аспектом онтологического представления является чёткое определение областей (Domain) и диапазонов (Range) для каждого «объяснительного» свойства. Например, для свойства hasWeatherType, доменом может быть класс Clothes (Одежда), а диапазоном – класс WeatherType (Тип погоды). Использование как позитивных, так и негативных ограничений на свойства (например, Sandal that Not hasWeatherType some Cold) позволяет создавать более точные и детализированные семантические описания в онтологии.

Процедура выбора свойств – это методологический подход, основанный на принципах и этапах для идентификации

наиболее релевантных семантических атрибутов в данных с целью генерации осмысленных объяснений:

- Глубокий анализ предметной области и привлечение экспертных знаний;
- Ориентация на пользователя и учёт ММ;
- Анализ данных и контекста задачи;
- Целевая направленность на улучшение объяснимости;
- Итеративность и адаптивность процесса;
- Учёт возможностей онтологического моделирования;
- Признание субъективности и контекстуальности.

Далее в главе заложена основа ***нечёткой объяснимости*** посредством распределённых ММ. Наше исследование привело к выводу, что ММ способствуют представлению нечёткой объяснимости и стремятся генерировать объяснения, которые:

- *Соответствуют когнитивным представлениям:* используя лингвистические переменные (“высокий”, “низкий”, “вероятно”) и степени уверенности, нечёткие объяснения могут напрямую отображаться на нечёткие концепты в ментальной модели пользователя.
- *Отражают реальную неопределенность:* вместо того чтобы скрывать или игнорировать неопределённость, присущую данным или работе модели, нечёткие объяснения делают её явной, что позволяет пользователю сформировать более адекватную ментальную модель ситуации.
- *Обеспечивают градации:* они показывают степень влияния того или иного фактора, что позволяет пользователю построить более тонкую и детальную ментальную модель причинно-следственных связей.

В той же главе рассмотрены подходы к оценке качества объяснений. Обосновано, что анализ объяснимости должен опираться на сочетание количественных и качественных метрик: первые характеризуют согласованность, различимость и устойчивость объяснений, вторые – их понятность, глубину, усвояемость и удобство для пользователя.

В **четвёртой главе** разработана формальная модель нечёткой объяснимости и предложена архитектура системы, реализующей данный подход. Нечёткая объяснимость определяется как способность XAI-системы генерировать объяснения, связывающие решение модели с семантическими концептами предметной области и одновременно отражающие степень уверенности в этих связях.

Соответствующая архитектура представлена на рисунке 3. В ней входными компонентами выступают наблюдения, ментальные модели и «объяснительные» свойства, центральным элементом является модуль нечёткого объяснения, а выходом - интерпретируемое решение, дополненное новыми нечёткими знаниями.

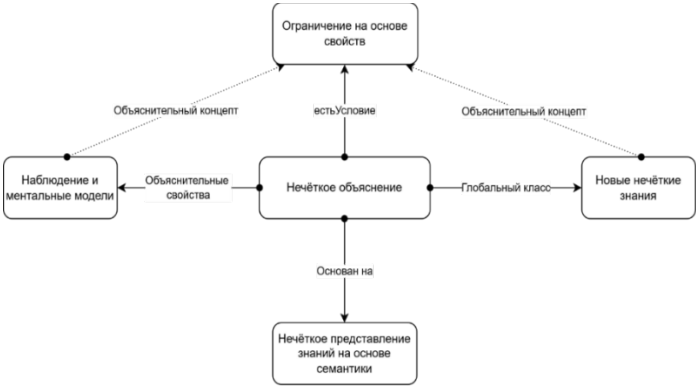

Рисунок 3. Дизайн системы, определяющий предложенные нечёткие объяснения на основе онтологии

Например, кредитный риск может быть основан различных объяснимых свойствах, которые получены на основе *наблюдений и отображают распределённый ММ*. В нашем семантически-ориентированном *нечётком представлении знаний* – онтология кредитного риска может определять такие концепты, как ShortDuration (короткая длительность), SmallCredit (небольшой кредит), LittleSaving (небольшие сбережения) и т.д., и связывать их с концептом кредитного риска. *Ограничения,*

основанные на свойствах (например, hasDuration, hasCreditAmount, hasSavingAccount и др.), уточняют объяснение посредством *нечётких условий*. На основе набора данных извлекаются *новые нечёткие знания* определяя *глобальный класс*. Вывод системы формирует *нечёткое объяснение*, где заёмщик обладает GoodRisk (0.94), что поддерживается свойствами ShortDuration (0.67), SmallCredit (0.55), LittleSaving (0.54).

Необходимость фаззификации «объяснительных» свойств обусловлена нечёткостью исходных данных, неопределённостью экспертных знаний и потребностью в градуированном представлении факторов, влияющих на вывод модели. В отличие от бинарного подхода, такая формализация позволяет описывать ситуации, в которых объект имеет класс или свойство лишь в определённой степени.

Цель нечёткой объяснимости – генерировать объяснения, которые не только связывают прогнозы модели с семантическими концептами, но и отражают степень уверенности в этих связях и присущую им неопределённость. Это особенно важно в средах с нечёткими знаниями, где бинарные ответы неадекватны.

Формально ***фреймворк нечёткой объяснимости*** был определён как $FE = \left(M, O_f, P_{\text{exp}}, K_f\right)$, где M представляет общую распределённую ММ, O_f обозначает нечёткую онтологию, P_{exp} – множество объяснительных свойств, а K_f соответствует нечётким знаниям, извлечённым из набора данных. ММ способствуют нашему представлению нечёткой объяснимости соответствуя когнитивным представлениям знаний пользователя и эксперта; отражая реальную неопределённость знаний; обеспечивая градации и влияния факторов.

Нечёткое объяснение для решения или события y выражается как $E_f(y) = \{\left(x_i, \mu(x_i)\right) \mid x_i \in X, \mu(x_i) \in [0,1]\}$, где $X = x_1, x_2, \dots, x_n$ – это множество объясняющих факторов, а $\mu(x_i)$ представляет степень вклада фактора x_i в объяснение y.

Для практической реализации предложенного подхода сформирована архитектура, основанная на использовании Fuzzy

OWL2, нечётких «объяснительных» свойств, нечётких аксиом для объяснений, механизмов вывода степеней уверенности и процедур представления объяснений пользователю[11].

В качестве механизма нечёткого логического вывода рассматривается fuzzyDL, поддерживающий нечёткие дескриптивные логики и обработку онтологий в формате Fuzzy OWL2. Алгоритм fuzzyDL основан на комбинации табло-алгоритма и решения задачи оптимизации. На таблице 2 изображены некоторые правила нечёткого ALC с пустой ABox[12].

Таблица 2
Логический вывод нечёткого ALC с пустой ABox

Правило	**Предварительное условие**	**Действие**
$(\bot)$	$\bot \in \mathcal{L}(v)$	$C = C \cup \{x_{v:\bot} = 0\}$
$(\top)$	$\top \in \mathcal{L}(v)$	$C = C \cup \{x_{v:\top} = 1\}$
$(\neg)$	$\neg A \in \mathcal{L}(v)$	$C = C \cup \{x_{v:\neg C} = \ominus x_{v:C}\}$
$(\sqcap)$	$C_1 \sqcap C_2 \in \mathcal{L}(v)$	$\mathcal{L}(v) = \mathcal{L}(v) \cup \{C_1, C_2\}$ $C = C \cup \{x_{v:C} \otimes x_{v:D} = x_{v:C_1 \sqcap C_2}\}$
$(\sqcup)$	$C_1 \sqcup C_2 \in \mathcal{L}(v)$	$\mathcal{L}(v) = \mathcal{L}(v) \cup \{C_1, C_2\}$ $C = C \cup \{x_{v:C} \oplus x_{v:D} = x_{v:C_1 \sqcup C_2}\}$
$(\exists)$	$\exists R.C \in \mathcal{L}(v)$	создать новый узел w $\mathcal{L}(\langle v,w\rangle) = \mathcal{L}(\langle v,w\rangle) \cup \{R\}$, и $\mathcal{L}(w) = \mathcal{L}(w) \cup \{C\}$, и $C = C \cup \{x_{\langle v,w\rangle:R} \otimes x_{w:C} = z, z \geq x_{v:\exists R.C}\}$
$(\forall)$	$\forall R.C \in \mathcal{L}(v)$ $R \in \mathcal{L}(\langle v,w\rangle)$	$\mathcal{L}(w) = \mathcal{L}(w) \cup \{C\}$ $C = C \cup \{x_{v:\forall R.C} \geq z, z = x_{(v,w):R} \Rightarrow x_{w:C}\}$

[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.

Ключевым архитектурной программной реализации является парадигма «модель на свойство». Вместо одной монолитной модели используется ансамбль, включающий глобальный классификатор и набор отдельных классификаторов, каждый из которых отвечает за оценку одного из «объяснительных» свойств. Это позволяет непосредственно связать результаты работы моделей с онтологическим представлением знаний. Рисунок 4 представляет предложенную новую архитектуру. Такое построение системы позволяет не только повысить модульность, но и сделать процесс принятия решения более прозрачным. Результаты работы любой отдельной модели являются отдельным блоком для итогового объяснения.

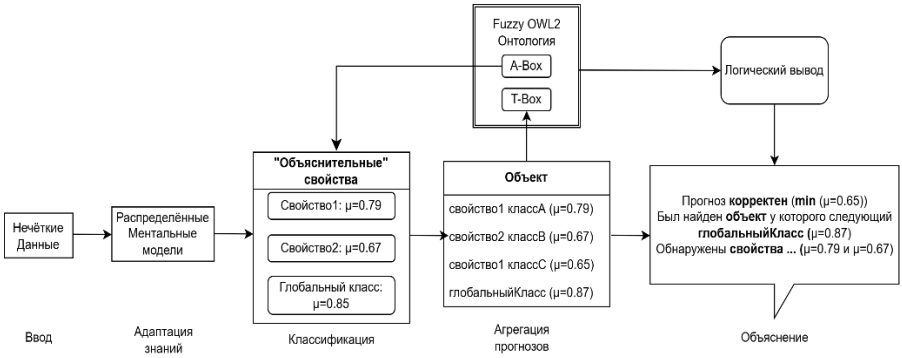

Рисунок 4. Предложенная архитектура для новой XAI системы

Предлагаемая схема функционирования новой системы. Программный комплекс работает по чётко определённому принципу, который можно разбить на следующие шаги:

Шаг 1. Предобработка данных: на вход система получает данные об объекте (например, строка из таблицы или изображение). Данные проходят этап подготовки: для данных это может включать кодирование категориальных признаков, нормализацию или дискретизацию числовых значений для приведения их к интервальным форматам, описанным в

онтологии (например, «Низкий», «Средний», «Высокий» возраст). Также происходит интеграция распределённых ментальных моделей.

Шаг 2. Параллельная классификация: подготовленный вектор признаков одновременно подаётся на вход глобальному классификатору и всем N классификаторам свойств. Каждый из них возвращает свой прогноз (целевой класс и значения для каждого из «объяснительных» свойств).

Шаг 3. Наполнение онтологии: полученные прогнозы используются для программного создания экземпляра в ABox онтологии. Например, для нового пациента создаётся индивид класса Patient, и ему через объектные свойства присваиваются связи с соответствующими индивидами классов-свойств. Процесс идёт в сочинение с распределёнными ментальными моделями.

Шаг 4. Представление нечётких знаний: на этом этапе происходит интеграция нечёткости. Степени принадлежности, полученные от МО моделей или вычисленного алгоритмом нечёткой кластеризации, используются для аннотирования созданных утверждений в онтологии.

Шаг 5. Логический вывод и проверка согласованности: после наполнения онтологии данными о конкретном индивиде запускается логический вывод. Он выполняет ключевую функцию, а именно проверяет, не противоречат ли новые утверждения существующим в онтологии аксиомам и ограничениям

Шаг 6. Генерация объяснения: на заключительном этапе система агрегирует результаты. Извлекается из онтологии итоговый класс, все *«объяснительные» свойства*, и их степени уверенности посредством нечётких значений. Эта информация форматируется в структурированный вид, который далее может быть передан в любой пользовательский интерфейс для визуализации.

Таким образом, разработанный программный комплекс представляет собой гибкую и масштабируемую платформу. Он эффективно решает задачу построения объяснимых систем,

способных работать с неполными и нечёткими знаниями. Такая платформа позволяет объяснять результаты классификации, полученные с помощью модели типа «чёрный ящик» и также задач кластеризации. Это можно рассматривать как важный шаг к созданию более надёжных и заслуживающих доверия систем XAI.

В **пятой главе** представлены результаты экспериментальной апробации, верификации и оценка предложенного нечётко-семантического подхода.

Верификация методологии была выполнена на 6 наборах данных, относящихся к различным областям: Fashion MNIST (компьютерное зрение), Glass Identification (материаловедение), German Credit Risk (кредитное дело), Heart Failure Clinical Records (прогнозирование в медицине), Network Traffic Data for Malicious Activity Detection (кибербезопасность в компьютерной сети), Student Performance (успеваемость студентов). Такой выбор позволил оценить применимость предложенного подхода к разнородным типам данных и различным сценариям интерпретации.

Экспериментальная апробация проводилась в три этапа. На первом этапе верифицировалась базовая концепция «объяснимых» свойств в среде классической OWL2-онтологии. На втором этапе исследовалась гибридная схема, в которой чёткая онтология дополнялась количественным представлением неопределённости. На третьем этапе осуществлён полный переход к Fuzzy OWL2, что позволило напрямую интерпретировать выходы моделей как степени принадлежности для нечётких аксиом. Сравнение результатов трёх этапов показало, что именно полный нечёткий вариант обеспечивает наибольшую выразительность, гибкость и содержательность формируемых объяснений.

На ***первом этапе*** верификации проверялась ***базовая концепция*** «объяснительных» свойств. Для этого использовались онтология OWL2 и стандартный механизм логического вывода. На рисунке 5 приведён пример структуры соответствующей онтологии, полученной в ходе эксперимента[9].

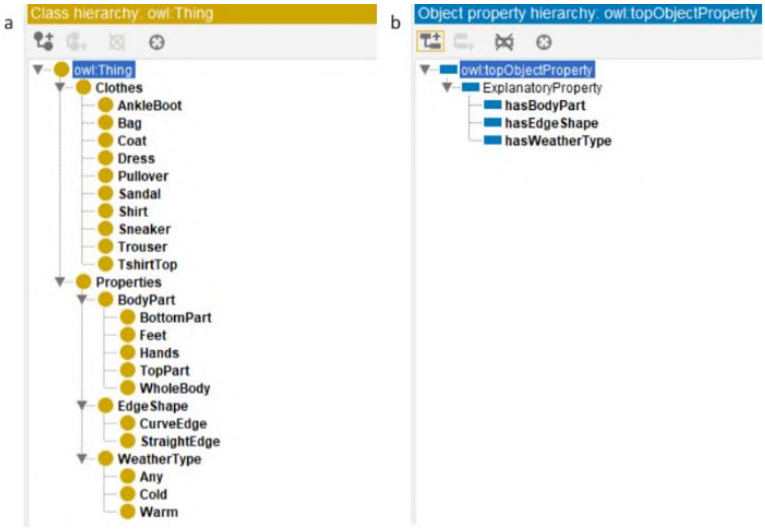

Рисунок 5. (a) Иерархии классов и (b) свойств объектов для данных Fashion MNIST

Модель «чёрного ящика» провела классификацию набора данных. В итоге была подтверждена принципиальная возможность представления «объяснительных» свойств средствами классической OWL2-онтологии и стандартного логического вывода для объяснения работы классификатора.

Однако объяснение было перегружено большом количеством информации и были учтены значения свойств как со стороны принадлежности к классу, так и со стороны непринадлежности. Данная система не имела нечёткие значения что и привело к такой перегрузке. Пример объяснения далее:

*«**Image_0** является классом TshirtTop и является **корректным** в онтологии. **Не** имеет WholeBody в BodyPart, имеет TopPart in BodyPart, **не** имеет BottomPart в BodyPart, **не** имеет Feet в BodyPart, **не** имеет Hands в BodyPart, **не** имеет Cold в WeatherType, имеет Warm в WeatherType, **не** имеет Any в WeatherType, **не** имеет StraightEdge in EdgeShape, имеет CurveEdge в EdgeShape”.»*

Ещё одним подтверждением гибкости свойств является эксперимент[13] на наборе данных “Glass Identification”, где требовалось классифицировать тип стекла на основе его химического состава. Была создана онтология, где “объяснительными” свойствами выступали сами химические элементов.

На ***втором этапе*** был реализован ***гибридный вариант подхода***, при котором чёткая онтология OWL2 дополнялась количественным представлением неопределённости. Для хранения степени принадлежности использовалось специальное свойство данных, выступающее контейнером для нечётких характеристик внутри формально чёткой структуры. Пример онтологии этого типа приведён на Рисунке 6.

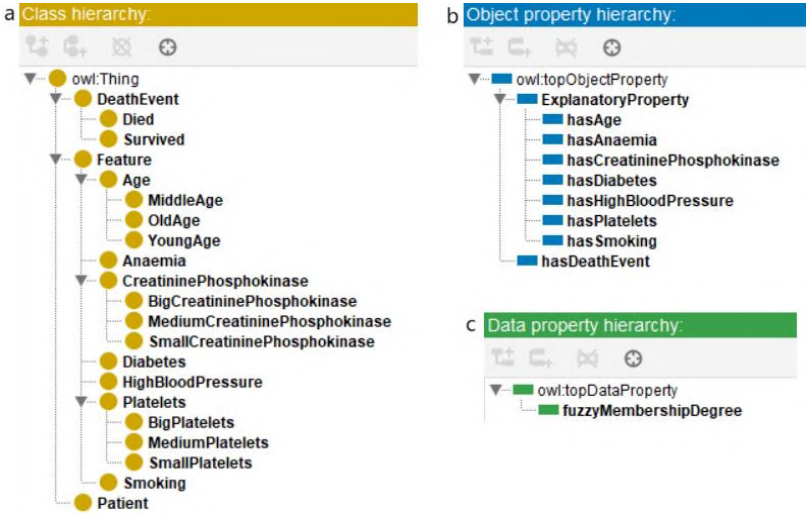

Рисунок 6. Иерархии классов (a), свойств объектов (b) и свойств данных (c) для набора данных о сердечной недостаточности

[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.

Для помощи логического вывода также используются правила семантической сети (англ. Semantic Web Rule Language, SWRL). Полноценная объяснимость достигается за счёт того, что правила включают все существующие в онтологии «объяснительные» свойства (Рисунок 7).

```
Patient(?p) ∧
        hasAge(?p, OldAge) ∧
        hasAnaemia(?p, true) ∧
        hasCreatininePhosphokinase(?p, MediumCreatininePhosphokinase) ∧
        hasDiabetes(?p, true) ∧
        hasHighBloodPressure(?p, true) ∧
        hasPlatelets(?p, SmallPlatelets) ∧
        hasSmoking(?p, true)
→ hasDeathEvent(?p, Died)
```

```
Patient(?p) ∧
        hasAge(?p, MiddleAge) ∧
        hasAnaemia(?p, false) ∧
        hasCreatininePhosphokinase(?p, SmallCreatininePhosphokinase) ∧
        hasDiabetes(?p, false) ∧
        hasHighBloodPressure(?p, false) ∧
        hasPlatelets(?p, MediumPlatelets) ∧
        hasSmoking(?p, false)
→ hasDeathEvent(?p, Survived)
```

Рисунок 7. Возможные SWRL правила для логического вывода в онтологии для набора данных о сердечной недостаточности[7]

Доказано, что введение количественной характеристики неопределённости в чёткую онтологическую структуру позволяет повысить гибкость представления факторов объяснения, хотя сама модель ещё не обеспечивает полноценной нечёткой семантики. Пример объяснение приведён ниже:

*«**Entry_0** для класса Patient является класс Survived. Он **корректен** в онтологии. Объяснительные свойства следующие: MiddleAge для hasAge, false длч hasAnaemia, SmallCreatininePhosphokinase для hasCreatininePhosphokinase, false для hasDiabetes, false для hasHighBloodPreassure, MediumPlatelets для hasPlatelets, false для hasSmoking. Целевой класс является survived для hasDeathEvent, и имеет нечёткую принадлежность fuzzyMembershipDegree как 0.8.»*

Благодаря нечёткой функции принадлежности получилось получить меньше информации в объяснение сохранив основную суть. Однако отсутствие нечёткой принадлежности для «объяснительных» свойств оставляет проблему представление неточных знаний и делает объяснения недостаточно ясными.

Завершающий этап соответствовал полному переходу к ***нечёткой объяснимости*** на базе Fuzzy OWL2. В качестве экспериментальной базы использовался набор данных Student Performance, содержащий признаки, естественным образом допускающие лингвистическую и нечёткую интерпретацию. В этой конфигурации количественные атрибуты представлялись через лингвистические переменные и нечёткие термы, а выходы моделей непосредственно интерпретировались как степени принадлежности для нечётких аксиом[10].

Программная конфигурация системы претерпела ключевые изменения:

- *Нечёткая онтология (Fuzzy OWL2)* позволила внедрить следующие фундаментальные нечёткие конструкции:
  - *Нечёткие аксиомы:* каждое утверждения в ABox теперь с определённой степенью уверенности, которая напрямую соответствует выходному значению модели МО;
  - *Лингвистические переменные:* количественные атрибуты (например, “время подготовки”) были представлены как лингвистические переменные с нечёткими термами (“недостаточное”, “умеренное”, “значительное”). Процесс определения этих термов и их функций принадлежности (например, трапециевидных или треугольных) был основан на экспертных знаниях в области педагогики, что позволило заложить в модель человеческое понимание предметной области.
- *Архитектура***:** сохранился принцип “модель на свойство”, но теперь выходные результаты моделей (например, вероятность 0.85 для свойства hasStudyTime) напрямую и органично интерпретировались как степени принадлежности для нечётких аксиом, создавая интеграцию между выводом МО и семантическим представлением знаний.

Было установлено, что использование Fuzzy OWL2 обеспечивает наиболее естественное и полное представление

степеней принадлежности, благодаря чему объяснения становятся более содержательными, градуированными и интерпретируемыми.

Пример нечёткой аксиомы для «объяснительного» свойства приведён на рисунке 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>
```

Рисунок 8. Пример “объяснимого” свойства с нечёткой аксиомой Fuzzy OWL2 на примере данных учащихся

Результаты показали, что в полном нечётком варианте система формирует более содержательные и гибкие объяснения, в которых каждое «объяснительное» свойство сопровождается мерой уверенности. Это позволяет не только указать, какие факторы повлияли на решение, но и отразить степень их участия в формировании итогового вывода[12]. Пример вывода системы:

«*Учащийся (Student_0) классифицирован как GoodAcademicPerformance (с уверенностью 0.88). Это решение основано на следующих факторах: HighStudyTime (0.85), HighGrades (0.79), GoodFamRel (0.67), LittleAbsences (0.62) и MediumGoOut (0.58)*».

Этот пример демонстрирует ключевые преимущества нашего подхода, так как система не только выделяет релевантные характеристики, повлиявшие на положительное решение, но и количественно оценивает вклад каждого фактора.

В главе было отмечено, что хотя точность классификации в нашей системе не превзошла существующие решения, основной целью исследования было именно повышение интерпретируемости предсказаний. В этом отношении нечёткие

онтологические структуры показали существенное преимущество по сравнению с классическими подходами.

Сравнительный анализ разработаны нечёткого и чёткого подходов. Чтобы наглядно продемонстрировать разницу в объяснениях, был проведён сравнительный анализ двух подходов на основе «объяснительных» свойств наборе данных German Credit Risk.

В *чётком подходе* используется стандартная онтология OWL2. Для каждого «объяснительного» свойства (концепта) была обучена отдельная модель МО. Выходные данные моделей (например, предсказанный класс LittleSaving) преобразовывались в абсолютные, бинарные утверждения, которые добавлялись в ABox (рисунок 9).

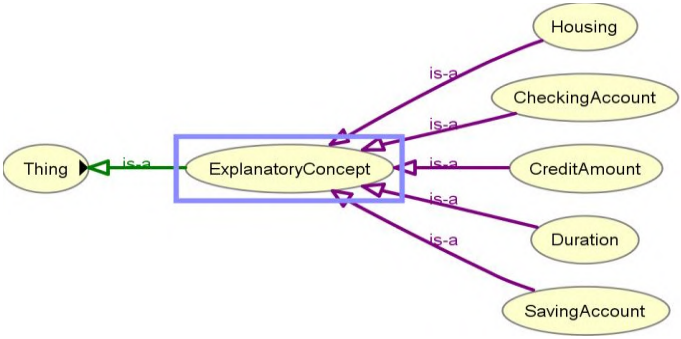

Рисунок 9. “Объяснимый” концепт для набора данных кредитного риска

В *нечёткий подход* применялась нечёткая онтология Fuzzy OWL2. Это позволило представлять такие понятия, как SmallCredit (небольшой кредит) или LongDuration (длительный срок) не как дискретные категории, а как лингвистические переменные с плавающими степенями принадлежности. Процесс формирования объяснения заключался в агрегации этих нечётких утверждений и их степеней уверенности.

Для поддержки логического вывода были разработаны 15 правил на языке SWRL (9 для GoodRisk и 6 для BadRisk), которые

связывали комбинации свойств с финальным решением. Например: «*LoanTaker(?x) ^ hasCheckingAccount(?x, NAChecking) ^ hasSavingAccount(?x, LittleSaving) -> hasRisk(?x, Bad)*».

Далее была проведена ***оценка и сравнительный анализ разработанного метода с другими подходами***. Основной целью является верификация эффективности и надёжности предложенных методов объяснения.

Для обеспечения сопоставимости результатов были зафиксированы одни и те же исходные условия: один и тот же элемент выборки, одно и то же предсказание модели и одно и то же пространство свойств. Тем самым исключалось влияние смены объекта наблюдения, набора факторов и итогового прогноза, а сравнение переводилось в плоскость различий между самими механизмами построения объяснения.

В качестве количественных критериев использовались евклидово расстояние, коэффициент Жаккара, взвешенное нечёткое среднее, байесовская вероятность согласованности объяснения и также модель взвешенной суммы на основе Z-чисел. Были сравнены между собой LIME, SHAP, контрфактические объяснения, объяснения на основе нечётких правил, и объяснения на основе нечётких деревьев решений.

Тем не менее для достоверной оценки также была опрошена группа экспертов и пользователей для того, чтобы понять отображается ли вся информация в полученных объяснениях. Была проведена оценка понятности, удовлетворённости, доверия, действенности и действенности.

Полученные результаты подтверждают целесообразность перехода от чёткой семантической модели к нечётко-онтологической архитектуре как направлению развития XAI-систем. Предложенный подход обеспечивает более глубокие, контекстно насыщенные и когнитивно совместимые объяснения, а верификация доказывает, что интеграция онтологий и нечёткой логики создаёт формальную основу для повышения доверия к результатам машинного обучения и умение работать с нечёткими знаниями. Рабочая версия научно-исследовательского кода

предлагаемой системы и инструкция по применению были выложены на онлайн ресурсе GitHub[14].

В **заключение** были приведены научные и практические выводы, полученные в ходе выполнения работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Основные результаты полностью соответствуют поставленным задачам и заключаются в следующем:

1. Проведён анализ современных подходов к построению систем объяснимого ИИ, в результате которого выявлены ограничения существующих алгоритмических и онтологических методов объяснения, связанные соответственно с недостаточной семантической глубиной и невозможностью адекватного учёта неопределённости реальных данных.
2. Разработана и формализована концепция объяснительных свойств как средства онтологического представления экспертных знаний, обеспечивающего интерпретацию результатов интеллектуального анализа данных с учётом логических зависимостей, функционального назначения объектов и особенностей восприятия пользователя.
3. Обоснована целесообразность использования концепции ментальных моделей пользователя при построении объяснений и показано, что согласование объяснения с когнитивными представлениями пользователя повышает его понятность и прикладную ценность для различных категорий пользователей.
4. Предложена и теоретически обоснована методология нечёткой объяснимости, основанная на интеграции семантических технологий и методов нечёткой логики для

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

представления и обработки неточных знаний в задачах объяснимого искусственного интеллекта.

5. Разработан метод представления объяснительных свойств и онтологических аксиом с использованием нечётких онтологий Fuzzy OWL2, обеспечивающий формирование объяснений с количественной оценкой степени уверенности.
6. Разработана процедура обработки результатов нечёткой кластеризации, позволяющая автоматически формировать утверждения о свойствах объектов в нечёткой онтологии на основе степеней принадлежности, полученных из данных.
7. Разработан механизм генерации нечётких семантических объяснений на основе аппарата нечётких дескрипционных логик, обеспечивающий получение логически обоснованных и интерпретируемых объяснений результатов работы моделей типа «чёрный ящик» без модификации самих моделей.
8. Проведена экспериментальная апробация предложенной методологии на шести наборах данных из различных предметных областей. Полученные результаты подтвердили универсальность, масштабируемость и практическую эффективность разработанного подхода.

Предложенная методология закладывает научный и практический фундамент для нового поколения систем XAI. Было доказано, что система, интегрирующая семантическую сеть и нечёткую логику способны адекватно отображать реальную неопределённость данных из реального мира. Также позволяет создавать не только более прозрачные, но и более надёжные и заслуживающие доверия системы искусственного интеллекта.

Основные результаты диссертации опубликованы в следующих научных работах:

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.

Личный вклад соискателя в трудах, опубликованных в соавторстве:

[1], [3], [5], [6], [7], [9] – теоретические исследования, программная имплементация, проведение верификации, и оформление исследований в статейном формате.

Защита диссертации состоится 05 мая 2026 года в 14:00 на заседании Диссертационного совета FD 2.48, действующего на базе Азербайджанского Государственного Университета Нефти и Промышленности.

Адрес: AZ 1010, Баку, проспект Азадлыг, 34.

С диссертацией можно ознакомиться в библиотеке Азербайджанского Государственного Университета Нефти и Промышленности.

Электронная версия автореферата размещена на официальном сайте Азербайджанского Государственного Университета Нефти и Промышленности.

Автореферат разослан по
соответствующим адресам 02 апреля 2026 года.

Подписано в печать: 12.03.2026

Формат бумаги: А5

Объём: 37 500

Тираж: 70