

Сетевой научно-практический журнал
**ИНТЕЛЛЕКТУАЛЬНЫЙ
ТРАНСПОРТ**

№3 (39)
сентябрь 2026



ОБЗОРНЫЕ РАБОТЫ
ПО СОВРЕМЕННЫМ
НАУЧНЫМ ДОСТИЖЕНИЯМ
В ОБЛАСТИ ТРАНСПОРТА

ИСКУССТВЕННЫЙ
ИНТЕЛЛЕКТ
И МАШИННОЕ
ОБУЧЕНИЕ

ИНТЕЛЛЕКТУАЛЬНЫЕ
ТРАНСПОРТНЫЕ
СИСТЕМЫ

ИНТЕЛЛЕКТУАЛЬНЫЙ ТРАНСПОРТ

Издается с 2017 года

Издатель:



АО «НИИАС»

Email: journal@vniias.ru
Телефон: +7 (495) 967-77-01

Сайт: <http://www.intelligent-transport.ru/>

Журнал включен в базу данных
Российского индекса научного
цитирования (РИНЦ)

Полная или частичная перепечатка,
сканирование любого материала текущего
номера возможны только с письменного
разрешения редакции.

Адрес редакции:

109029, г. Москва, Нижегородская ул.,
д. 27, стр. 1

Сетевой научно-практический журнал
«Интеллектуальный транспорт» публикует
основные научные результаты соискателей
учёной степени кандидата и доктора наук по
следующим специальностям и
соответствующим им отраслям науки:

- 1.2.1. Искусственный интеллект и машинное обучение (технические науки)
- 2.9.1. Транспортные и транспортно-технологические системы страны, ее регионов и городов, организация производства на транспорте (технические науки)
- 2.9.4. Управление процессами перевозок (технические науки)
- 2.9.8. Интеллектуальные транспортные системы (технические науки)

Зарегистрирован в Федеральной службе
по надзору в сфере связи,
информационных технологий
и массовых коммуникаций (Роскомнадзор),
свидетельство о регистрации
Эл № ФС77-91565 от 6 мая 2026 г.

ГЛАВНЫЙ РЕДАКТОР

Суханов Андрей Валерьевич

к.т.н., доцент, главный эксперт АО «НИИАС», член
Российской ассоциации искусственного
интеллекта, эксперт комиссии по вопросам
научно-технологического развития и
образования Общественного совета при
Минтрансе России

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Алиев Вугар Амирович

д.ф.-м.н., профессор, Генеральный директор
AMIR Technical Services Company, Республика
Азербайджан

Бородин Андрей Федорович

д.т.н., профессор, начальник отдела АО «ИЭРТ»

Бочков Константин Афанасьевич

д.т.н., профессор, профессор кафедры «БелГУТ»,
научный руководитель – заведующий НИЛ
«БЭМС ТС», Республика Беларусь

Броневиц Андрей Георгиевич

д.ф.-м.н., главный научный сотрудник
АО «НИИАС»

Бутакова Мария Александровна

д.т.н., профессор, профессор РГУПС

Грибова Валерия Викторовна

д.т.н., чл.-корр. РАН, заместитель директора по
научной работе ДВО РАН

Гришаев Сергей Юрьевич

к.т.н., заместитель Генерального директора –
директор Ростовского филиала АО «НИИАС»

Гуров Юрий Владимирович

к.т.н., начальник отделения АО «НИИАС»

Зуев Александр Валерьевич

д.т.н., начальник управления – заместитель
главного ученого секретаря ДВО РАН,
заведующий лабораторией аэрокосмических и
морских робототехнических комплексов ИАПУ
ДВО РАН

Ковалев Сергей Михайлович

д.т.н., профессор, научный руководитель
Ростовского филиала АО «НИИАС»

Коваленко Николай Иванович

д.т.н., профессор, профессор РУТ (МИИТ)

Котенко Алексей Геннадьевич

д.т.н., профессор, руководитель центра ИИ,
РГУПС

Куренков Петр Владимирович

д.э.н., профессор, профессор РУТ (МИИТ)

Ли Инь

к.ф.-м.н., доцент Харбинского технологического
института, КНР

Ольгейзер Иван Александрович

к.т.н., доцент, заместитель начальника
опытно-конструкторского бюро АО «НИИАС»

Папич Любиша

д.т.н., профессор, директор Исследовательского
центра DQM, Республика Сербия

Попов Павел Александрович

к.т.н., заместитель директора
Научно-исследовательского института
транспорта РУТ (МИИТ)

Сабанов Алексей Геннадьевич

д.т.н., доцент, главный эксперт АО «НИИАС»

Соколов Сергей Викторович

д.т.н., профессор, профессор РГУПС, главный
научный сотрудник АО «НИИАС»

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА

Розенберг Ефим Наумович

д.т.н., профессор, первый заместитель
Генерального директора АО «НИИАС»

Хабаров Валерий Иванович

д.т.н., профессор, академик Академии
транспорта РФ, заведующий кафедрой СГУПС

Хатламаджиян Агоп Ервандович

к.т.н., доцент, заместитель Генерального
директора АО «НИИАС»

Чжан Дачжи

профессор Харбинского технологического
института, КНР

Шабалин Николай Григорьевич

д.т.н., профессор, научный руководитель
АО «НИИАС»

Шубинский Игорь Борисович

д.т.н., профессор, главный эксперт
АО «НИИАС», заместитель генерального
директора ООО «ИБ Транс»

Юдин Дмитрий Александрович

к.т.н., старший научный сотрудник,
заведующий лабораторией МФТИ

Яковлев Константин Сергеевич

к.ф.-м.н., ведущий научный сотрудник ФИЦ
ИУ РАН

РЕДАКЦИОННЫЙ СОВЕТ

Долгий Александр Игоревич

Председатель редакционного совета, к.т.н.,
Генеральный директор АО «НИИАС»

Бочков Александр Владимирович

Заместитель Председателя редакционного
совета, д.т.н., ученый секретарь АО «НИИАС»

Зажигалкин Александр Владимирович

Заместитель Председателя редакционного
совета, д.э.н., главный эксперт АО «НИИАС»

Калинина Илона Владимировна

Заместитель Председателя редакционного
совета, Исполнительный директор
АО «НИИАС»

Павловский Андрей Александрович

Заместитель Председателя редакционного
совета, к.т.н., заместитель Генерального
директора АО «НИИАС»

Чернуха Алексей Викторович

Заместитель Председателя редакционного
совета, к.э.н., начальник Департамента
АО «НИИАС»

РЕДАКЦИЯ

Корабельников Андрей Владимирович

Руководитель редакции, заместитель
Исполнительного директора АО «НИИАС»

Соловьева Юлия Геннадьевна

Ответственный секретарь, главный специалист
АО «НИИАС»

Муллагулова Анастасия Робертовна

Дизайн журнала и сайта, начальник отдела
АО «НИИАС»

Выпускающий редактор

Левшина Анна Андреевна

Специалист АО «НИИАС»

СОДЕРЖАНИЕ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

А.А. Агафонов, А.В. Пономарев

Сравнительный анализ методов повышения точности RAG-систем на специализированных корпусах нормативно-технической документации	4
--	----------

М.И. Кривошей, Д.А. Юдин, М.Н. Зарипов, О.В. Буличев, К.Ф. Муравьев, А.И. Панов

Нейросетевые методы распознавания подземных объектов в данных георадара мобильного колесного робота	23
--	-----------

К.Ф. Муравьев, В.И. Романенко

Построение компактных графов сцены на основе топологической карты для автономной навигации мобильного робота	37
---	-----------

И.С. Михайлов, Мьо Хлайн Вин, К.О. Сидоров

Сравнительное исследование методов обнаружения выбросов в данных мониторинга нефтяных скважин на основе SVM и DBSCAN	53
---	-----------

ИНТЕЛЛЕКТУАЛЬНЫЕ ТРАНСПОРТНЫЕ СИСТЕМЫ

Ю.В. Трофимов, А.Н. Аверкин, А.В. Шевченко, Е.М. Кузнецов, А.Д. Лебедев

Интерпретируемое управление инициацией треков с заранее заданными границами вмешательства в системах восприятия беспилотных летательных аппаратов	69
--	-----------

Р.Д. Давыдов, Г.Д. Ахмадов, М.В. Жабриков

Методика оценки полноты, достоверности и качества данных мониторинга инфраструктуры интеллектуальных транспортных систем	85
---	-----------

Н.В. Павлов, А.А. Свистельников, А.А. Вальковская, А.Р. Платова

Совместная разработка документов транспортного планирования и интеллектуальных транспортных систем	100
---	------------

ОБЗОРНЫЕ РАБОТЫ

А.В. Бочков, А.И. Долгий, А.В. Зажигалкин, Е.Н. Розенберг, А.В. Суханов, А.Е. Хатламаджиян, А.В. Чернуха

Современное состояние систем интеллектуализации, цифровизации и роботизации отечественного железнодорожного транспорта	115
---	------------

УДК 004.896

DOI: 10.24412/3033-6007-2026-339-4-22

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ПОВЫШЕНИЯ ТОЧНОСТИ RAG-СИСТЕМ НА СПЕЦИАЛИЗИРОВАННЫХ КОРПУСАХ НОРМАТИВНО-ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ

Агафонов А.А.^{1, a}, Пономарев А.В.^{1, b}

¹ Санкт-Петербургский Федеральный исследовательский центр Российской академии наук (СПб ФИЦ РАН), Санкт-Петербург, Россия.

E-mail: ^a agafonov.a@spcras.ru, ^b ponomarev@iias.spb.su

Аннотация. В условиях цифровизации промышленности возрастает потребность в интеллектуальных системах поддержки принятия решений, способных эффективно извлекать знания из массивов нормативно-технической документации. В статье рассматриваются интеллектуальные вопрос-ответные системы на основе технологии генерации, дополненной поиском (Retrieval-Augmented Generation, RAG), применительно к специализированному корпусу нормативно-технических документов железнодорожной отрасли. Проводится сравнительный анализ ряда методов повышения точности RAG-систем, включая стратегии обработки запросов (HyDE, Step-Back, Multi-Query, Decomposition, Pseudo-Relevance Feedback, Recursive Refinement), двухэтапное извлечение с использованием кросс-энкодера текстовых фрагментов (вопроса и нормативного пункта), а также дообучение моделей эмбединга текстовых фрагментов на синтетических парах «вопрос – нормативный пункт». Эксперименты показывают, что наиболее существенный прирост качества поиска обеспечивает предметное дообучение моделей эмбединга. Применение кросс-энкодера также дает положительный эффект, особенно для моделей с изначально менее точным ранжированием. В то же время продвинутые стратегии обработки запросов не приводят к значимому повышению качества извлечения по сравнению с базовым подходом. Сформированные корпус документов, наборы вопросов и программный код размещены в открытом доступе.

Ключевые слова: RAG, большие языковые модели, семантический поиск, модели эмбединга, кросс-энкодер, нормативно-техническая документация, железнодорожный транспорт, дообучение.

Введение

В условиях цифровой трансформации промышленности и роста объема неструктурированных текстовых данных возрастает потребность в интеллектуальных системах поддержки принятия решений (СППР), способных оперативно предоставлять лицу, принимающему решение, релевантную информацию. На железнодорожном транспорте, характеризующемся высокими требованиями к безопасности и надежности, особенно остро стоит задача

Для цитирования:

Агафонов А.А., Пономарев А.В. Сравнительный анализ методов повышения точности RAG-систем на специализированных корпусах нормативно-технической документации. *Интеллектуальный транспорт*. 2026. № 3(39). С. 4–22. DOI: 10.24412/3033-6007-2026-339-4-22.

эффективного извлечения знаний из массивов нормативно-технической документации, регламентирующей процессы эксплуатации, обслуживания и ремонта подвижного состава и инфраструктуры. Традиционные поисковые системы, основанные на точном совпадении ключевых слов, зачастую не справляются с семантической сложностью запросов, что обуславливает применение методов интеллектуального анализа текстов на основе современных подходов обработки естественного языка.

В настоящее время одним из наиболее распространенных подходов к построению текстовых СППР является технология генерации, дополненной поиском (Retrieval-Augmented Generation, RAG), впервые предложенная в работе [1]. Современные подходы к построению RAG-систем и направления их развития систематизированы в обзорной работе [2]. Применение RAG в СППР рассматривается в различных предметных областях, в частности в здравоохранении [3], сельском хозяйстве [4], электронной коммерции [5] и розничной торговле [6]. Эта технология обеспечивает сочетание механизмов семантического поиска по внешним базам знаний с генеративными способностями больших языковых моделей (БЯМ), что позволяет формировать аргументированные ответы на основе актуальных фактологических данных, а не только знаний, заложенных в веса языковой модели. Благодаря этой особенности RAG-системы становятся важным компонентом современных отраслевых СППР, обеспечивая баланс между гибкостью диалогового взаимодействия и достоверностью информации.

Однако практическая реализация RAG-систем в предметных областях со строгими требованиями к информационной безопасности и высокой степенью лингвистической специализации сталкивается с рядом вызовов, которые существенно ограничивают прямое применение стандартных решений. Среди основных проблем создания отраслевых текстовых СППР можно отметить следующие:

1. наличие различных категорий закрытой информации, которая не должна покидать контур организации, что ограничивает применение наиболее мощных моделей, доступных по API, обладающих как большим контекстным окном, позволяющим одновременно анализировать большой объем источников, так и достаточно развитыми языковыми возможностями;
2. наличие специфической отраслевой терминологии, которая может снижать эффективность работы моделей, предназначенных для широкого использования, опирающегося преимущественно на общую лексику.

Данные трудности оказываются взаимосвязаны – отраслевая терминология, вполне вероятно, более точно обрабатывалась бы передовыми БЯМ, однако необходимость предотвращения утечки данных ограничивает выбор теми инструментами, которые могут быть развернуты на собственной вычислительной инфраструктуре предприятия, которая может быть достаточно ограниченной. Таким образом, несмотря на рост эффективного контекстного окна передовых БЯМ¹ [7], технология RAG, центральным элементом которой является управление размером контекста, доступного БЯМ при формировании ответа, оказывается по-прежнему востребованной на практике.

¹Claude Blog. 1M Context Is Now Generally Available for Opus 4.6 and Sonnet 4.6 [Электронный ресурс]. URL: <https://claude.com/blog/1m-context-ga> (дата обращения: 13.07.2026).

Технология RAG предлагает множество вариантов практической реализации, различающихся методом выделения фрагментов («чанков») и отображения их в векторное пространство для семантического поиска, стратегией формирования и обработки поискового запроса, используемыми языковыми моделями [2]. Существенное внимание также уделяется разработке методов оценки качества RAG-систем, учитывающих как качество поиска релевантных документов, так и корректность формируемых ответов [8]. Несмотря на значительное количество исследований, посвященных развитию отдельных компонентов RAG, выбор эффективной конфигурации RAG для специализированных предметных областей остается открытой научно-практической задачей. В статье проводится практическое исследование ряда распространенных техник, используемых при реализации технологии RAG, на примере специализированного корпуса нормативно-технической документации в области железнодорожного транспорта. В рамках реализации данного исследования был сформирован корпус нормативно-технической документации в области железнодорожного транспорта и тестовое множество, состоящее из набора вопросов с ожидаемыми ответами, реализован базовый вариант технологии RAG, а также ряд распространенных стратегий обработки запросов. Кроме того, была проведена оценка перспективности дообучения модели отображения текста в векторное пространство низкой размерности (модели эмбединга текстовых фрагментов) для повышения точности извлечения фрагментов. Реализованные варианты оценены как по показателям, характеризующим качество извлечения релевантных фрагментов, так и по правильности формируемых ответов. Сформированный корпус, база фрагментов, а также программный код, воспроизводящий реализованные алгоритмы и результаты экспериментов, размещены в открытом доступе².

Статья организована следующим образом. В разделе 1 рассматриваются общая схема технологии RAG и исследуемые способы повышения качества поиска, включая стратегии обработки запросов, переранжирование с использованием кросс-энкодера и предметное дообучение моделей эмбединга. В разделе 2 описаны сформированный корпус нормативно-технической документации и наборы вопросов, использованные в экспериментах. В разделе 3 приведены методика экспериментального исследования и результаты сравнительной оценки рассмотренных подходов. В заключении сформулированы основные выводы работы.

1. Технология RAG

Общая схема технологии RAG изображена на рисунке 1. Весь комплекс операций, осуществляемых в рамках данной технологии, можно разделить на две фазы: фаза подготовки и фаза выполнения. На фазе подготовки производится идентификация корпуса документов, содержащих факты о предметной области, которые должны быть использованы при ответах на запросы пользователя. Документы разбиваются на фрагменты, которые обрабатываются моделью эмбединга текстовых фрагментов для отображения в семантическое векторное пространство таким образом, что близкие по смыслу документы соответствуют близким (по значению определенной меры –

²Репозиторий проекта rag-intrans-paper [Электронный ресурс]. URL: <https://gitverse.ru/avponomarev/rag-intrans-paper> (дата обращения: 28.08.2026).

например, косинусного сходства) векторам. Пары, состоящие из фрагмента и соответствующего ему вектора, помещаются в векторное хранилище, обеспечивающее возможность эффективного поиска по заданному вектору. На фазе выполнения поступивший от пользователя запрос также преобразуется в векторное представление и в векторном хранилище производится поиск фрагментов, обладающих схожими векторными представлениями. Исходный запрос пользователя и найденные в векторном хранилище фрагменты включаются в промпт к БЯМ, содержащий инструкции по формированию ответа на запрос на основе представленной информации.

При практической реализации этой общей схемы необходимо принять ряд инженерных решений, определяющих:

1. Стратегию разбиения корпуса на фрагменты. Эта стратегия является результатом поиска компромисса между богатством контекста (чем больше фрагмент, тем больше информации он содержит, что может быть полезно как для его семантической интерпретации, так и для построения ответа) и «сфокусированностью» фрагмента (облегчающей поиск конкретных фактов, релевантных запросу, и позволяющей сократить объем контекста БЯМ, формирующей итоговый ответ).
2. Модель эмбединга. Существует множество моделей, характеризующихся различным качеством и требованиями к вычислительным ресурсам.
3. Стратегию обработки запроса. В качестве поискового образа не обязательно должен выступать сам запрос, для повышения качества извлечения он может быть предварительно обработан (например, переформулирован в другом стиле).
4. БЯМ, осуществляющую формирование окончательного ответа.

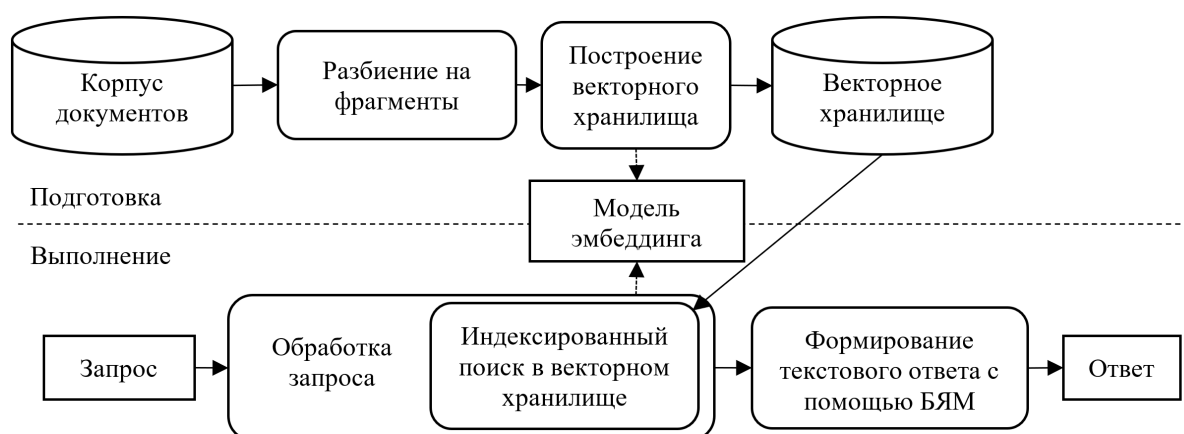


Рис. 1. Общая схема технологии RAG

В статье представлены результаты экспериментального исследования, нацеленного на сравнительную оценку эффективности решений, касающихся нескольких ключевых компонентов этой схемы, а именно:

- стратегии обработки запросов;
- стратегии извлечения результатов из векторного хранилища (одноэтапная или двухэтапная);

- модели эмбединга и возможности ее дообучения для учета предметной терминологии;
- БЯМ, используемой для формирования текстовых ответов.

1.1. Стратегии обработки запросов

Прямое использование исходного запроса пользователя в качестве поискового образа не всегда является оптимальным: запросы могут быть сформулированы нестрого, содержать избыточные детали или, напротив, быть слишком краткими для точного семантического сопоставления. Кроме того, сложные многосоставные вопросы могут требовать извлечения информации из нескольких источников, каждый из которых релевантен лишь отдельному аспекту запроса. В связи с этим в практике построения RAG-систем широкое распространение получили методы предварительной обработки запросов, направленные на трансформацию исходного пользовательского вопроса в форму, более удобную для семантического поиска. В настоящей работе исследован ряд таких стратегий, различающихся как по принципу действия, так и по вычислительной сложности.

HyDE (Hypothetical Document Embeddings) [9] – вместо прямого использования пользовательского запроса сначала с помощью большой языковой модели генерируется гипотетический документ-ответ на заданный вопрос, после чего этот синтезированный текст используется в качестве поискового запроса для извлечения из векторного хранилища. Предполагается, что гипотетический ответ, даже если он содержит неточности, лучше отражает семантическую структуру релевантных документов, чем исходный вопрос.

Step-Back (шаг назад) [10] – метод, основанный на идее абстрагирования. Стратегия заключается в том, чтобы сформулировать более общую версию исходного вопроса, сохраняющую его суть, но абстрагирующуюся от частных деталей. Затем выполняется поиск как по исходному, так и по обобщенному запросу, а их результаты объединяются.

Multi-Query (мультизапросы) [11] – стратегия, заключающаяся в генерации нескольких альтернативных формулировок одного и того же вопроса с помощью языковой модели. Каждая из сгенерированных формулировок используется для отдельного поиска, после чего результаты всех запросов объединяются. Данный подход призван компенсировать возможные неудачные формулировки исходного вопроса и повышать полноту извлечения.

Decomposition (декомпозиция) [12] – метод, предназначенный для обработки сложных многосоставных вопросов, ответ на которые требует информации из нескольких источников. В рамках этой стратегии исходный вопрос с помощью БЯМ разбивается на несколько более простых подвопросов, для каждого из которых выполняется независимый поиск, после чего найденные фрагменты объединяются. Такой подход особенно эффективен для многошаговых вопросов, релевантные факты для которых распределены по разным документам.

PRF (Pseudo-Relevance Feedback) [13] – сначала выполняется поиск по исходному запросу, из найденных документов извлекаются наиболее частотные ключевые слова (за исключением стоп-слов), после чего исходный запрос расширяется этими ключевыми словами и выполняется повторный

поиск. Данный подход не требует обращения к БЯМ и может быть реализован с минимальными вычислительными затратами.

PRF+LLM [14] – модификация предыдущего метода, в которой извлечение ключевых слов из найденных документов выполняется не на основе частотного анализа, а с помощью большой языковой модели. Такой подход, хотя и более вычислительно затратный, потенциально позволяет выделять семантически значимые термины, не являющиеся простыми повторяющимися словами.

Recursive Query Refinement (рекурсивная/многошаговая переформулировка) [15] – итеративная стратегия, при которой на каждом шаге на основе предыдущего запроса и найденных по нему документов языковая модель генерирует уточненную версию запроса. Результаты поиска со всех шагов объединяются. Данный подход позволяет последовательно сужать область поиска, двигаясь от общих формулировок к более точным, и особенно полезен в случаях, когда релевантная информация не находится с первой попытки.

1.2. Переранжирование фрагментов с использованием кросс-энкодера

Одним из наиболее распространенных способов повышения качества поиска в RAG-системах является использование двухэтапной схемы извлечения фрагментов [16]. На первом этапе выполняется быстрый поиск кандидатов с использованием модели эмбединга, после чего найденные фрагменты повторно ранжируются более точной, но вычислительно затратной моделью – кросс-энкодером, который получает на вход одновременно пару «запрос – фрагмент» и непосредственно оценивает степень их семантического соответствия. В рамках работы для повторного ранжирования использовалась открытая модель `qilowoq/bge-reranker-v2-m3-en-ru`³, представляющая собой русскоязычную адаптацию (усеченную по токенам до поддержки только русского и английского языков) многоязычной модели `BAAI/bge-reranker-v2-m3` [17].

Алгоритм поиска состоит из следующих этапов:

1. формирование эмбединга пользовательского запроса с использованием модели эмбединга;
2. поиск наиболее релевантных фрагментов в векторном хранилище;
3. отбор первых N фрагментов-кандидатов;
4. вычисление оценки релевантности для каждой пары «запрос – фрагмент» с помощью кросс-энкодера;
5. повторное ранжирование найденных фрагментов по полученным оценкам и выбор первых k результатов.

1.3. Дообучение модели эмбединга

Подход к обучению моделей для получения семантически значимых эмбедингов предложений представлен в работе [18], а методы обучения

³Здесь и далее указание названия модели машинного обучения в виде X/Y означает, что используется модель из репозитория HuggingFace, доступная по URL: <https://huggingface.co/X/Y>.

моделей эмбединга для задач информационного поиска рассмотрены в [19]. Несмотря на высокое качество современных моделей эмбединга общего назначения [20], они обучаются на текстовых корпусах широкой тематики и не учитывают специфику нормативно-технической документации железнодорожного транспорта. Вследствие этого семантически связанные запросы и фрагменты документов могут располагаться в пространстве эмбедингов недостаточно близко, что приводит к ухудшению качества поиска релевантного контекста.

Целью дообучения является адаптация модели эмбединга к предметной области железнодорожного транспорта таким образом, чтобы вопросы пользователей располагались в векторном пространстве ближе к соответствующим нормативным пунктам. Предполагается, что подобная адаптация позволит повысить качество компонента поиска без изменения архитектуры RAG-системы.

Особенностью предлагаемого подхода является использование для обучения только простых вопросов и вопросов средней сложности, каждый из которых в большинстве случаев имеет единственный релевантный нормативный пункт. При этом оценка качества модели производится на независимом тестовом множестве сложных вопросов, для ответа на которые требуется извлечение нескольких взаимосвязанных нормативных пунктов. Данный подход позволяет оценить способность модели к обобщению и определить, способствует ли дообучение на простых примерах улучшению поиска для более сложных пользовательских запросов.

Для исследования были выбраны три современные модели, поддерживающие русский язык и широко применяемые в задачах семантического поиска:

- **ai-forever/ru-en-RoSBERTa** [21] выбрана как относительно компактная двуязычная модель, обеспечивающая высокую скорость вычисления эмбедингов. Использование данной модели позволяет оценить, насколько эффективной может быть предметная адаптация модели небольшого размера и способен ли ее дообученный вариант конкурировать с более современными моделями эмбединга;
- **BAAI/bge-m3** [17] представляет собой одну из наиболее современных открытых моделей, специально разработанных для задач информационного поиска. Модель изначально оптимизирована для этой задачи и демонстрирует высокое качество поиска без дополнительного обучения, поэтому представляет интерес с точки зрения оценки потенциала ее дальнейшей предметной адаптации;
- **intfloat/multilingual-e5-large** [22] относится к числу современных моделей эмбединга и широко применяется при построении RAG-систем.

Дообучение выполнялось с использованием библиотеки SentenceTransformers. В качестве обучающей выборки использовались пары вида (q_i, d_i) , где q_i – вопрос пользователя, а d_i – соответствующий нормативный пункт, содержащий правильный ответ.

В рамках подхода к дообучению для каждого положительного образца остальные фрагменты текущего батча автоматически рассматриваются как отрицательные образцы. Таким образом, модель обучается увеличивать коси-

нусное сходство между вопросом и соответствующим нормативным пунктом, одновременно уменьшая сходство между вопросом и всеми остальными фрагментами батча.

Дообучение всех моделей выполнялось по одинаковой схеме. Использовались две эпохи обучения, размер батча 32, скорость обучения $2 \cdot 10^{-5}$, коэффициент затухания весов 0,01. Обучение осуществлялось на графическом процессоре RTX 3090.

2. Корпус документов и наборы вопросов

В рамках исследования был сформирован корпус нормативно-технических документов на железнодорожную тематику, который был использован тремя способами: для формирования тестового набора данных в ходе оценки различных вариантов реализации RAG, для наполнения векторной базы данных (компонент извлечения в рамках RAG осуществлял поиск непосредственно среди множества текстовых фрагментов данного корпуса) и для формирования набора данных при дообучении модели эмбединга. В корпус вошли основные публичные нормативно-технические документы железнодорожной отрасли [23]:

- Устав железнодорожного транспорта Российской Федерации⁴;
- Правила технической эксплуатации железных дорог Российской Федерации (ПТЭ)⁵;
- Инструкция по организации движения поездов и маневровой работы на железнодорожном транспорте Российской Федерации (Приложение 2 к ПТЭ);
- Приложения 1–20 к Инструкции по организации движения (содержащие текстовые данные, остальные приложения содержат преимущественно формы);
- Правила перевозок грузов, порожних грузовых вагонов железнодорожным транспортом⁶.

Актуальные на дату написания статьи версии документов были загружены с сайта РЖД по ссылкам, приведенным выше (для Инструкции по организации движения и приложений к ней ссылки не указаны, так как данные документы доступны по ссылке на ПТЭ) в форматах HTML и DOCX.

Следует отметить, что, несмотря на то что одним из важных мотивов, определяющих актуальность проводимого исследования, является наличие множества документов, предназначенных для служебного пользования (ДСП), в корпус сознательно включались только открытые документы. С одной стороны, это позволило сделать сам корпус и все его производные открытыми, что способствует обеспечению воспроизводимости исследования и облегчает его развитие, с другой стороны, основные проблемы (в частности,

⁴Устав железнодорожного транспорта Российской Федерации // Веб-портал РЖД [Электронный ресурс]. URL: <https://company.rzd.ru/ru/9353/page/105104?id=154> (дата обращения: 13.07.2026).

⁵Правила технической эксплуатации железных дорог Российской Федерации // Веб-портал РЖД [Электронный ресурс]. URL: <https://company.rzd.ru/ru/9353/page/105104?id=1827> (дата обращения: 13.07.2026).

⁶Правила перевозок грузов, порожних грузовых вагонов железнодорожным транспортом // Веб-портал РЖД [Электронный ресурс]. URL: <https://company.rzd.ru/ru/9353/page/105104?id=1758> (дата обращения: 13.07.2026).

специфическая отраслевая терминология) проявляются и на сформированном корпусе документов, поэтому авторы считают, что результаты, полученные с его помощью, могут быть легко обобщены.

Первым шагом обработки корпуса при применении технологии RAG является разбиение его на фрагменты («чанки»), которые сохраняются в векторном хранилище и в дальнейшем служат фактографической базой при формировании ответов на вопросы пользователя (см. раздел 1). Особенностью нормативной документации и, в частности, всех документов, включенных в корпус, является высокая степень структуризации – документы разбиты на разделы и пункты (или статьи – в зависимости от вида документа), каждый из которых посвящен определенной ситуации, сценарию или действию. Подобная структуризация, наряду с высокой тематической сфокусированностью текста пунктов и их относительно небольшим размером, позволяет непосредственно использовать их в качестве фрагментов. В некоторых случаях текст пунктов, в свою очередь, бывает разбит на абзацы – при обработке корпуса данное разбиение также было учтено и сохранено в структурированном представлении корпуса, это может быть полезно при проведении экспериментов с альтернативными стратегиями разбиения на фрагменты, однако во всех экспериментах, проведенных в статье, в качестве фрагмента использовался весь текст пункта (или статьи).

Для оценки различных вариантов реализации RAG был сформирован набор вопросов трех уровней сложности:

- Простые вопросы, относящиеся к одному факту, например, «Для чего предназначен базовый размер исчисления сборов и штрафов?».
- Вопросы средней сложности, относящиеся к нескольким фактам, расположенным в одном или нескольких абзацах нормативного пункта, например, «В чем различие операций, выполняемых в местах общего пользования и в местах необщего пользования?» или «По каким основаниям перевозчик имеет право отказать в согласовании заявки на перевозку груза?».
- Вопросы повышенной сложности, требующие извлечения нескольких взаимосвязанных нормативных положений, например, «Каковы обязательства перевозчика и заявителя переадресовки в случае изменения грузополучателя и необходимости информирования сторон о переадресовке, если перевозчик является владельцем инфраструктуры?».

Всего в тестовое множество входит 150 вопросов, из которых 37 простых, 13 средней сложности и 100 повышенной сложности. Для оценки дообученных моделей эмбединга использовались только 100 вопросов повышенной сложности, тогда как в остальных экспериментах использовалось полное тестовое множество из 150 вопросов.

Формирование вопросов повышенной сложности выполнялось автоматически с использованием модели GPT-4o-mini. Генерация проводилась не по отдельным пунктам, а по заранее сформированным группам тематически связанных фрагментов. Для каждого нормативного пункта вычислялось векторное представление с помощью модели *intfloat/multilingual-e5-large*, после чего для случайно выбранного опорного пункта находились ближайшие по косинусному сходству фрагменты. Из найденных кандидатов формировались

группы размером два или три пункта. В зависимости от конфигурации эксперимента применялись три режима отбора: пункты из одного раздела одного документа; пункты из одного документа, но необязательно одного раздела; пункты без ограничения по документу и разделу.

Для формирования вопросов, отражающих различные информационные потребности пользователей, использовались три пользовательские роли:

1. поездной диспетчер, рассматривающий эксплуатационные ситуации, условия, обязанности и порядок действий;
2. специалист по нормативному контролю, сопоставляющий общие правила, специальные условия, ограничения и исключения;
3. разработчик технологического процесса, анализирующий последовательность операций и распределение ответственности.

Для каждой роли задавался предпочтительный набор типов вопросов, включая вопросы вида «условие – действие», описание процедуры, сопоставление общего правила и исключения, агрегацию требований и восстановление причинно-следственной последовательности.

Для предметного дообучения моделей эмбединга был сформирован отдельный синтетический набор пар «вопрос – релевантный нормативный пункт». БЯМ (GPT-4o-mini) передавались сведения об источнике, версии документа, разделе и номере пункта, а также полный текст всех его абзацев. Вопросы формировались в двух категориях сложности: простые и средней сложности.

Генерация выполнялась с использованием структурированного вывода. Температура генерации была установлена равной 0,2, что обеспечивало достаточное разнообразие формулировок при сохранении устойчивости и воспроизводимости. В результате было сформировано 3159 вопросов, из которых 2071 относились к категории простых, 1088 – к категории вопросов средней сложности.

3. Экспериментальное исследование

3.1. Методика

Целью экспериментального исследования являлась оценка эффективности различных способов повышения точности RAG-системы, предназначенной для работы с нормативно-технической документацией железнодорожного транспорта. Экспериментальное исследование проводилось последовательно в несколько этапов.

На первом этапе выполнялось сравнение трех базовых моделей эмбединга (ai-forever/ru-en-RoSBERTa, BAAI/bge-m3 и intfloat/multilingual-e5-large) без дополнительной предметной адаптации.

На втором этапе исследовалось влияние переранжирования документов с использованием кросс-энкодера qilowoq/bge-reranker-v2-m3-en-ru. Для каждой модели эмбединга оценивалось качество поиска при различном количестве фрагментов-кандидатов, подлежащих повторному ранжированию.

На третьем этапе исследовалось влияние стратегий предварительной обработки пользовательских запросов на качество поиска.

На четвертом этапе выполнялось предметное дообучение моделей эмбединга на сформированном наборе простых вопросов и вопросов средней сложности. Исследовалось влияние объема обучающей выборки (500, 1000 и 3000 вопросов) на качество поиска. Оценка проводилась на независимом тестовом множестве вопросов повышенной сложности, что позволило определить способность модели переносить приобретенные знания на более сложные пользовательские запросы.

На заключительном этапе проводилась оценка качества генерации ответов БЯМ с использованием модели эмбединга, выбранной по результатам экспериментов по оценке качества поиска.

Для количественной оценки качества поиска использовались следующие показатели:

- **Recall@10** характеризует полноту поиска и показывает долю релевантных фрагментов, найденных среди первых десяти результатов;
- **MRR@10** (Mean Reciprocal Rank) характеризует качество ранжирования и определяется как среднее значение обратного ранга первого найденного релевантного фрагмента;
- **NDCG@10** (Normalized Discounted Cumulative Gain) характеризует качество ранжирования с учетом как полноты поиска, так и позиций всех релевантных фрагментов.

Для оценки качества генерации рассматривались две БЯМ одного семейства – Qwen3 14B и Qwen3 32B [24]. В обоих случаях генерация выполнялась без режима рассуждения (Thinking), а в качестве контекста языковой модели передавались десять наиболее релевантных фрагментов, найденных с использованием модели эмбединга multilingual-e5-large, дообученной на выборке из 3000 синтетических вопросов.

Качество генерации оценивалось с использованием библиотеки Ragas⁷ [8], реализующей подход LLM-as-a-Judge, в рамках которого в качестве модели-судьи использовалась БЯМ GPT-4o-mini. Оценка выполнялась автоматически по следующим показателям:

- **Достоверность ответа** (Faithfulness) – степень, в которой утверждения, содержащиеся в ответе, подтверждаются предоставленным контекстом;
- **Релевантность ответа** (Answer Relevancy) – степень соответствия сгенерированного ответа поставленному вопросу;
- **Семантическое сходство** (Semantic Similarity) – степень семантической близости сгенерированного ответа к эталонному ответу независимо от совпадения формулировок.

3.2. Результаты

⁷Официальный сайт библиотеки Ragas [Электронный ресурс]. URL: <https://www.ragas.io/> (дата обращения: 13.07.2026).

Базовый поиск с использованием моделей эмбединга

На первом этапе была проведена оценка качества семантического поиска с использованием трех моделей эмбединга без дополнительного предметного дообучения. Результаты представлены в таблице 1.

Таблица 1. Результаты оценки базовых моделей эмбединга

Модель эмбединга	Recall@10	MRR@10	NDCG@10
ru-en-RoSBERTa	0,640	0,582	0,543
multilingual-e5-large	0,712	0,672	0,608
bge-m3	0,724	0,638	0,597

Полученные результаты показывают, что модели эмбединга multilingual-e5-large и bge-m3 существенно превосходят ru-en-RoSBERTa по всем рассматриваемым показателям качества поиска. Наибольшее значение Recall@10 продемонстрировала модель bge-m3 (0,724), что свидетельствует о ее более высокой способности находить релевантные фрагменты. Вместе с тем по показателям MRR@10 (0,672) и NDCG@10 (0,608), характеризующим качество ранжирования поисковой выдачи, лучшие результаты показала multilingual-e5-large, что означает, что данная модель чаще располагает релевантные фрагменты на первых позициях поисковой выдачи.

Результаты использования кросс-энкодера

Для оценки влияния повторного ранжирования документов была реализована двухэтапная схема поиска. На первом этапе модель эмбединга извлекала заданное число фрагментов-кандидатов (20, 50 или 100), после чего они повторно ранжировались кросс-энкодером. Полученные результаты представлены на рисунке 2. Значение 0 на горизонтальной оси соответствует базовому варианту поиска без использования переранжирования.

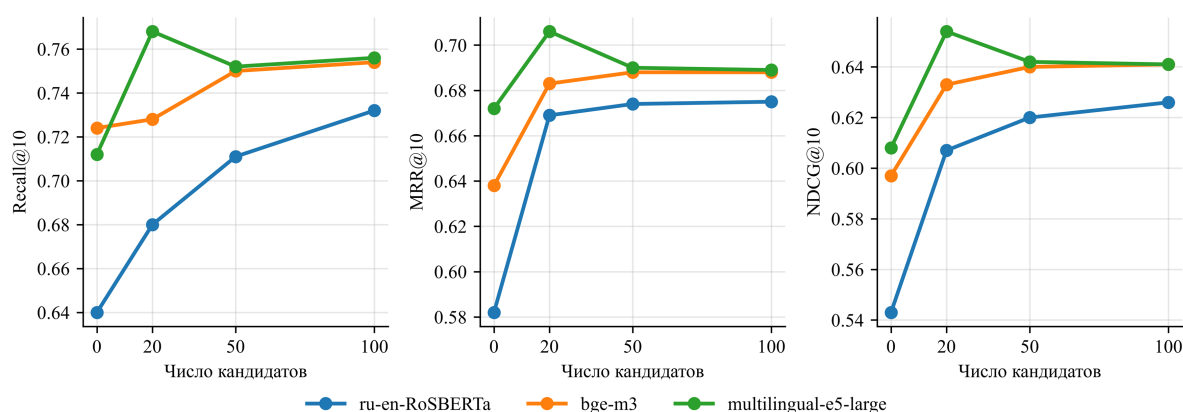


Рис. 2. Влияние переранжирования кросс-энкодером на качество семантического поиска

Полученные результаты показывают, что использование кросс-энкодера позволяет повысить качество поиска для всех рассмотренных моделей эмбединга, однако величина прироста существенно зависит от качества исходного ранжирования. Наибольший эффект наблюдается для модели ru-en-RoSBERTa. По сравнению с базовым вариантом значение Recall@10

увеличилось с 0,640 до 0,732, MRR@10 – с 0,582 до 0,675, а NDCG@10 – с 0,543 до 0,626. При этом увеличение числа фрагментов-кандидатов с 20 до 100 приводит к последовательному улучшению всех рассматриваемых показателей.

Для модели bge-m3 также наблюдается положительное влияние увеличения числа документов-кандидатов. При переходе от 20 к 100 кандидатам значение Recall@10 возрастает с 0,728 до 0,754, а показатели MRR@10 и NDCG@10 демонстрируют умеренный, но устойчивый рост.

Иная картина наблюдается для multilingual-e5-large. Наилучшие результаты достигаются уже при повторном ранжировании первых 20 фрагментов-кандидатов (Recall@10 = 0,768, MRR@10 = 0,706, NDCG@10 = 0,654). Дальнейшее увеличение числа кандидатов до 50 и 100 приводит к небольшому снижению всех трех показателей. Вероятно, это связано с тем, что исходное ранжирование модели уже обладает высоким качеством, а включение менее релевантных документов увеличивает количество потенциальных ошибок повторного ранжирования.

Таким образом, эффективность использования кросс-энкодера определяется не только его собственной способностью различать релевантные фрагменты, но и качеством первоначальной поисковой выдачи. Для моделей с менее точным ранжированием (ru-en-RoSBERTa) увеличение числа кандидатов обеспечивает существенный прирост качества, тогда как для более сильной модели эмбединга (multilingual-e5-large) оптимальным оказывается повторное ранжирование сравнительно небольшого числа фрагментов.

Результаты применения различных стратегий обработки запроса

Результаты применения различных стратегий обработки запроса приведены в таблице 2. В данном эксперименте использовалась модель эмбединга multilingual-e5-large без переранжирования кросс-энкодером, а для формулировки альтернативных (или дополнительных) версий запроса – БЯМ GPT OSS 20B. Вопреки ожиданиям авторов статьи, продвинутые методы обработки запросов не привели к заметному повышению качества поиска релевантных фрагментов. Большинство методов показали примерно тот же уровень качества извлечения, что и базовая стратегия без дополнительной обработки. По результатам эксперимента наиболее перспективной выглядит стратегия StepBack, однако ее преимущество не является значительным, применение же стратегии HyDE привело к существенному падению качества результатов.

Таблица 2. Результаты использования различных стратегий обработки запроса

Стратегия обработки	Recall@10	MRR@10	NDCG@10
Без доп. обработки	0,712	0,672	0,608
HyDE	0,599	0,558	0,499
StepBack	0,711	0,688	0,611
MultiQuery	0,692	0,642	0,577
Decomposition	0,723	0,603	0,571
PRF+LLM	0,700	0,641	0,590
PRF без LLM	0,704	0,648	0,599
Recursive	0,682	0,659	0,590

Результаты дообучения моделей эмбединга

Результаты дообучения моделей эмбединга представлены на рисунке 3. На графиках показано изменение значений Recall@10, MRR@10 и NDCG@10 в зависимости от размера обучающей выборки.

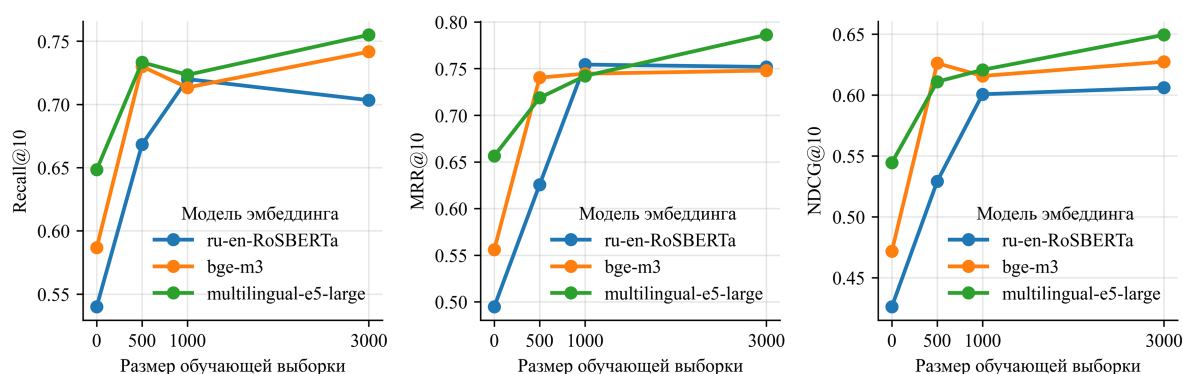


Рис. 3. Влияние дообучения моделей эмбединга на качество семантического поиска

Дообучение приводит к существенному повышению качества поиска для всех исследуемых моделей. При этом наибольший прирост наблюдается уже при использовании сравнительно небольших обучающих выборок, тогда как дальнейшее увеличение их объема обеспечивает значительно менее выраженный эффект.

Наиболее заметное влияние предметная адаптация оказала на модель ai-forever/ru-en-RoSBERTa. По сравнению с базовым вариантом значение Recall@10 увеличилось с 0,540 до 0,720 при обучении на 1000 вопросах, MRR@10 – с 0,495 до 0,754, а NDCG@10 – с 0,426 до 0,601. Дальнейшее увеличение размера обучающей выборки до 3000 вопросов не привело к существенному улучшению качества. Вероятно, это связано с тем, что для относительно компактной модели потенциал предметной адаптации оказался практически исчерпан уже при обучении на 1000 вопросах. Для модели BAAI/bge-m3 также наблюдается устойчивое улучшение всех исследуемых показателей.

Наиболее стабильный рост качества демонстрирует модель multilingual-e5-large. При обучении на 3000 вопросах модель достигает лучших результатов среди всех рассмотренных вариантов: Recall@10 составляет 0,755, MRR@10 – 0,786, а NDCG@10 – 0,649.

В целом, результаты подтверждают эффективность предложенного подхода к предметному дообучению моделей эмбединга. При этом наиболее важным является тот факт, что обучение выполнялось исключительно на относительно простых вопросах, тогда как оценка производилась на независимом наборе вопросов повышенной сложности, что говорит о способности моделей переносить знания, полученные на относительно простых примерах, на более сложные сценарии поиска, требующие извлечения нескольких взаимосвязанных фрагментов.

Результаты оценки качества генерации ответов

Полученные результаты представлены в таблице 3. В таблице приведены средние значения показателей и соответствующие им стандартные отклонения.

Таблица 3. Результаты оценки качества генерации ответов БЯМ

БЯМ	Faithfulness	Answer Relevancy	Semantic Similarity
Qwen3 14B	0,839 ± 0,216	0,913 ± 0,100	0,939 ± 0,020
Qwen3 32B	0,833 ± 0,202	0,896 ± 0,136	0,931 ± 0,021
<i>p</i> -значение	0,667	0,078	< 0,001

Модель Qwen3 14B продемонстрировала несколько более высокие значения по всем рассматриваемым показателям по сравнению с Qwen3 32B. Для оценки статистической значимости различий между моделями использовался парный критерий Уилкоксона. Статистически значимых различий по показателям Faithfulness ($p = 0,667$) и Answer Relevancy ($p = 0,078$) не выявлено. Для показателя Semantic Similarity различие оказалось статистически значимым ($p < 0,001$), однако его абсолютная величина невелика (разница составляет 0,008), поэтому данное различие не имеет существенной практической значимости.

Заключение

В работе проведен сравнительный анализ методов повышения точности RAG-систем на корпусе нормативно-технической документации в области железнодорожного транспорта. Исследованы три группы методов: стратегии обработки запросов, двухэтапное извлечение с использованием кросс-энкодера и дообучение моделей эмбединга.

Установлено, что современные модели эмбединга существенно превосходят более компактные аналоги по всем показателям качества поиска. Применение кросс-энкодера для повторного ранжирования фрагментов позволяет повысить качество поиска, однако величина эффекта определяется качеством исходного ранжирования: для моделей с менее точной выдачей прирост значителен, тогда как для сильных моделей оптимальным является ранжирование небольшого числа кандидатов.

Стратегии предварительной обработки запросов не обеспечили значимого повышения качества извлечения по сравнению с базовым подходом, а в ряде случаев привели к ухудшению результатов.

Наибольшую эффективность среди всех исследованных методов показало предметное дообучение моделей эмбединга на синтетических парах «вопрос – нормативный пункт». Обучение на относительно простых вопросах обеспечило устойчивое улучшение поиска по сложным запросам, что свидетельствует о способности моделей к обобщению. Сравнение языковых моделей разного размера не выявило существенного преимущества более крупной модели при формировании ответов на основе предоставленного контекста.

Таким образом, в условиях ограниченной вычислительной инфраструктуры и необходимости работы с закрытыми документами, содержащими

специфическую отраслевую терминологию, наиболее практичным инструментом повышения точности RAG-систем является предметное дообучение модели эмбединга. Данный подход позволяет существенно повысить качество поиска без дополнительных вычислительных затрат на этапе выполнения и без риска утечки данных. Предложенная методология может быть обобщена на другие предметные области со схожей структурой нормативно-технической документации.

Список литературы

- [1] Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks / P. Lewis, E. Perez, A. Piktus [et al.] // *Advances in Neural Information Processing Systems*. – 2020. – Vol. 33. – P. 9459–9474.
- [2] Retrieval-Augmented Generation for Large Language Models: A Survey / Y. Gao, Y. Xiong, X. Gao [et al.] // *arXiv*. – 2023. – arXiv:2312.10997. – DOI 10.48550/arXiv.2312.10997.
- [3] A Review of Large Language Models in Medical Education, Clinical Decision Support, and Healthcare Administration / J. Vrdoljak, Z. Boban, M. Vilović [et al.] // *Healthcare*. – 2025. – Vol. 13, № 6. – Art. 603. – DOI 10.3390/healthcare13060603.
- [4] A Comprehensive Survey of Retrieval-Augmented Large Language Models for Decision Making in Agriculture: Unsolved Problems and Research Opportunities / A. Vizniuk, G. Diachenko, I. Laktionov [et al.] // *Journal of Artificial Intelligence and Soft Computing Research*. – 2025. – Vol. 15, № 2. – P. 115–146. – DOI 10.2478/jaiscr-2025-0007.
- [5] Ершов, М. А. Интеллектуальная СППР для оценки конкурентоспособности в электронной коммерции: методология интеграции RAG-системы / М. А. Ершов // *Экономика строительства*. – 2025. – № 11. – С. 518–521.
- [6] Самофалов, Г. С. Интеграция больших языковых моделей (LLM) и RAG в контуры поддержки принятия решений (DSS) для задач ценообразования и управления спросом в розничной торговле / Г. С. Самофалов // *Интернаука*. – 2026. – № 16–1. – С. 38–42.
- [7] Cioplea, R. Memory in the Age of AI: How LLMs Remember, Forget, and Leak : Technical Report, ver. 3.0. – 2026. – 37 p. – DOI 10.17605/OSF.IO/9WYG3.
- [8] RAGAS: Automated Evaluation of Retrieval-Augmented Generation / S. Es, J. James, L. E. Anke, S. Schockaert // *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. – 2024. – P. 150–158. – DOI 10.18653/v1/2024.eacl-demo.16.
- [9] Precise Zero-Shot Dense Retrieval without Relevance Labels / L. Gao, X. Ma, J. Lin, J. Callan // *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. – 2023. – P. 1762–1777. – DOI 10.18653/v1/2023.acl-long.99.
- [10] Take a Step Back: Evoking Reasoning via Abstraction in Large Language Models / H. S. Zheng, S. Mishra, X. Chen [et al.] // *International Conference on Learning Representations (ICLR 2024)*. – 2024.
- [11] DMQR-RAG: Diverse Multi-Query Rewriting for RAG / Z. Li [et al.] // *arXiv*. – 2024. – arXiv:2411.13154. – DOI 10.48550/arXiv.2411.13154.
- [12] Ammann, P. J. L. Question Decomposition for Retrieval-Augmented Generation / P. J. L. Ammann, J. Golde, A. Akbik // *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*. – 2025. – P. 497–507. – DOI 10.18653/v1/2025.acl-srw.32.
- [13] Salton, G. Improving Retrieval Performance by Relevance Feedback / G. Salton, C. Buckley // *Journal of the American Society for Information Science*. – 1990. – Vol. 41, № 4. – P. 288–297.
- [14] Mackie, I. Generative relevance feedback with large language models / I. Mackie, S. Chatterjee, J. Dalton // *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. – 2023. – P. 2026–2031.

- [15] Wang, Y. MaFeRw: Query Rewriting with Multi-Aspect Feedbacks for Retrieval-Augmented Large Language Models / Y. Wang // arXiv. – 2024. – arXiv:2408.17072. – DOI 10.48550/arXiv.2408.17072.
- [16] Nogueira, R. Passage Re-Ranking with BERT / R. Nogueira, K. Cho // arXiv. – 2019. – arXiv:1901.04085. – DOI 10.48550/arXiv.1901.04085.
- [17] M3-Embedding: Multi-Linguality, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation / J. Chen [et al.] // arXiv. – 2024. – arXiv:2402.03216. – DOI 10.48550/arXiv.2402.03216.
- [18] Reimers, N. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks / N. Reimers, I. Gurevych // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). – 2019. – P. 3982–3992. – DOI 10.18653/v1/D19-1410.
- [19] Text Embeddings by Weakly-Supervised Contrastive Pre-Training / L. Wang, N. Yang, X. Huang [et al.] // arXiv. – 2022. – arXiv:2212.03533. – DOI 10.48550/arXiv.2212.03533.
- [20] MTEB: Massive Text Embedding Benchmark / N. Muennighoff, N. Tazi, L. Magne, N. Reimers // arXiv. – 2022. – arXiv:2210.07316. – DOI 10.48550/arXiv.2210.07316.
- [21] The Russian-focused embedders' exploration: ruMTEB benchmark and Russian embedding model design / A. Snegirev [et al.] // arXiv. – 2024. – arXiv:2408.12503. – DOI 10.48550/arXiv.2408.12503.
- [22] Multilingual E5 Text Embeddings: A Technical Report / L. Wang [et al.] // arXiv. – 2024. – arXiv:2402.05672. – DOI 10.48550/arXiv.2402.05672.
- [23] Паршин, К. А. Исследование применения признаков терминов железнодорожной отрасли при формировании классификатора / К. А. Паршин, М. С. Подгорный // Теория и практика современной науки. – 2019. – № 3. – С. 229–232.
- [24] Qwen3 Technical Report / A. Yang [et al.] // arXiv. – 2025. – arXiv:2505.09388. – DOI 10.48550/arXiv.2505.09388.

COMPARATIVE ANALYSIS OF METHODS FOR IMPROVING RAG SYSTEM ACCURACY ON SPECIALIZED CORPORA OF REGULATORY AND TECHNICAL DOCUMENTATION

Agafonov A.A.^{1, a}, Ponomarev A.V.^{1, b}

¹ St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), St. Petersburg, Russia.

E-mail: ^a agafonov.a@spcras.ru, ^b ponomarev@iiias.spb.su

Abstract. The digital transformation of industry drives a growing demand for intelligent decision support systems capable of efficiently extracting knowledge from regulatory and technical documentation. This paper examines Retrieval-Augmented Generation (RAG) technology applied to a specialized corpus of railway regulatory documents. A comparative analysis is conducted for several methods aimed at improving RAG accuracy, including query processing strategies (HyDE, Step-Back, Multi-Query, Decomposition, Pseudo-Relevance Feedback, Recursive Refinement), two-stage retrieval with a cross-encoder, and fine-tuning of embedding models on synthetic question-regulatory clause pairs. The experiments demonstrate that domain-specific

For citation:

Agafonov A.A., Ponomarev A.V. Comparative Analysis of Methods for Improving RAG System Accuracy on Specialized Corpora of Regulatory and Technical Documentation. *Intelligent transport*, 2026, no. 3(39), pp. 4–22. DOI: 10.24412/3033-6007-2026-339-4-22. (in Russian)

fine-tuning of embedding models yields the most significant improvement in retrieval quality. Cross-encoder reranking also provides a positive effect, particularly for models with initially less accurate ranking. At the same time, advanced query processing strategies do not lead to a substantial improvement in retrieval quality compared to the baseline approach. The constructed document corpus, question sets, and source code are made publicly available.

Keywords: RAG, large language models, semantic search, embedding models, cross-encoder, regulatory and technical documentation, railway transport, fine-tuning.

References

- [1] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474.
- [2] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., et al. (2023). *Retrieval-augmented generation for large language models: A survey* (arXiv:2312.10997). arXiv. <https://doi.org/10.48550/arXiv.2312.10997>
- [3] Vrdoljak, J., Boban, Z., Vilović, M., Kumrić, M., & Božić, J. (2025). A review of large language models in medical education, clinical decision support, and healthcare administration. *Healthcare*, 13(6), Article 603. <https://doi.org/10.3390/healthcare13060603>
- [4] Vizniuk, A., Diachenko, G., Laktionov, I., Siwocha, A., Xiao, M., & Smola, J. (2025). A comprehensive survey of retrieval-augmented large language models for decision making in agriculture: Unsolved problems and research opportunities. *Journal of Artificial Intelligence and Soft Computing Research*, 15(2), 115–146. <https://doi.org/10.2478/jaiscr-2025-0007>
- [5] Ershov, M. A. (2025). Intelligent DSS for competitiveness assessment in e-commerce: Methodology of RAG-system integration. *Ekonomika stroitelstva*, (11), 518–521. (in Russian)
- [6] Samofalov, G. S. (2026). Integration of large language models (LLM) and RAG into decision support systems (DSS) for pricing and demand management in retail. *Internauka*, (16-1), 38–42. (in Russian)
- [7] Cioplea, R. (2026). *Memory in the age of AI: How LLMs remember, forget, and leak* (Technical Report ver. 3.0). <https://doi.org/10.17605/OSF.IO/9WXG3>
- [8] Es, S., James, J., Anke, L. E., & Schockaert, S. (2024). RAGAS: Automated evaluation of retrieval-augmented generation. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 150–158). <https://doi.org/10.18653/v1/2024.eacl-demo.16>
- [9] Gao, L., Ma, X., Lin, J., & Callan, J. (2023). Precise zero-shot dense retrieval without relevance labels. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1762–1777). <https://doi.org/10.18653/v1/2023.acl-long.99>
- [10] Zheng, H. S., Mishra, S., Chen, X., Cheng, H.-T., Chi, E. H., Le, Q. V., & Zhou, D. (2024). Take a step back: Evoking reasoning via abstraction in large language models. In *International Conference on Learning Representations (ICLR 2024)*.
- [11] Li, Z., et al. (2024). *DMQR-RAG: Diverse multi-query rewriting for RAG* (arXiv:2411.13154). arXiv. <https://doi.org/10.48550/arXiv.2411.13154>
- [12] Ammann, P. J. L., Golde, J., & Akbik, A. (2025). Question decomposition for retrieval-augmented generation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)* (pp. 497–507). <https://doi.org/10.18653/v1/2025.acl-srw.32>

- [13] Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41(4), 288–297.
- [14] Mackie, I., Chatterjee, S., & Dalton, J. (2023). Generative relevance feedback with large language models. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2026–2031).
- [15] Wang, Y. (2024). *MaFeRw: Query rewriting with multi-aspect feedbacks for retrieval-augmented large language models* (arXiv:2408.17072). arXiv. <https://doi.org/10.48550/arXiv.2408.17072>
- [16] Nogueira, R., & Cho, K. (2019). *Passage re-ranking with BERT* (arXiv:1901.04085). arXiv. <https://doi.org/10.48550/arXiv.1901.04085>
- [17] Chen, J., et al. (2024). *M3-Embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation* (arXiv:2402.03216). arXiv. <https://doi.org/10.48550/arXiv.2402.03216>
- [18] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982–3992). <https://doi.org/10.18653/v1/D19-1410>
- [19] Wang, L., Yang, N., Huang, X., Jiao, B., Yang, L., Jiang, D., et al. (2022). *Text embeddings by weakly-supervised contrastive pre-training* (arXiv:2212.03533). arXiv. <https://doi.org/10.48550/arXiv.2212.03533>
- [20] Muennighoff, N., Tazi, N., Magne, L., & Reimers, N. (2022). *MTEB: Massive text embedding benchmark* (arXiv:2210.07316). arXiv. <https://doi.org/10.48550/arXiv.2210.07316>
- [21] Snegirev, A., et al. (2024). *The Russian-focused embedders' exploration: ruMTEB benchmark and Russian embedding model design* (arXiv:2408.12503). arXiv. <https://doi.org/10.48550/arXiv.2408.12503>
- [22] Wang, L., et al. (2024). *Multilingual E5 text embeddings: A technical report* (arXiv:2402.05672). arXiv. <https://doi.org/10.48550/arXiv.2402.05672>
- [23] Parshin, K. A., & Podgornyy, M. S. (2019). A study of the application of railway industry term features in the formation of a classifier. *Teoriya i praktika sovremennoy nauki*, (3), 229–232. (in Russian)
- [24] Yang, A., et al. (2025). *Qwen3 technical report* (arXiv:2505.09388). arXiv. <https://doi.org/10.48550/arXiv.2505.09388>

НЕЙРОСЕТЕВЫЕ МЕТОДЫ РАСПОЗНАВАНИЯ ПОДЗЕМНЫХ ОБЪЕКТОВ В ДАННЫХ ГЕОРАДАРА МОБИЛЬНОГО КОЛЕСНОГО РОБОТА

Кривошей М.И.^{1,а}, Юдин Д.А.^{1,2,б}, Зарипов М.Н.^{1,с}, Буличев О.В.^{1,д},
Муравьев К.Ф.^{1,3,е}, Панов А.И.^{1,2,ф}

¹ Московский физико-технический институт (МФТИ), Москва, Россия.

² Институт искусственного интеллекта AIRI, Москва, Россия.

³ Федеральный исследовательский центр «Информатика и управление» РАН (ФИЦ ИУ РАН), Москва, Россия.

E-mail: ^а krivoshei.mi@mipt.ru, ^б yudin.da@mipt.ru, ^с zaripov.mn@mipt.ru, ^д obulichev@yandex.ru,

^е kirill.mouraviev@yandex.ru, ^ф panov.ai@mipt.ru

Аннотация. В настоящее время для решения задач, связанных с поиском и распознаванием объектов под землёй, часто используется георадар. Это позволяет получать богатую информацию об исследуемой местности на различной глубине, однако ручная интерпретация радарограмм связана с большими трудозатратами. В данной работе предлагается метод GeoRTt для автоматического нахождения подземных объектов интереса на основе нейронной сети оригинальной архитектуры с механизмом внимания GeoAttention, на вход которой подаются А-сканы георадара, установленного на колёсном роботе, перемещающемся в автоматическом либо полуавтоматическом режиме. Предлагаемая нейросетевая модель позволяет учитывать пространственные взаимосвязи между А-сканами георадара даже в тех случаях, когда траектория движения робота отличается от прямой линии. С использованием мобильного колёсного робота и георадара, антенна которого установлена на небольшой высоте над анализируемой поверхностью, собран и размечен уникальный набор данных для распознавания объектов нескольких категорий, находящихся под поверхностью земли. Эксперименты с разработанной нейросетевой моделью на собранном наборе данных продемонстрировали F1-меру классификации свыше 40%, что превосходит по качеству методы-аналоги. В ходе сравнительного анализа различных вариаций архитектуры модели выделены две конфигурации для разных условий решения задачи георадарного исследования.

Ключевые слова: *мобильный колёсный робот, А-сканы, георадар, нейронная сеть, трансформер, распознавание подземных объектов.*

Введение

Для задач, связанных с анализом и распознаванием подземных объектов как природного, так и искусственного происхождения, использование георадара предпочтительно именно потому, что данный тип сенсора даёт исследователю информацию о подповерхностном пространстве куда более богатую, чем металлоискатель и магнитометр, а в ряде случаев (детектирование объектов, документация на которые отсутствует или утеряна, суждения

Для цитирования:

Кривошей М.И., Юдин Д.А., Зарипов М.Н., Буличев О.В., Муравьев К.Ф., Панов А.И. Нейросетевые методы распознавания подземных объектов в данных георадара мобильного колесного робота. *Интеллектуальный транспорт*. 2026. № 3(39). С. 23–36. DOI: 10.24412/3033-6007-2026-339-23-36.

об опорных и др. физических свойствах грунта [1] на основе отраженного сигнала и т.д.) георадар незаменим. Соответственно, актуальность задачи создания надежной и универсальной технологии анализа радарограмм очевидна. На сегодняшний день автоматический анализ радарограмм георадара сопряжён с рядом трудностей. В частности, существующие модели распознавания подземных объектов склонны к ложным срабатываниям, когда в исследуемой области находится много металлического мусора. С другой стороны, если объект содержит мало металлических компонентов или не содержит их вовсе, модель может его не обнаружить. Неизбежно возникающие при движении робота по пересечённой местности отклонения от прямого участка траектории и качания трансмиссии искажают радарограммы [2], что особенно критично для методов, анализирующих радарограммы в формате В-сканов – последовательностей А-сканов по ходу движения робота [3].

Предлагаемая технология позволяет как находить с высокой точностью объекты разной степени локальности (точечные, конечного размера, протяженные с линейным размером, превосходящим размеры исследуемой площади), так и компенсировать возможную неравномерность и криволинейность движения робота.

Для решения данных проблем предлагается трансформерная архитектура GeoRTr, которая на первом этапе, с целью придания модели быстродействия и уменьшения потребления видеопамати, разбивает исследуемую поверхность на независимые секторы, а на втором – применяет к каждому из секторов классифицирующую модель на основе трансформерного энкодера, кодирующего каждый А-скан совместно с его локальными координатами. Использование механизма внимания позволяет решить проблему методов детекции объектов на двумерных изображениях на основе архитектуры YOLO [4], неспособных информативно кодировать пространственное положение каждого А-скана [2]. В то же время, разбиение на секторы решает проблемы быстродействия и потребления видеопамати, актуальные для трансформерных моделей [3], а использование архитектуры с небольшим числом параметров решает проблему переобучения модели при малом размере обучающей выборки.

Основной вклад нашей статьи состоит в следующем:

- предлагается оригинальная архитектура нейросетевого классификатора А-сканов на основе механизма внимания, названная GeoRTr, учитывающая пространственные взаимосвязи А-сканов, разбиение исследуемого поля на секторы и содержащая относительно малое число параметров, что позволяет бороться с переобучением при малом размере обучающей выборки;
- с использованием мобильного колёсного робота и георадара собран и размечен уникальный набор данных для распознавания объектов нескольких категорий, находящихся под поверхностью земли;
- эксперименты с разработанной нейросетевой моделью на собранном наборе данных продемонстрировали F1-меру классификации выше 40%, что превосходит по качеству методы-аналоги. В ходе сравнительного анализа различных вариаций архитектуры модели выделены две конфигурации для разных условий решения задачи георадарного исследования.

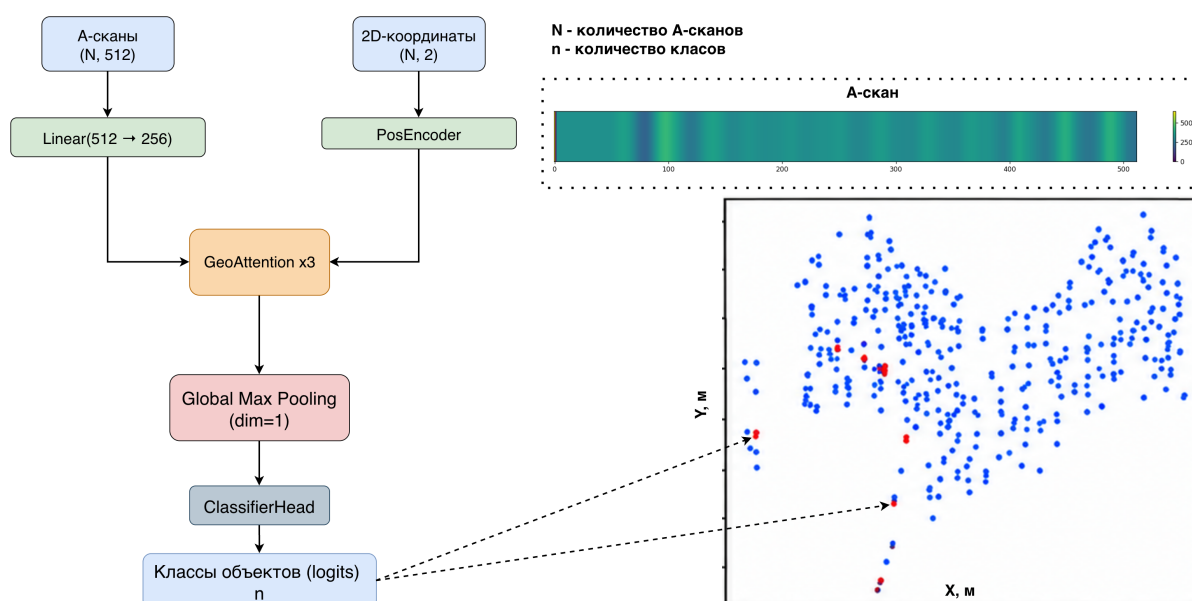


Рис. 1. Архитектура предложенного метода GeoRTg для классификации А-сканов, полученных от георадара

1. Анализ предметной области

1.1. Анализ методов обнаружения подземных объектов на В-сканах георадара

Для обнаружения подземных объектов зачастую используются модели детекции на базе YOLO, в частности модифицированные архитектуры, такие как HFL-YOLOv8 [5], TLAD-YOLO [6], ME-YOLO [7], FLS-YOLO [8]. Для повышения устойчивости модели к шуму и чувствительности к геометрическим особенностям радарограмм применяются специальные методы предобработки данных – такие как топологический анализ данных TDA [9]. Предварительное выделение на изображении структурных признаков (форма гипербол, замкнутые контуры) повышает точность работы YOLOv5 на В-сканах. На примере YOLOv5 показано, что при использовании TDA $mAP@0,5$ увеличивается с 0,975 до 0,993, $mAP@0,5:0,95$ – с 0,757 до 0,801.

Ещё одним способом адаптации YOLO для детекции объектов на В-сканах является замена стандартных свёрточных слоёв на специальные – такие как GhostConv [10] и C3Ghost [10]. Основная идея этих модификаций в том, чтобы сохранить максимальное разнообразие карт признаков при малом числе параметров модели и высоком быстродействии. Вместо того, чтобы использовать стандартную свёртку с m фильтрами, GhostConv применяет стандартную свёртку лишь с $n < m$ фильтрами, а остальные $m - n$ фильтров представлены простыми линейными преобразованиями. Экспериментально показано, что использование GhostConv в базовой модели YOLOv8n снижает число параметров с 3,6М до 1,9М и GFLOP с 8,3 до 4,9, а mAP на одном из наборов данных увеличивается с 0,715 до 0,723.

Метод MCGA-Net [11] для детекции дефектов автомобильных дорог представляет собой архитектуру YOLOv8, адаптированную для обработки данных многоантенных георадаров. Модель последовательно обрабатывает тензор из нескольких В-сканов с использованием произведения Эйнштейна, а затем детектирует дефекты модулем на базе YOLOv8. На собранном авторами

наборе данных модель показала метрики: Precision = 0,928, Recall = 0,94, mAP@50 = 0,967.

Для обнаружения подземных объектов также используются специальные легковесные свёрточные архитектуры [12]. Один из методов основан на архитектуре VGG16 с добавлением слоя Global Average Pooling для уменьшения числа параметров и борьбы с переобучением. Решается задача нахождения подземных полостей путём детекции максимумов гипербола на В-сканах. На реальных данных модель показала точность обнаружения подземных полостей, равную 1,0.

В модели GPR-TSBiNet [13] для обнаружения подземных заземляющих линий используется трансформерная архитектура GPR-Trans в комбинации с модифицированной SC-BiFPN и функцией потерь Shape-IoU. На реальных данных модель имеет метрику APsmall = 0,494.

Существует ряд методов, таких как MHInvNet [1] и LeakInv-CUNet [14] на базе U-Net и Pix2pixGAN [15] на базе генеративно-состязательной модели, решающих смежную с распознаванием подземных объектов задачу – построение карты диэлектрической проницаемости подповерхностного пространства.

1.2. Анализ методов обнаружения подземных объектов с учётом 2D-координат

Метод GPR-Former [2] использует для обнаружения объектов ванильный трансформер, что позволяет учитывать взаимное расположение координат А-сканов. Авторы рассмотрели три варианта позиционного кодирования: синусоидальное, rotary, линейный слой, и лучшие метрики модель выдала на синусоидальном. Все А-сканы исследуемой поверхности разделяются на группы с помощью K-D дерева. Наибольшие метрики модель показала при размере группы, равном 1024: AR = 0,4834, F2-score = 0,7277.

Существует подход, использующий специальную рекуррентную архитектуру Mamba [3] для обнаружения объектов. На первом этапе 6-канальный пространственный кодировщик на основе ResNet-18 кодирует каждые 2 соседних В-скана в один тензор, аналогичный по размерности исходным В-сканам, который затем подаётся в Mamba как последовательность А-сканов. На собранном авторами датасете метрики составили: RMSE = 0,06 м, R^2 = 0,94.

Методы, реализующие обнаружение подземных объектов с учётом 2D-координат, имеют более высокую обобщающую способность в сравнении с опирающимися на В-сканы (которые, как правило, основаны на обученном с нуля YOLO). Однако такие методы обладают зачастую более сложной и ресурсоёмкой трансформерной архитектурой, что может быть не оправдано в тех случаях, когда объекты интереса сильно контрастируют на радарограммах.

2. Создание обучающего набора георадарных данных, содержащего подземные объекты интереса

Для создания обучающей и валидационной выборок использовался георадар, способный синхронизировать свои данные с точным геодезическим приемником сантиметровой точности, то есть была применена следующая схема:

Неподвижный геодезический навигационный приемник, играющий роль

«базы», то есть источника дифференциальных поправок к навигационным данным, передаваемых по беспроводному каналу на навигационный приемник-ровер», георадар с закрепленным на нем навигационным приемником-ровером, прикрепленный к роботу, и собственно робот, совершающий развертку исследуемой площадки так, чтобы вблизи каждой точки площадки георадар прошел хотя бы раз. Удаление ровера от базы не превышало 500 м. На выходе георадара получаем амплитуду отраженного импульса как функцию времени, прошедшего с момента излучения импульса, а также точные (сантиметровая точность) координаты георадара на момент излучения импульса. Производится примерно 100 сканирований в секунду, для каждого сканирования измеряется интенсивность отраженного сигнала в 500–2000 равноотстоящих друг от друга моментах времени (количество выставлялось в настройках георадара в зависимости от структуры площадки), здесь чем более поздний момент – тем более длинному пути сигнала он соответствует, и, соответственно, может быть вычислена глубина, на которой он был отражен.

В качестве носителя использовалась модифицированная колесная платформа AgileX Scout 2. На ней были установлены георадар GSSI SIR 3000 с антенной 400 МГц, геодезический приемник Javad Maxor, фронтальная RGB-D камера Intel RealSense D455i и средства колесной одометрии. Антенна георадара закреплялась перед платформой и перемещалась на небольшой высоте над поверхностью. Колесное шасси обеспечивало плавное движение и устойчивое положение антенны на неровном грунте, что уменьшало искажения радарограмм, связанные с изменением скорости и вертикальными колебаниями сенсора.

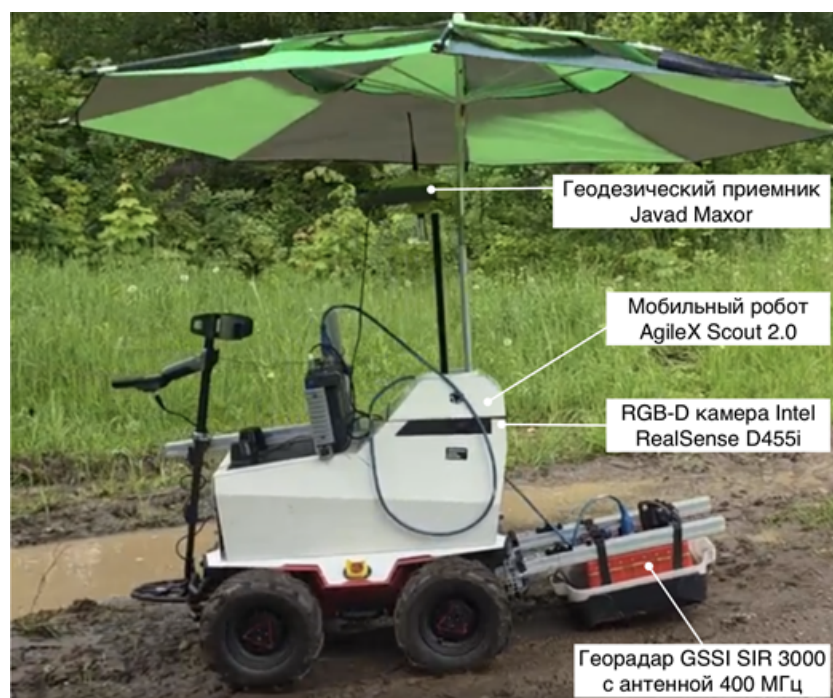


Рис. 2. Мобильный робототехнический комплекс с георадаром GSSI SIR 3000 и антенной 400 МГц

Сканирование выполнялось по системе параллельных профилей с шагом около 0,3 м. Данные записывались преимущественно при движении в одном направлении, чтобы уменьшить влияние направления движения и взаимного

положения антенны и платформы на форму сигнала. После завершения очередного профиля оператор переводил комплекс на соседнюю линию, а прямолинейный участок робот проходил в полуавтоматическом режиме.

Управление прямолинейным движением осуществлялось по данным колесной одометрии. Для ограничения накапливаемой ошибки использовалась визуальная коррекция по контрастному ориентиру, установленному в конце профиля и поднятому над уровнем растительности. RGB-D камера определяла направление и расстояние до ориентира.

Для определения объекта с камеры использовалась дообученная YOLOv8n. При его обнаружении оценка положения уточнялась, после чего робот останавливался на заданном безопасном расстоянии. Такая схема обеспечивала воспроизводимое прохождение профилей без использования спутниковых координат в контуре управления.



Рис. 3. Полуавтоматическое движение робота с визуальной коррекцией по ориентиру в конце профиля

Навигация робота и пространственная привязка измерений решали разные задачи. Система одометрии и камера использовались для управления движением, тогда как приемник-ровер получал дифференциальные поправки от неподвижной базы и формировал координаты георадарных трасс сантиметровой точности. Записи георадара и навигационного приемника снабжались временными метками и синхронизировались при последующей обработке. В результате каждый А-скан сопоставлялся с координатами на площадке независимо от способа управления платформой.

Данные георадара характерны тем, что они несут отпечаток не только одного непосредственно выделенного объекта, но и всех объектов, оказавшихся на пути электромагнитного импульса. Так, импульс, отраженный от сильно заглубленного объекта, несет в себе не только информацию о данном

объекте, но и дифракционные изменения от всех более близких к георадару локальных объектов на его пути, что позволяет судить об их форме и материале.

Затем, точно зная подземное строение площадки, сканирование которой производилось, полученный материал, то есть файлы с выдачей георадара, привязанные к точному времени и геодезическим координатам с навигационного приемника, нарезали на фрагменты с маркировкой соответствующей им подземной структуры и создавали обучающую выборку. Для создания выборки использовалась грунтовая поверхность с невысокой травой; почва умеренно влажная.

Исследуемая поверхность представляет собой поле размером 80×60 метров со средней диэлектрической проницаемостью подповерхностного пространства $\varepsilon = 8$, на котором закопаны разнородные археологические объекты на глубине в несколько десятков сантиметров. Всего закопано 13 объектов в разных частях поля.

Для быстрого действия модели и меньших затрат видеопамати, а также для решения проблемы небольшого размера обучающей выборки, исследуемая поверхность размера 80×60 метров разбивается на сектора 30×30 сантиметров – считается, что в каждом из таких секторов А-сканы имеют сильную пространственную взаимосвязь. В каждом из секторов находится не более 5000 А-сканов. Набор данных разбивается на обучающую и валидационную выборки в пропорции 8:2 путём сэмплирования А-сканов из всей исследуемой поверхности.

3. Предлагаемый подход GeoRTr для распознавания подземных объектов в данных георадара мобильного колесного робота

3.1. Общая постановка задачи

А-скан георадара представляет собой 512-мерный вектор d , каждая компонента которого представляет собой амплитуду на определённом временном отсчёте:

$$d = (d_1, \dots, d_{512}). \quad (1)$$

Входной вектор модели GeoRTr представляет собой сектор D размера 30×30 сантиметров со всеми находящимися в нём А-сканами:

$$D = \{(d_i, x_i, y_i)\}_{x_D-0,15 \leq x_i \leq x_D+0,15}^{y_D-0,15 \leq y_i \leq y_D+0,15}, \quad (2)$$

где (x_D, y_D) – координаты центра сектора.

Для модели f_θ задача классификации формулируется следующим образом:

$$f_\theta : \mathbb{R}^{514 \times T_i} \rightarrow [0, 1]^{T_i}, \quad (3)$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left[-\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^{T_i} (y_{i,k} \log f_\theta(d_{i,k}) + (1 - y_{i,k}) \log (1 - f_\theta(d_{i,k}))) + \lambda R(\theta) \right], \quad (4)$$

где T_i – число А-сканов в секторе i , N – число секторов, $d_{i,k}$ – k -й по счёту

А-скан в i -м секторе, $y_{i,k} \in \{0, 1\}$ – метка его класса, $R(\theta)$ – регуляризация с коэффициентом λ .

В случае большой погрешности определения координат объектов интереса в обучающей выборке, модель может обучаться классифицировать не каждый А-скан, а секторы целиком:

$$f_{\theta} : \mathbb{R}^{514 \times T_i} \rightarrow [0, 1], \quad (5)$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \left[-\frac{1}{N} \sum_{i=1}^N (y_i \log f_{\theta}(D_i) + (1 - y_i) \log (1 - f_{\theta}(D_i))) + \lambda R(\theta) \right], \quad (6)$$

где D_i – i -й по счёту сектор.

3.2. Подход к решению

Антенна георадара имеет частоту 400 МГц. На глубине в несколько десятков сантиметров расположены подземные объекты интереса: они имеют различные размеры и материалы. Объекты могут состоять из керамики, железа, других материалов. Требуется для каждого А-скана определить, находится ли в его зоне видимости (используемая модель георадара может обнаруживать объекты на глубине 40–60 сантиметров в радиусе 30 сантиметров) какой-либо объект. Каждый А-скан может принадлежать к одному из двух классов: «объект» или «не объект».

Модель представляет собой оригинальную трансформерную архитектуру, применяемую независимо к каждому из секторов. 512-мерные сканы георадара и 2D-координаты кодируются в единое 256-мерное признаковое пространство, размерность которого сохраняется до применения модуля классификации. Архитектура включает в себя 3 оригинальных слоя внимания GeoAttention, Global Max Pooling для независимой агрегации признаков для каждой глубины (условно, каждой глубине соответствует определённое время приёма отражённого сигнала в наносекундах) и двухслойный модуль классификации. Модель содержит 2,5М параметров.

3.3. Кодирование входных векторов

А-сканы и 2D-координаты приводятся в единое 256-мерное признаковое пространство. Для кодирования А-сканов используется линейный слой, для кодирования координат – двухслойная полносвязная нейронная сеть с функцией активации ReLU. Позиционный эмбединг в неизменном виде подаётся на вход каждого из слоёв внимания.

3.4. Механизм внимания

Используется оригинальный механизм внимания GeoAttention. На каждом слое он принимает на вход 2 вектора: эмбединг предыдущего слоя или модуля и позиционный эмбединг. Это позволяет сохранять на всех слоях точную информацию о 2D-координатах каждого А-скана. Поступающие на вход векторы суммируются и подаются четырёхголовому стандартному механизму самовнимания (self-attention), выход которого обрабатывается

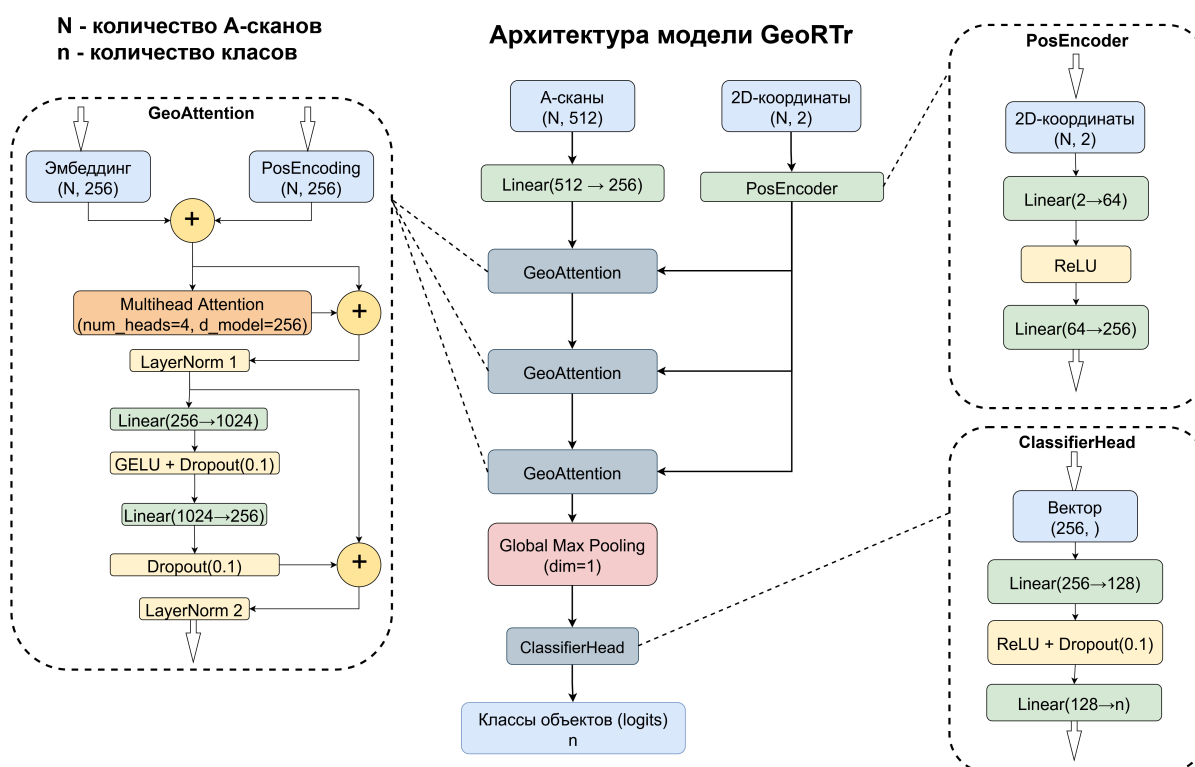


Рис. 4. Укрупненная схема предлагаемого подхода GeoRT обнаружения подземных объектов при помощи классификации А-сканов. Он отличается оригинальным механизмом внимания GeoAttention, сохраняющим на всех слоях точную информацию о взаимном расположении А-сканов и имеющим несколько голов внимания

двухслойной нейронной сетью с функцией активации GELU, dropout и нормализацией.

3.5. Количественная оценка качества предлагаемого подхода

Для оценки качества модели основными метриками являются precision, recall, F1-score. F1-score используется как общая оценка точности метода, несмещённая относительно баланса классов, Recall (полнота) – метрика наиболее важная для задач, где имеет критическое значение нахождение как можно большего числа объектов при допущении ложных обнаружений, Precision (точность) – наибольшее значение имеет, если ложные обнаружения должны быть сведены к минимуму.

4. Результаты экспериментов

Эксперименты с предложенным подходом поиска подземных объектов GeoRT осуществлялись на собранном наборе данных. Тестировались различные варианты архитектуры метода – с различным числом слоёв GeoAttention с 4 головами внимания.

Первый вариант метода с 4 головами внимания и 2 слоями энкодера отличается не очень высокими метриками. Модель с 4 головами и 4 слоями отличается тем, что не имеет на валидационной выборке ложных срабатываний ($\text{precision} = 1,0$) при низкой полноте нахождения объектов ($\text{F1-score} = 0,29$, $\text{recall} = 0,17$). Модель с 4 головами и 3 слоями показывает

наиболее высокое значение $F1\text{-score} = 0,44$ ($\text{precision} = 0,67$), хотя имеет не самую высокую полноту обнаружения ($\text{recall} = 0,33$). Для сравнения, многослойный перцептрон имеет самое низкое значение $F1\text{-score} = 0,19$, а универсальный и достаточно эффективный для задач классификации метод градиентного бустинга CatBoost имеет метрики, схожие с неоптимальными вариантами архитектуры GeoRTr.

Таблица 1. Метрики качества разработанного метода GeoRTr при использовании разного числа голов и слоёв в сравнении с другими существующими методами

Модель	Precision	Recall	F1-score
GeoRTr 2 слоя, 4 головы	0,50	0,17	0,25
GeoRTr 3 слоя, 4 головы	0,67	0,33	0,44
GeoRTr 4 слоя, 4 головы	1,00	0,17	0,29
MLP	0,11	1,00	0,19
CatBoost	1,00	0,17	0,29

На рисунке 5 показаны результаты инференса различных моделей на смешанных данных из тестовой, валидационной и части обучающей выборок. Видно, что многослойный перцептрон MLP склонен давать ложноположительные срабатывания. Остальные модели имеют схожую картину предсказания, но имеют разную точность работы при объезде скоплений объектов.

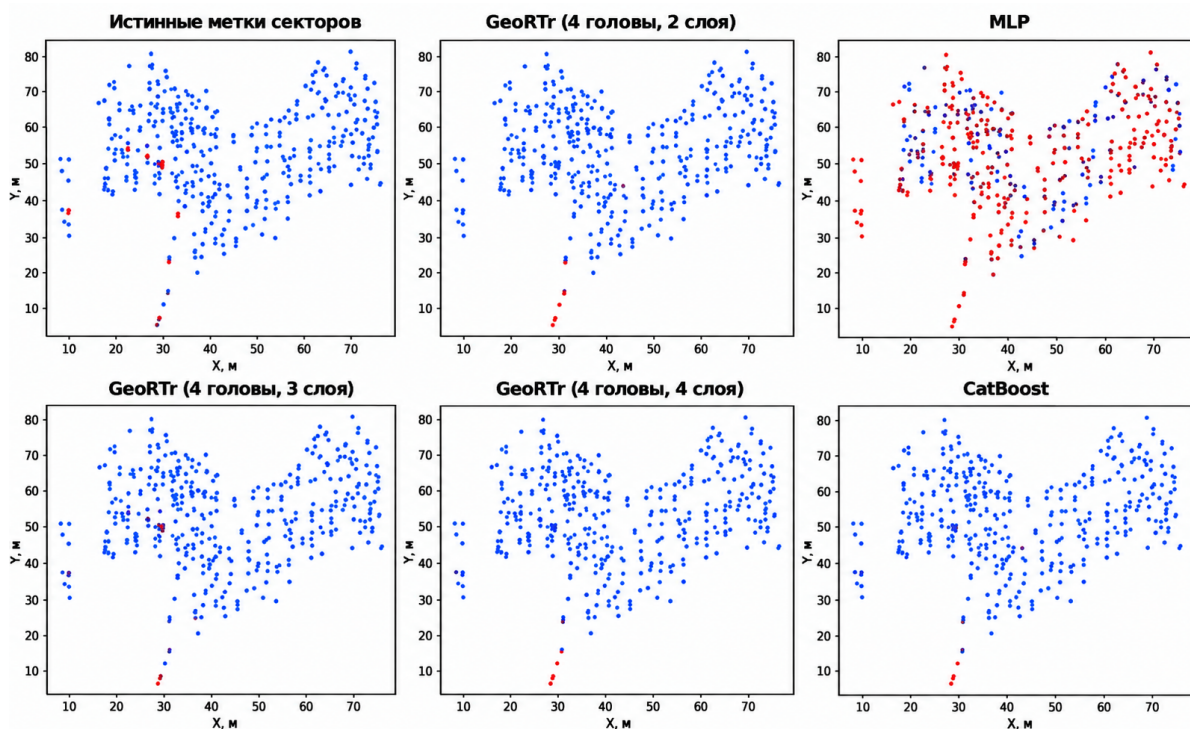


Рис. 5. Визуализация работы метода GeoRTr с различным числом слоёв GeoAttention, а также многослойного перцептрона и CatBoost. Красные точки – объект есть, синие – объекта нет. Каждая точка – А-скан валидационной выборки

Заключение

В настоящей статье рассмотрен нейросетевой вариант решения задачи распознавания объектов под землёй в данных георадара мобильного колёсного робота, основанный на применении специального механизма внимания GeoAttention к каждому из секторов, на которые разбивается исследуемая поверхность. Разработанный подход демонстрирует высокую точность работы.

Отметим, что программная реализация была испытана на пересеченной местности с сильными перепадами высоты, что вносило возмущения в прямолинейность и равномерность движения робота, а также были непроходимые участки, которые робот объезжал. Это показывает устойчивость предлагаемой технологии по отношению к неблагоприятному рельефу участка, а также к неравномерному и непрямолинейному движению робота.

В дальнейшем планируется собрать больше разнообразных данных для обучения модели, доработать архитектуру метода, добавив кодировщик вертикальных колебаний георадара и оценок диэлектрической проницаемости исследуемой поверхности. Это позволит как достичь более высокой точности модели в пределах одной исследуемой поверхности, так и сделать модель более устойчивой к изменению типа исследуемого грунта и неровностям поверхности. Также планируется добавить возможность поиска глубоко расположенных объектов с оценкой глубины их залегания, а также возможность детектирования нескольких объектов одинаковой или разной природы, расположенных непосредственно один над другим на разной глубине, которая также будет вычисляться моделью.

Благодарности

Работа частично выполнена в рамках НИР «Разработка метода одновременной локализации и картирования TerraCognitaSLAM на основе георадара и визуально-инерциальных сенсоров для условий недоступного спутникового сигнала» (по Договору между Центральным университетом и МФТИ).

Авторы выражают благодарность Центру робототехники Сбера за предоставление робота AgileX Scout 2.0 для проведения экспериментов, а также Шишкову Д. Л. за предоставление георадарного оборудования для сбора набора данных, описанного в настоящей статье.

Список литературы

- [1] Wu, M. Two-stage GPR image inversion method based on multi-scale dilated convolution and hybrid attention gate / M. Wu, Q. Liu, S. Ouyang // Remote Sensing. – 2025. – Vol. 17. – No. 2. – Art. 322. – DOI 10.3390/rs17020322.
- [2] Lee, K. GPR-Former: Context-Aware Moisture Detection / K. Lee, C. Feng // ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. – Montreal : IAARC Publications, 2025. – Vol. 42. – P. 1–8. – DOI 10.22260/isarc2025/0002.
- [3] Jeon, M. Efficient and Robust Temporal Modeling for LGPR Odometry With Mamba / M. Jeon, H. Lee, J. Lee // IEEE Access. – 2026. – Vol. 14. – P. 38775–38792. – DOI 10.1109/ACCESS.2026.3672677.
- [4] Automated railway subgrade defect detection using GPR and an enhanced YOLOv11 framework / M. T. Tegegne, R. Yang, S. Huang [et al.] // Transportation Geotechnics. – 2026. – Vol. 62. – Art. 102105. – DOI 10.1016/j.trgeo.2026.102105.

- [5] Lei, W. HFL-YOLOv8: a hyperbolic feature-enhanced lightweight network for object detection in ground penetrating radar images / W. Lei, S. Yu, T. Zhang // *Applied Soft Computing*. – 2026. – Vol. 188. – Art. 114403. – DOI 10.1016/j.asoc.2025.114403.
- [6] TLAD-YOLO: Lightweight network for intelligent detection of railway tunnel lining anomalies using ground penetrating radar / P. Zheng, A. Zhang, Z. Shi [et al.] // *Journal of Applied Geophysics*. – 2025. – Vol. 241. – Art. 105869. – DOI 10.1016/j.jappgeo.2025.105869.
- [7] ME-YOLO: A novel real-time detection network for pavement interlayer distress using ground-penetrating radar / S. Cao, C. Lu, X. Wang [et al.] // *Journal of Applied Geophysics*. – 2026. – Vol. 245. – Art. 106057. – DOI 10.1016/j.jappgeo.2025.106057.
- [8] FLS-YOLO: a multi-scale subsurface defect detection network via heterogeneous grouping and spatial reconstruction / W. Zhang, Z. Liu, Q. Yang, X. Qiu // *Signal, Image and Video Processing*. – 2026. – Vol. 20. – No. 5. – Art. 262. – DOI 10.1007/s11760-026-05352-z.
- [9] A Novel Shape-Aware Topological Representation for GPR Data with DNN Integration / M. Kang, S. Kaji, S.-Y. Lee [et al.] // *arXiv*. – 2025. – arXiv:2506.06311. – DOI 10.48550/arXiv.2506.06311.
- [10] Arvas, M. M. Detection of underground objects from GPR data using a lightweight YOLO-based approach / M. M. Arvas, A. Çınar // *Scientific Reports*. – 2026. – DOI 10.1038/s41598-026-59135-0.
- [11] Intelligent recognition of GPR road hidden defect images based on feature fusion and attention mechanism / H. Lv, Y. Zhang, J. Dai [et al.] // *IEEE Transactions on Geoscience and Remote Sensing*. – 2025. – Vol. 63. – P. 1–17.
- [12] Lightweight CNN model for automatic detection and depth estimation of subsurface voids using GPR B-scan data / A. Mojahid, D. El Ouai, K. El Amraoui [et al.] // *Natural Hazards Research*. – 2025. – Vol. 5. – No. 2. – P. 432–446. – DOI 10.1016/j.nhres.2025.02.001.
- [13] GPR-TSBiNet: An information gradient enrichment model for GPR B-scan small target detection / C. Wang, Y. Guan, M. Chi [et al.] // *Sensors*. – 2025. – Vol. 25. – No. 7. – Art. 2223. – DOI 10.3390/s25072223.
- [14] Attention-Module-Guided Time-Lapse Leakage Plume Imaging Driven by LeakInv-CUNet GPR Inversion Framework / H. Wang [et al.] // *IEEE Access*. – 2025. – Vol. 13. – P. 122514–122529.
- [15] Enhanced Pix2pixGAN with Spatial-Channel Attention for Underground Medium Inversion from GPR / S. Yang, L. Guo, Y. Yang, H. Ye // *Remote Sensing*. – 2026. – Vol. 18. – No. 3. – Art. 448. – DOI 10.3390/rs18030448.

NEURAL NETWORK METHODS FOR RECOGNIZING UNDERGROUND OBJECTS IN THE DATA OF A MOBILE WHEELED ROBOT'S GROUND-PENETRATING RADAR

**Krivoshey M. I.^{1, a}, Yudin D. A.^{1, 2, b}, Zaripov M. N.^{1, c}, Bulichev O. V.^{1, d},
Muravyev K. F.^{1, 3, e}, Panov A. I.^{1, 2, f}**

¹ Moscow Institute of Physics and Technology (MIPT), Moscow, Russia.

² Artificial Intelligence Research Institute (AIRI), Moscow, Russia.

³ Federal Research Center «Computer Science and Control» of the RAS (FRC CSC RAS), Moscow, Russia.

E-mail: ^a krivoshei.mi@mipt.ru, ^b yudin.da@mipt.ru, ^c zaripov.mn@mipt.ru, ^d obulichev@yandex.ru,

^e kirill.mouraviev@yandex.ru, ^f panov.ai@mipt.ru

For citation:

Krivoshey M. I., Yudin D. A., Zaripov M. N., Bulichev O. V., Muravyev K. F., Panov A. I. Neural Network Methods for Recognizing Underground Objects in the Data of a Mobile Wheeled Robot's Ground-Penetrating Radar. *Intelligent transport*, 2026, no. 3(39), pp. 23–36. DOI: 10.24412/3033-6007-2026-339-23-36. (in Russian)

Abstract. Currently, ground-penetrating radar (GPR) is often used to solve problems related to searching for and recognizing underground objects. This allows for obtaining rich information about the studied area at various depths, but manual interpretation of radargrams is labor-intensive. This paper proposes the GeoRTr method for automatically detecting underground objects of interest. It is based on an original neural network architecture with a GeoAttention mechanism. The network's input is A-scans from a GPR mounted on a wheeled robot moving in automatic or semi-automatic mode. The proposed neural network model accounts for spatial relationships between A-scans even when the robot's trajectory deviates from a straight line. Using a mobile wheeled robot and a GPR with an antenna mounted at a low altitude above the surface being analyzed, a unique dataset was collected and labeled for recognizing several categories of subsurface objects. Experiments with the developed neural network model on the collected dataset demonstrated a classification F1-score of over 40%, surpassing comparable methods. A comparative analysis of various model architecture variations identified two configurations for different ground-penetrating radar object search conditions.

Keywords: *mobile wheeled robot, A-scans, ground-penetrating radar, neural network, transformer, underground object recognition.*

References

- [1] Wu, M., Liu, Q., & Ouyang, S. (2025). Two-stage GPR image inversion method based on multi-scale dilated convolution and hybrid attention gate. *Remote Sensing*, 17(2), Article 322. <https://doi.org/10.3390/rs17020322>
- [2] Lee, K., & Feng, C. (2025). GPR-Former: Context-aware moisture detection. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction* (pp. 1–8). IAARC Publications. <https://doi.org/10.22260/isarc2025/0002>
- [3] Jeon, M., Lee, H., & Lee, J. (2026). Efficient and robust temporal modeling for LGPR odometry with Mamba. *IEEE Access*, 14, 38775–38792. <https://doi.org/10.1109/ACCESS.2026.3672677>
- [4] Tegegne, M. T., Yang, R., Huang, S., Xiao, H., Yang, Z., Zhang, Q., Liu, Y., & Qi, W. (2026). Automated railway subgrade defect detection using GPR and an enhanced YOLOv11 framework. *Transportation Geotechnics*, 62, Article 102105. <https://doi.org/10.1016/j.trgeo.2026.102105>
- [5] Lei, W., Yu, S., & Zhang, T. (2026). HFL-YOLOv8: A hyperbolic feature-enhanced lightweight network for object detection in ground penetrating radar images. *Applied Soft Computing*, 188, Article 114403. <https://doi.org/10.1016/j.asoc.2025.114403>
- [6] Zheng, P., Zhang, A., Shi, Z., Wang, S., Ma, Y., & Liu, Z. (2025). TLAD-YOLO: Lightweight network for intelligent detection of railway tunnel lining anomalies using ground penetrating radar. *Journal of Applied Geophysics*, 241, Article 105869. <https://doi.org/10.1016/j.jappgeo.2025.105869>
- [7] Cao, S., Lu, C., Wang, X., Zhang, P., Jin, G., & Cai, W. (2026). ME-YOLO: A novel real-time detection network for pavement interlayer distress using ground-penetrating radar. *Journal of Applied Geophysics*, 245, Article 106057. <https://doi.org/10.1016/j.jappgeo.2025.106057>
- [8] Zhang, W., Liu, Z., Yang, Q., & Qiu, X. (2026). FLS-YOLO: A multi-scale subsurface defect detection network via heterogeneous grouping and spatial reconstruction. *Signal, Image and Video Processing*, 20(5), Article 262. <https://doi.org/10.1007/s11760-026-05352-z>

- [9] Kang, M., Kaji, S., Lee, S.-Y., Kim, T., Ryu, H.-H., & Choi, S. (2025). *A novel shape-aware topological representation for GPR data with DNN integration* (arXiv:2506.06311). arXiv. <https://doi.org/10.48550/arXiv.2506.06311>
- [10] Arvas, M. M., & Çınar, A. (2026). Detection of underground objects from GPR data using a lightweight YOLO-based approach. *Scientific Reports*. <https://doi.org/10.1038/s41598-026-59135-0>
- [11] Lv, H., Zhang, Y., Dai, J., Wu, H., Wang, J., & Wang, D. (2025). Intelligent recognition of GPR road hidden defect images based on feature fusion and attention mechanism. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–17.
- [12] Mojahid, A., El Ouai, D., El Amraoui, K., El-Hami, K., Aitbenamer, H., Verrelst, J., & Barone, P. M. (2025). Lightweight CNN model for automatic detection and depth estimation of subsurface voids using GPR B-scan data. *Natural Hazards Research*, 5(2), 432–446. <https://doi.org/10.1016/j.nhres.2025.02.001>
- [13] Wang, C., Guan, Y., Chi, M., Shen, F., Yu, Z., Chen, Q., & Chen, C. (2025). GPR-TSBiNet: An information gradient enrichment model for GPR B-scan small target detection. *Sensors*, 25(7), Article 2223. <https://doi.org/10.3390/s25072223>
- [14] Wang, H., et al. (2025). Attention-module-guided time-lapse leakage plume imaging driven by LeakInv-CUNet GPR inversion framework. *IEEE Access*, 13, 122514–122529.
- [15] Yang, S., Guo, L., Yang, Y., & Ye, H. (2026). Enhanced Pix2pixGAN with spatial-channel attention for underground medium inversion from GPR. *Remote Sensing*, 18(3), Article 448. <https://doi.org/10.3390/rs18030448>

УДК 004.896

DOI: 10.24412/3033-6007-2026-339-37-52

ПОСТРОЕНИЕ КОМПАКТНЫХ ГРАФОВ СЦЕНЫ НА ОСНОВЕ ТОПОЛОГИЧЕСКОЙ КАРТЫ ДЛЯ АВТОНОМНОЙ НАВИГАЦИИ МОБИЛЬНОГО РОБОТА

Муравьев К.Ф.^{1, а}, Романенко В.И.^{2, b}

¹ Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), Москва, Россия.

² Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), Москва, Россия.

E-mail: ^а kmuravev@frccsc.ru, ^b viktorsamr@yandex.ru

Аннотация. Автономная навигация мобильного робота в человеко-ориентированных средах требует наличия карты, которая является не только геометрической моделью среды для планирования маршрутов, но также содержит информацию об объектах среды (двери, мебель, офисная техника и т.д.). Такое представление карты обеспечивают графы сцены, в которых вершинами являются помещения, локации и объекты, а ребра кодируют пространственные связности или отношения между объектами. Большинство современных методов построения графов сцены обладают высокой вычислительной сложностью, и построенные ими графы являются избыточными для обеспечения автономной навигации робота. В данной работе предлагается метод построения компактных графов сцены (Compact Scene Graph, CSG), в основе которого лежит вычислительно эффективный метод топологического картирования PRISM-ТороМар и привязка семантических объектов к локациям топологической карты. Построенный граф сцены позволяет строить маршруты до объектов по локациям топологической карты и осуществлять локализацию с высокой точностью. Предложенный метод был экспериментально исследован в симуляционной среде Habitat. Результаты экспериментов показали, что предложенный CSG потребляет значительно меньше памяти, чем традиционные метрические карты и графы сцен, и обеспечивает надежную локализацию по топологической карте за счет привязки к семантическим объектам.

Ключевые слова: *граф сцены, топологическая карта, семантическая сегментация, автономная навигация, симуляционная среда.*

Введение

Автономная навигация является одним из ключевых аспектов автономного функционирования мобильного робота. Способность самостоятельно перемещаться в незнакомой или динамически изменяющейся среде без участия оператора определяет практическую ценность робототехнических систем в самых разных областях – от сервисной робототехники до промышленной автоматизации. Для реализации автономной навигации роботу требуется карта окружающего пространства, которая служит основой для планирования маршрутов, локализации и принятия решений о движении.

Для цитирования:

Муравьев К.Ф., Романенко В.И. Построение компактных графов сцены на основе топологической карты для автономной навигации мобильного робота. *Интеллектуальный транспорт*. 2026. № 3(39). С. 37–52. DOI: 10.24412/3033-6007-2026-339-37-52.

В зависимости от способа представления среды различают несколько типов карт. Метрические карты, такие как сетки занятости или трёхмерные облака точек, обеспечивают высокую геометрическую точность и детализацию, однако потребляют значительный объём памяти и подвержены накоплению ошибок по мере движения робота. Примерами наиболее распространенных метрических подходов к картированию являются RTAB-Map [1] и GLIM [2]. Такие методы удобны тем, что строят плотное геометрическое представление среды и комплексировать данные с разных сенсоров робота, однако из-за высокого потребления памяти и накопления ошибки они пригодны лишь для сред небольшой площади, что подтверждается экспериментами, проведенными в работах [3] и [4].

Альтернативой выступают топологические карты, которые представляют среду в виде графа, где вершины соответствуют локациям среды, а рёбра – пространственной связности между ними. Такие карты компактны и эффективны для планирования маршрутов, однако зачастую лишены структурной детализации и семантической информации, что снижает их различительную способность и затрудняет локализацию в визуально монотонных средах. К числу топологических методов относятся PRISM-ТороМар [5], строящий компактный граф локаций на основе распознавания мест и сопоставления сканов, а также TSGM [6], интегрирующий семантическую информацию в топологическое представление для навигации по целевому изображению. Метод PRISM-ТороМар строит компактные карты сцен большой площади, пригодные для долгосрочной навигации, однако без добавления семантической информации навигация возможна лишь по заданным точным координатам внутри локации. Метод TSGM позволяет задавать цель навигации в виде изображения, однако не предназначен для долгосрочной навигации и подвержен ошибкам распознавания локаций.

Свойства метрических и топологических карт объединяют графы сцены – представления, сочетающие геометрическую, семантическую и топологическую информацию об окружающей среде. В таких графах вершины могут соответствовать помещениям, локациям и отдельным объектам, а рёбра кодируют пространственные или семантические отношения между ними. Графы сцены обеспечивают богатое структурированное описание среды, пригодное для высокоуровневого понимания сцены, социально-ориентированной навигации и семантических рассуждений. Однако существующие подходы к их построению, как правило, предназначены для полного понимания сцены (вплоть до отношений вида «кружка стоит на столе»), а не для навигации [7]. В результате построенные ими графы остаются плотными и громоздкими, их построение требует больших вычислительных затрат, и их поддержание на роботизированных платформах с ограниченными ресурсами затруднительно (см. рис. 1). Среди таких методов можно выделить Hydra [8] – систему для построения иерархических трёхмерных графов сцен, объединяющую метрические сетки с семантической информацией, и Khronos [9] – расширение Hydra для динамических сред с моделированием временной размерности.

В данной работе предлагается метод построения компактного графа сцены (Compact Scene Graph, CSG) – промежуточного представления, оптимизированного для автономной навигации мобильного робота в крупных помещениях. Основная идея заключается в использовании вычислительно эффективного метода топологического картирования PRISM-ТороМар в

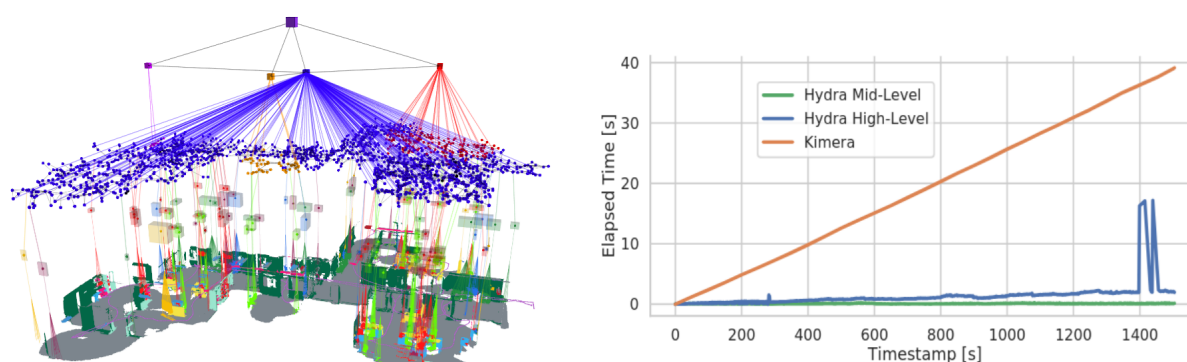


Рис. 1. Граф сцены, построенный методом Hydra (слева), и график времени обновления карты (справа). Граф для представленного помещения занимает 200 МБ, на графике виден линейный рост времени обновления карты. Источник [8]

качестве структурной основы, что обеспечивает компактность и масштабируемость графа. Навигационные вершины, соответствующие посещённым локациям, дополняются облегчённым семантическим слоем: обнаруженные объекты (двери, мебель, офисная техника и т.д.) привязываются к локациям топологической карты, с которых они наблюдались, с обязательной дедупликацией – каждый физический объект представлен ровно одной вершиной, даже если он виден с нескольких локаций. Такое представление позволяет использовать семантическую информацию для переранжирования кандидатов на этапе локализации, повышая её надёжность в визуально неразличимых областях, при этом потребление памяти на хранение и поддержание карты остается значительно более низким, чем у метрических методов картирования и традиционных методов построения графов сцен. Для локализации по построенному CSG в данной работе предлагается использовать коэффициент Жаккара, IDF-взвешивание (IDF – Inverse Document Frequency) и позиционно-осознанный режим.

Экспериментальная оценка предложенного метода проводилась в симуляторе Habitat [10] на сценах, моделирующих помещения большой площади. Результаты показали, что CSG потребляет значительно меньше памяти, чем плотные метрические карты, и превосходит полностью топологические подходы по надёжности локализации в визуально монотонной среде большой площади. Использование семантического переранжирования снизило среднюю ошибку локализации с 2,10 м до 1,20 м при сохранении компактности графового представления. Полученные результаты подтверждают эффективность предложенного подхода как компромисса между геометрической точностью, семантической насыщенностью и вычислительной эффективностью.

1. Обзор литературы

1.1. Методы топологического картирования

Топологические методы представляют среду в виде графа, где вершины соответствуют локациям, а рёбра – пространственной связности между ними. Такие подходы обеспечивают компактность представления и эффективность планирования маршрутов, однако зачастую не содержат семантической информации, необходимой для надёжной локализации. Одним из наиболее эффективных онлайн-методов в этой области является PRISM-ТороМар [5], который строит граф локаций с помощью обучаемого мультимодального

модуля распознавания мест и сопоставления сканов, не полагаясь на точные глобальные координаты робота и поддерживая компактность графа. В работе [6] предлагается топологический семантический граф памяти (TSGM), адаптированный для навигации до целевого изображения; этот метод объединяет структуру обучаемой памяти с навигационными политиками, однако может порождать рёбра между несмежными местоположениями, что снижает его надёжность. Другой подход [11] реализует модульную систему автономного исследования и семантического картирования крупных помещений, в которой на основе построенной метрической карты занятости формируется топологический граф объектов, обновляемый при последующих обходах. Помимо перечисленных, существуют методы, основанные на адаптивной резонансной теории (Adaptive Resonance Theory [12]), иерархическом обучении с подкреплением (TopoNav [13]) и семантическом обнаружении циклов (SemanticTopoLoop [14]), которые расширяют возможности топологических карт в задачах долговременной навигации и локализации в сложных условиях.

1.2. Методы построения графов сцены

На настоящий момент разработано множество методов построения графов сцен, которые различаются по целевой задаче, вычислительной сложности и области применения. В зависимости от режима работы их можно разделить на офлайн-методы, работающие с уже реконструированными данными (например, облаками точек и видеопоследовательностями), и онлайн-методы, выполняющие построение графа в реальном времени по мере поступления сенсорной информации.

Офлайн-методы предназначены для обработки заранее собранных данных, таких как трёхмерные реконструкции помещений или облака точек, и, как правило, обеспечивают высокое качество построения графа за счёт возможности использования глобальной информации. Одним из первых подходов в этом направлении стала работа [15], в которой авторы предлагают алгоритм, строящий семантические графы сцен из реконструированных трёхмерных облаков точек с использованием обучаемых признаков объектов и графовых нейронных сетей для предсказания отношений между ними. Альтернативную стратегию предлагает метод [16], рассматривающий рёбра как первичные сущности: здесь используется рёберно-центрированная агрегация для итеративного уточнения предсказаний отношений, что позволяет повысить согласованность построенного графа по сравнению с вершинно-центрированными архитектурами. Помимо указанных работ, в последние годы появилось множество других офлайн-подходов, основанных на трансформерных архитектурах (например, SGFormer++ [17]), кросс-модальном выравнивании с использованием CLIP [18], а также самоконтролируемом обучении (SGRec3D [19]) и языковых онтологиях. Все эти методы демонстрируют высокую точность на эталонных датасетах, однако их вычислительная сложность и зависимость от качества входной реконструкции ограничивают применение в реальных робототехнических системах.

В отличие от офлайн-подходов, онлайн-методы ориентированы на построение графов сцены в реальном времени по мере движения робота, что делает их пригодными для использования на бортовых вычислительных системах. Система Hydra [8] строит иерархический трёхмерный граф,

включающий здания, комнаты, места, объекты и метрическую карту в виде трёхмерной полигональной сетки (меша), объединяя геометрию и семантику в единое структурированное представление. Расширением этой идеи является Khronos [9], который адаптирует построение графов для динамических сред, моделируя фоновую геометрию, фрагменты объектов, полные объекты и позиции робота в четырёх измерениях (пространство и время). Графы сцен, построенные подобными методами, содержат полную информацию о сцене, однако являются очень громоздкими (например, граф одного этажа здания, построенный методом Hydra, занимает 200 МБ памяти при хранении). Такие методы подходят только для картирования небольших помещений, поскольку, как показано авторами метода Hydra, время на одну итерацию обновления графа растёт линейно со временем работы системы.

В отличие от перечисленных систем, строящих полное описание сцены, метод Clío [7] строит граф, ориентированный на конкретную задачу: используя текстовое описание задач, система сохраняет только те объекты и области, которые релевантны для достижения поставленной цели, что позволяет значительно сократить размер карты и вычислительные затраты. К числу других онлайн-подходов относятся vS-Graphs [20], S-Graphs+ [21] и ConceptGraphs [22], которые интегрируют понимание сцены непосредственно в процесс картирования и обеспечивают совместную оптимизацию метрических, топологических и семантических уровней представления. Методы vS-Graphs и S-Graphs+ строят компактное описание сцены с элементами семантической информации (определение стен помещений), однако в графе отсутствуют многие полезные семантические классы (двери, элементы мебели). Метод ConceptGraphs, напротив, даёт подробное описание сцены с логическими связями, однако для построения этих связей нужны большие языковые модели, использование которых на реальных роботах затруднительно ввиду ограниченности вычислительных ресурсов.

Таким образом, с учетом ограничений всех вышеперечисленных методов, в рамках данной работы был разработан новый метод построения компактных, но семантически информативных графов сцены путем добавления информации о семантических объектах в топологический метод PRISM-ТороМар.

2. Постановка задачи

В данной работе рассматривается задача построения компактного графа сцены, пригодного для локализации мобильного робота в помещениях, по данным с бортовых датчиков робота. Предполагается, что на первом этапе робот строит граф сцены в реальном времени при проезде по заранее заданному маршруту, охватывающему всю доступную для навигации среду. Далее робот осуществляет навигацию в среде, локализуясь с помощью построенного графа, без использования внешних источников позиционирования и глобальных метрических карт.

На каждом шаге t доступны следующие данные с датчиков робота:

- $O_t = (I_t, D_t)$ – **наблюдение**, состоящее из панорамного изображения (угол обзора 360°) и его карты глубин;
- $\Delta p_t \in SE(3)$ – **одометрия** (оценка перемещения робота от шага $t - 1$ к

шагу t).

Требуется разработать алгоритм *Mapping*, который по таким данным будет строить граф $G = (V, E)$.

Граф $G = (V, E)$ содержит два типа вершин и ребер:

- $V = V_{nav} \cup V_{sem}$, где V_{nav} – это множество **навигационных вершин** (локаций), V_{sem} – множество вершин, соответствующих семантическим объектам. Каждая локация $v \in V_{nav}$ имеет свою систему координат с началом координат (v^x, v^y) и углом поворота v^θ в глобальной системе координат. Каждый **семантический объект** $w \in V_{sem}$ задается координатами центра объекта в глобальной системе координат и меткой класса w^c .
- $E = E_{nav} \cup E_{sem}$, где ребра $(v_i, v_j) \in E_{nav}$ связывают пространственно смежные локаций из V_{nav} , а ребра $(v_i, w_j) \in E_{sem}$ связывают объект w_j из V_{sem} с локацией v_i из V_{nav} , с которой он был обнаружен. Один и тот же семантический объект может быть связан ребрами с несколькими локациями.

Алгоритм должен работать в реальном времени, т. е. на каждом шаге t выдавать обновленный граф G_t :

$$Mapping(G_{t-1}, O_t, \Delta p_t) = G_t. \quad (1)$$

Также требуется разработать алгоритм локализации *Loc*, который по графу G и данным с робота $(O_t, \Delta p_t)$ будет определять позицию робота в графе – локацию v_{cur}^t , в которой находится робот в момент t , и положение робота $\delta_{cur}^t = (\delta x, \delta y, \delta \theta)$ в системе координат локации:

$$Loc(G, O_t, \Delta p_t) = (v_{cur}^t, \delta_{cur}^t). \quad (2)$$

Целевой метрикой является ошибка локализации – расхождение между истинным положением робота и положением робота, определенным алгоритмом локализации и переведенным в глобальную систему координат:

$$error = \left\| \begin{pmatrix} x_t \\ y_t \end{pmatrix} - \begin{pmatrix} x_{cur} + \cos(\theta_{cur}) \delta_x + \sin(\theta_{cur}) \delta_y \\ y_{cur} - \sin(\theta_{cur}) \delta_x + \cos(\theta_{cur}) \delta_y \end{pmatrix} \right\|, \quad (3)$$

где (x_t, y_t) – истинные координаты робота в момент t ; $(x_{cur}, y_{cur}, \theta_{cur})$ – координаты и угол ориентации локации v_{cur}^t .

3. Предложенный метод построения компактного графа сцены

Предложенный метод в реальном времени строит граф сцены $G = (V, E)$, состоящий из навигационного подграфа (V_{nav}, E_{nav}) и семантического слоя (V_{sem}, E_{sem}) . Навигация и планирование маршрутов осуществляются по навигационному подграфу. Семантический слой задействуется на этапе локализации, а также может быть использован для задания целевой точки для навигации. Пример построенного графа с семантическим слоем приведен на рисунке 2.

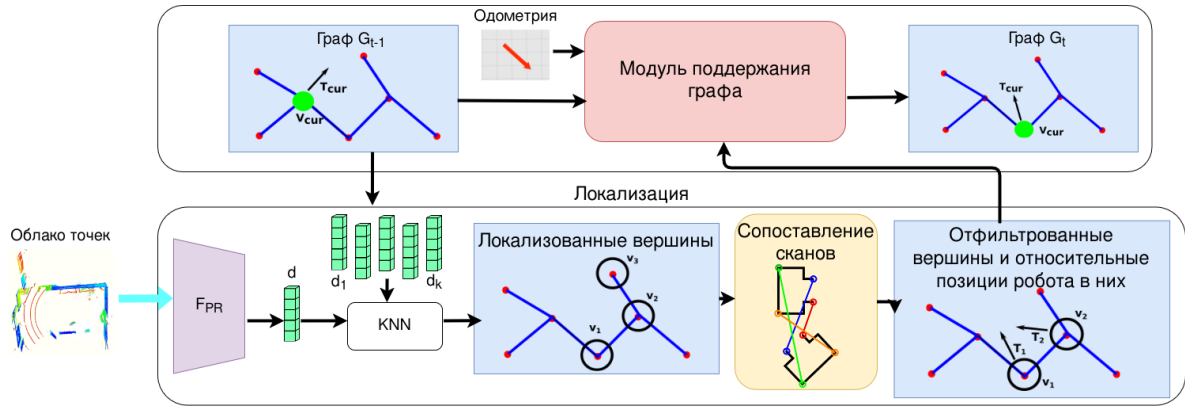


Рис. 3. Схема метода построения навигационного графа PRISM-ТороМар. Источник [5]

3.2. Семантический слой

Для создания графа сцены к навигационному графу, который строится методом PRISM-ТороМар, добавляется семантический слой. Найденные семантические объекты соединяются ребрами с локациями, из которых они были обнаружены. Для обнаружения семантических объектов используется нейросетевая модель GroundingDINO [23]. На вход модели подаются части панорамного изображения с камер робота (4 изображения с охватом 90 градусов каждое) и текстовый запрос, задающий перечень семантических классов. Модель возвращает ограничивающие прямоугольники объектов на изображении, текстовые метки классов и оценки уверенности. Для каждого обнаруженного объекта определяется его положение в пространстве – вычисляется медианная глубина в квадрате размером 5×5 пикселей вокруг центра ограничивающего прямоугольника, далее по глубине и положению на изображении вычисляется положение центра объекта в системе координат робота, которое затем переводится в глобальную систему координат.

Для исключения повторного добавления объектов в граф проводится дедупликация следующим образом. После вычисления глобальных координат объекта (w_x, w_y) проводится поиск ближайшего объекта того же класса $w_{nearest}$ в семантическом подграфе V_{sem} . Если он найден и расстояние до него не превышает порог r_{dup} , то новый семантический объект не создается, а найденный объект $w_{nearest}$ соединяется ребром с локацией v_{cur} . В противном случае создается новый семантический объект w_{new} с координатами (w_x, w_y) , добавляется в семантический подграф и соединяется с локацией v_{cur} .

3.3. Локализация в графе сцены

Модуль локализации на каждом шаге определяет вершину $v_{cur} \in V_{nav}$, в которой находится робот, и его относительное положение $\delta_{cur}^t = (\delta x, \delta y, \delta \theta)$ внутри локации v_{cur} . Локализация проводится в два этапа; первый этап совпадает с базовым методом PRISM-ТороМар. В ходе первого этапа отбираются k локаций-кандидатов с помощью метода распознавания мест. Далее найденные кандидаты переупорядочиваются с использованием объединенной оценки:

$$score(v_i) = (1 - \alpha) (-\|f_{cur} - f_{v_i}\|_2) + \alpha s(v_i), \quad (4)$$

где $\alpha \in [0, 1]$ – вес семантической составляющей, f_{cur} – дескриптор текущего наблюдения, f_{v_i} – дескриптор вершины v_i , $s(v_i)$ – семантическая оценка кандидата. Оценка вычисляется только для тех кандидатов, для которых было успешно найдено сопоставление сканов с помощью базового метода PRISM-ТороМар. Для вычисления величины $s(v_i)$ применяется несколько способов:

Коэффициент Жаккара (jaccard):

$$s(v_i) = J(C_{cur}, C_{v_i}) = \frac{|C_{cur} \cap C_{v_i}|}{|C_{cur} \cup C_{v_i}|}, \quad (5)$$

где C_{cur} – множество классов объектов, обнаруженных в текущем кадре, C_{v_i} – множество классов объектов, наблюдаемых из вершины v_i .

IDF-взвешенный коэффициент Жаккара (idf):

$$idf(c) = \log \frac{N + 1}{df(c) + 1}, \quad (6)$$

где N – число вершин графа, $df(c)$ – число вершин, содержащих хотя бы один объект класса c . Семантическая оценка вычисляется как взвешенный коэффициент Жаккара:

$$s(v_i) = \frac{\sum_{c \in C_{cur} \cap C_{v_i}} idf(c)}{\sum_{c \in C_{cur} \cup C_{v_i}} idf(c)}. \quad (7)$$

Позиционно-осознанный режим (position):

Для вычисления семантической оценки в данном режиме учитываются не только классы обнаруженных объектов, но и их координаты относительно робота. Для каждого объекта, обнаруженного в текущем кадре, зададим тройку (c, o_x, o_y) – семантический класс объекта и его координаты относительно робота. При рассмотрении вершины-кандидата v_i все семантические объекты из V_{sem} , соединенные с ней, переводятся в систему координат робота. Объект с текущего наблюдения $d = (c, o_x, o_y)$ считается сопоставленным с вершиной v_i , если в ней существует объект класса c с координатами относительно робота (p_x, p_y) , для которого выполняется условие сопоставления $matched(d, v_i)$:

$$matched(d, v_i) \Leftrightarrow \left\| \begin{pmatrix} p_x - o_x \\ p_y - o_y \end{pmatrix} \right\|_2 < r_{pos}, \quad (8)$$

где r_{pos} – порог допустимого отклонения координат. Семантическая оценка вычисляется как доля сопоставленных объектов:

$$s(v_i) = \frac{|\{d \in D_{cur} : matched(d, v_i)\}|}{|D_{cur}|}, \quad (9)$$

где D_{cur} – множество объектов, обнаруженных в текущем кадре. Если после переранжирования наилучшая вершина $v^* = \arg \max_{v \in V_{nav}} score(v)$ имеет два и более семантических объекта, сопоставленных с текущим кадром, то проводится дополнительное уточнение оценки положения робота относительно вершины v^* . По координатам объектов вершины v^* и координатам сопоставленных им объектов в текущем кадре вычисляется оптимальное положение робота относительно v^* с помощью метода наименьших квадратов:

$$\min_{x_m, y_m, \theta} \sum_{j=1}^k \left\| R^T(\theta) \begin{pmatrix} l_x^{(j)} - x_m \\ l_y^{(j)} - y_m \end{pmatrix} - \begin{pmatrix} o_x^{(j)} \\ o_y^{(j)} \end{pmatrix} \right\|, \quad (10)$$

где $(l^{(j)} \in \mathbb{R}^2, o^{(j)} \in \mathbb{R}^2)$ – множество сопоставленных пар двумерных координат объектов, $R^T(\theta) \in SO(2)$ – матрица поворота координат на угол θ .

4. Экспериментальное исследование

4.1. Постановка эксперимента

Эксперименты по оценке качества предложенного метода проводились в симуляционной среде, представляющей собой модель помещения большой площади. Эксперименты проводились в два этапа. Вначале выполнялся обход среды виртуальным агентом по заранее заданному маршруту, и во время обхода строился граф сцены с помощью предложенного метода. Затем выполнялась серия навигационных заданий с заданными точками старта и финиша, в ходе которых выполнялась локализация виртуального агента по построенному на первом этапе графу и оценивалась ошибка локализации. Навигация осуществлялась средствами симулятора по кратчайшему пути от старта до финиша. Ошибка локализации измерялась как среднее расхождение между истинным положением робота и вычисленным с помощью локализации на каждом шаге траектории.

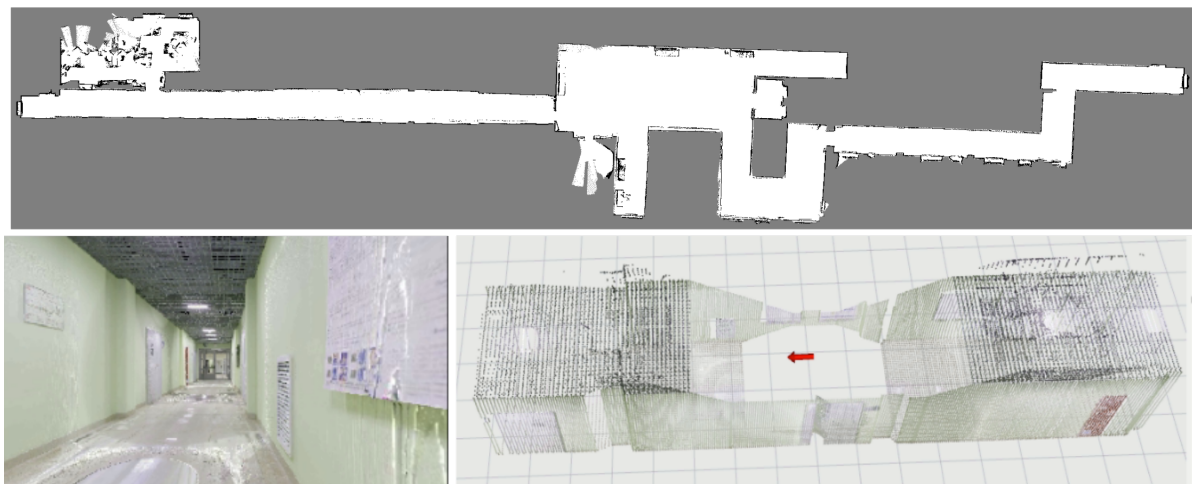


Рис. 4. Карта среды, используемой для экспериментов (вверху), изображение и облако точек из симулятора (внизу)

Для экспериментов использовался симулятор Habitat [10]. В качестве сцены использовалась модель пятого этажа корпуса «Цифра» Московского физико-технического института. Площадь сцены составила 1600 м², длина маршрута для обхода – 300 м. Пример сцены показан на рисунке 4. Ошибка локализации оценивалась как среднее отклонение истинной позиции робота от позиции, соответствующей состоянию робота в графе, определенному с помощью локализации. Движение и локализация выполнялись асинхронно – виртуальный агент двигался к целевой точке независимо от того, был ли завершен текущий шаг локализации.

4.2. Результаты

Результаты эксперимента представлены в таблице 1. Тестировалось четыре варианта локализации – без использования семантической информации и с семантическим переранжированием по всем трем вышеописанным формулам. Эксперименты проводились на компьютере со следующими характеристиками: 6-ядерный процессор Intel Core i5-11500, 32 ГБ оперативной памяти, видеокарта GeForce RTX 2060 с 6 ГБ видеопамати. Помимо ошибки локализации по построенному графу сцены, проводились измерения задержки локализации (времени от получения сенсорных данных до вычисления состояния робота в графе), а также размера построенного графа в хранилище.

Таблица 1. Результаты экспериментов с различным весом семантики и различными режимами переранжирования

α	Переранжирование	Ошибка локализации, м	Задержка локализации, с	Размер графа, МБ
0,0	–	2,10	0,80	0,374
0,3	Jaccard	1,20	0,87	0,403
0,3	IDF	1,36	0,87	0,403
0,3	Position	1,18	0,87	0,403

Как видно из таблицы, использование семантической информации снижает ошибку локализации на 35–44%, в зависимости от метода переранжирования. Наилучший результат показывает позиционно-осознанный режим переранжирования, сокращая ошибку локализации почти в два раза. Высокая ошибка локализации (более 2 м) без семантики обусловлена наличием в сцене длинных однообразных коридоров, разные участки которых имеют одинаковую геометрическую форму при проекции на плоскость. Добавление семантической информации позволяет различать участки коридоров по дверям, мебели, пожарным шкафам и т. д., что уменьшает число ложных локализаций и, как следствие, среднюю ошибку. При добавлении семантической информации задержка локализации и потребление памяти построенным графом увеличиваются незначительно. Граф сцены помещения площадью 1600 м² занимает всего 400 КБ дискового пространства – на порядки меньше, чем графы сцены, построенные традиционными методами, и метрические карты.

Заключение

В данной работе предложен метод построения компактного графа сцены, пригодный для автономной навигации мобильных роботов в помещениях большой площади. Метод строит граф, состоящий из двух слоев: навигационного подграфа, описывающего пространственную связность локаций, и семантического слоя, содержащего информацию об объектах, обнаруженных в локациях. Построение навигационного подграфа и локализация в нем выполняются методом PRISM-ТороМар, в котором локализация проводится в два этапа: поиск текущей локации методом распознавания мест и уточнение позиции методом сопоставления двумерных проекций сканов. Объекты семантического слоя используются для переупорядочивания локаций-кандидатов

навигационного графа, найденных с помощью распознавания мест. Предложено три режима переупорядочивания: на основе коэффициента Жаккара, IDF-взвешенного коэффициента и позиционно-осознанного сопоставления объектов. В последнем режиме также проводится уточнение положения робота с использованием позиций объектов.

Эксперименты, проведенные в симуляционной среде большой площади, показали, что граф сцены, построенный предложенным методом, занимает незначительно больше памяти, чем компактный топологический граф метода PRISM-ТороМар без семантики (400 КБ на сцену площадью 1600 м²), при этом ошибка локализации по графу с добавлением семантической информации уменьшается почти в два раза. В дальнейшем планируется доработка предложенного метода и экспериментальное сравнение с другими методами построения графов сцены, настройка и обучение весового коэффициента α , а также проведение экспериментов с навигацией по построенному графу сцены и переход из симуляции на реальные робототехнические платформы.

Список литературы

- [1] Labbé, M. RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation / M. Labbé, F. Michaud // *Journal of Field Robotics*. – 2019. – Vol. 36, No. 2. – P. 416–446. – DOI 10.1002/rob.21831.
- [2] GLIM: 3D range-inertial localization and mapping with GPU-accelerated scan matching factors / K. Koide, M. Yokozuka, S. Oishi, A. Banno // *Robotics and Autonomous Systems*. – 2024. – Vol. 179. – P. 104750. – DOI 10.1016/j.robot.2024.104750.
- [3] Muravyev, K. Evaluation of RGB-D SLAM in large indoor environments / K. Muravyev, K. Yakovlev // *International Conference on Interactive Collaborative Robotics*. – 2022. – P. 93–104. – DOI 10.1007/978-3-031-23609-9_9.
- [4] Muravyev, K. Evaluation of topological mapping methods in indoor environments / K. Muravyev, K. S. Yakovlev // *IEEE Access*. – 2023. – Vol. 11. – P. 132683–132698. – DOI 10.1109/ACCESS.2023.3335818.
- [5] PRISM-TopoMap: Online topological mapping with place recognition and scan matching / K. Muravyev, A. Melekhin, D. Yudin, K. Yakovlev // *IEEE Robotics and Automation Letters*. – 2025. – Vol. 10, No. 4. – P. 3126–3133. – DOI 10.1109/LRA.2025.3541454.
- [6] Topological semantic graph memory for image-goal navigation / N. Kim, O. Kwon, H. Yoo [et al.] // *Proceedings of the 6th Conference on Robot Learning (CoRL)*, PMLR. – 2023. – Vol. 205. – P. 393–402.
- [7] Clio: Real-time task-driven open-set 3D scene graphs / D. Maggio, Y. Chang, N. Hughes [et al.] // *IEEE Robotics and Automation Letters*. – 2024. – Vol. 9, No. 10. – P. 8921–8928. – DOI 10.1109/LRA.2024.3451395.
- [8] Hughes, N. Hydra: A real-time spatial perception system for 3D scene graph construction and optimization / N. Hughes, Y. Chang, L. Carlone // *arXiv*. – 2022. – arXiv:2201.13360. – DOI 10.48550/arXiv.2201.13360.
- [9] Khronos: Middleware for simplified time management in CPS / S. Peros, S. Delbruel, S. Michiels [et al.] // *Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems*. – 2019. – P. 127–138. – DOI 10.1145/3328905.3329507.
- [10] Habitat: A platform for embodied AI research / M. Savva, A. Kadian, O. Maksymets [et al.] // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. – 2019. – P. 9338–9346. – DOI 10.1109/ICCV.2019.00943.
- [11] A modular robotic system for autonomous exploration and semantic updating in large-scale indoor environments / S. H. Allu, I. Kadosh, T. Summers, Y. Xiang // *arXiv*. – 2024. – arXiv:2409.15493. – DOI 10.48550/arXiv.2409.15493.

- [12] Carpenter, G. A. Adaptive resonance theory : Tech. Rep. / G. A. Carpenter, S. Grossberg. – Boston, MA, USA : Boston University, Center for Adaptive Systems and Department of Cognitive and Neural Systems, 1993.
- [13] TopoNav: Topological navigation for efficient exploration in sparse reward environments / J. Hossain, A. Z. M. Faridee, N. Roy [et al.] // 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). – 2024. – P. 693–700. – DOI 10.1109/IROS58592.2024.10802380.
- [14] SemanticTopoLoop: Semantic loop closure with 3D topological graph based on quadric-level object map / Z. Cao, Q. Zhang, J. Guang [et al.] // IEEE Robotics and Automation Letters. – 2024. – Vol. 9, No. 5. – P. 4257–4264. – DOI 10.1109/LRA.2024.3374169.
- [15] Learning 3D semantic scene graphs from 3D indoor reconstructions / J. Wald, H. Dharmo, N. Navab, F. Tombari // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2020. – P. 3961–3970. – DOI 10.1109/CVPR42600.2020.00402.
- [16] Exploiting edge-oriented reasoning for 3D point-based scene graph analysis / C. Zhang, J. Yu, Y. Song, W. Cai // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2021. – P. 9705–9715. – DOI 10.1109/CVPR46437.2021.00958.
- [17] SGFormer++: Semantic graph transformer for incremental 3D scene graph generation / M. Qi, C. Lv, Z. Fu [et al.] // arXiv. – 2026. – arXiv:2606.15328. – DOI 10.48550/arXiv.2606.15328.
- [18] Learning transferable visual models from natural language supervision / A. Radford, J. W. Kim, C. Hallacy [et al.] // Proceedings of the 38th International Conference on Machine Learning, PMLR. – 2021. – Vol. 139. – P. 8748–8763.
- [19] SGR3D: Self-supervised 3D scene graph learning via object-level scene reconstruction / S. Koch, N. Vaskevicius, M. Colosi [et al.] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. – 2024. – P. 3392–3402. – DOI 10.1109/WACV57701.2024.00337.
- [20] vS-Graphs: Integrating visual SLAM and situational graphs through multi-level scene understanding / A. Tourani, S. Ejaz, H. Bavle [et al.] // arXiv. – 2025. – arXiv:2503.01783. – DOI 10.48550/arXiv.2503.01783.
- [21] S-Graphs+: Real-time localization and mapping leveraging hierarchical representations / H. Bavle, J. L. Sánchez-López, M. Shaheer [et al.] // IEEE Robotics and Automation Letters. – 2023. – Vol. 8, No. 8. – P. 4927–4934. – DOI 10.1109/LRA.2023.3290512.
- [22] ConceptGraphs: Open-vocabulary 3D scene graphs for perception and planning / Q. Gu, A. Kuwajerwala, S. Morin [et al.] // 2024 IEEE International Conference on Robotics and Automation (ICRA). – 2024. – P. 5021–5028. – DOI 10.1109/ICRA57147.2024.10610243.
- [23] Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection / S. Liu, Z. Zeng, T. Ren [et al.] // Computer Vision – ECCV 2024, Springer. – 2024. – Vol. 47. – P. 38–55. – DOI 10.1007/978-3-031-72970-6_3.

BUILDING COMPACT SCENE GRAPHS BASED ON A TOPOLOGICAL MAP FOR AUTONOMOUS NAVIGATION OF A MOBILE ROBOT

Muravyev K.F.^{1, a}, Romanenko V.I.^{2, b}

¹ Federal Research Center «Computer Science and Control» of the Russian Academy of Sciences (FRC CSC RAS), Moscow, Russia.

² National Research University Higher School of Economics (NRU HSE), Moscow, Russia.

E-mail: ^a kmuravev@frccsc.ru, ^b viktorsamr@yandex.ru

Abstract. Autonomous navigation of a mobile robot in human-centered environments requires a map that contains not only a geometric model of the environment for path planning, but also information about environmental objects (doors, furniture, office equipment, etc.). Scene graphs provide such a map representation, where nodes correspond to rooms, locations, and objects, and edges encode spatial connectivity or relationships between objects. Most modern scene graph construction methods have high computational complexity, and the graphs they produce are redundant for the purposes of autonomous robot navigation. This paper proposes a method for constructing Compact Scene Graphs (CSG), which is based on the computationally efficient topological mapping method PRISM-TopoMap and the association of semantic objects with locations on the topological map. The resulting scene graph enables route planning to objects via topological map locations and achieves high-precision localization. The proposed method was experimentally evaluated in the Habitat simulation environment. The experimental results demonstrate that the proposed CSG consumes significantly less memory than traditional metric maps and scene graphs, while providing reliable localization on the topological map through association with semantic objects.

Keywords: *scene graph, topological map, semantic segmentation, autonomous navigation, simulated environment.*

References

- [1] Labbé, M., & Michaud, F. (2019). RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of Field Robotics*, 36(2), 416–446. <https://doi.org/10.1002/rob.21831>
- [2] Koide, K., Yokozuka, M., Oishi, S., & Banno, A. (2024). GLIM: 3D range-inertial localization and mapping with GPU-accelerated scan matching factors. *Robotics and Autonomous Systems*, 179, Article 104750. <https://doi.org/10.1016/j.robot.2024.104750>
- [3] Muravyev, K., & Yakovlev, K. (2022). Evaluation of RGB-D SLAM in large indoor environments. In *International Conference on Interactive Collaborative Robotics* (pp. 93–104). https://doi.org/10.1007/978-3-031-23609-9_9

For citation:

Muravyev K.F., Romanenko V.I. Building Compact Scene Graphs Based on a Topological Map for Autonomous Navigation of a Mobile Robot. *Intelligent transport*, 2026, no. 3(39), pp. 37–52. DOI: 10.24412/3033-6007-2026-339-37-52. (in Russian)

- [4] Muravyev, K., & Yakovlev, K. S. (2023). Evaluation of topological mapping methods in indoor environments. *IEEE Access*, 11, 132683–132698. <https://doi.org/10.1109/ACCESS.2023.3335818>
- [5] Muravyev, K., Melekhin, A., Yudin, D., & Yakovlev, K. (2025). PRISM-TopoMap: Online topological mapping with place recognition and scan matching. *IEEE Robotics and Automation Letters*, 10(4), 3126–3133. <https://doi.org/10.1109/LRA.2025.3541454>
- [6] Kim, N., Kwon, O., Yoo, H., Choi, Y., Park, J., & Oh, S. (2023). Topological semantic graph memory for image-goal navigation. In *Proceedings of the 6th Conference on Robot Learning (CoRL)* (Vol. 205, pp. 393–402). PMLR.
- [7] Maggio, D., Chang, Y., Hughes, N., Trang, M., Griffith, D., Dougherty, C., Cristofalo, E., Schmid, L., & Carlone, L. (2024). Clio: Real-time task-driven open-set 3D scene graphs. *IEEE Robotics and Automation Letters*, 9(10), 8921–8928. <https://doi.org/10.1109/LRA.2024.3451395>
- [8] Hughes, N., Chang, Y., & Carlone, L. (2022). Hydra: A real-time spatial perception system for 3D scene graph construction and optimization (arXiv:2201.13360). arXiv. <https://doi.org/10.48550/arXiv.2201.13360>
- [9] Peros, S., Delbruel, S., Michiels, S., Joosen, W., & Hughes, D. (2019). Khronos: Middleware for simplified time management in CPS. In *Proceedings of the 13th ACM International Conference on Distributed and Event-based Systems* (pp. 127–138). <https://doi.org/10.1145/3328905.3329507>
- [10] Savva, M., Kadian, A., Maksymets, O., Zhao, Y., Wijmans, E., Jain, B., Straub, J., Liu, J., Koltun, V., Malik, J., Batra, D., & Mottaghi, R. (2019). Habitat: A platform for embodied AI research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9338–9346). <https://doi.org/10.1109/ICCV.2019.00943>
- [11] Allu, S. H., Kadosh, I., Summers, T., & Xiang, Y. (2024). A modular robotic system for autonomous exploration and semantic updating in large-scale indoor environments (arXiv:2409.15493). arXiv. <https://doi.org/10.48550/arXiv.2409.15493>
- [12] Carpenter, G. A., & Grossberg, S. (1993). *Adaptive resonance theory* (Technical Report). Boston University, Center for Adaptive Systems and Department of Cognitive and Neural Systems.
- [13] Hossain, J., Faridee, A. Z. M., Roy, N., Freeman, J., Gregory, T., & Trout, T. (2024). TopoNav: Topological navigation for efficient exploration in sparse reward environments. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 693–700). <https://doi.org/10.1109/IROS58592.2024.10802380>
- [14] Cao, Z., Zhang, Q., Guang, J., Wu, S., Hu, Z., & Liu, J. (2024). SemanticTopoLoop: Semantic loop closure with 3D topological graph based on quadric-level object map. *IEEE Robotics and Automation Letters*, 9(5), 4257–4264. <https://doi.org/10.1109/LRA.2024.3374169>
- [15] Wald, J., Dhama, H., Navab, N., & Tombari, F. (2020). Learning 3D semantic scene graphs from 3D indoor reconstructions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3961–3970). <https://doi.org/10.1109/CVPR42600.2020.00402>
- [16] Zhang, C., Yu, J., Song, Y., & Cai, W. (2021). Exploiting edge-oriented reasoning for 3D point-based scene graph analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9705–9715). <https://doi.org/10.1109/CVPR46437.2021.00958>
- [17] Qi, M., Lv, C., Fu, Z., Zhang, X., & Ma, H. (2026). SGFormer++: Semantic graph transformer for incremental 3D scene graph generation (arXiv:2606.15328). arXiv. <https://doi.org/10.48550/arXiv.2606.15328>
- [18] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning* (Vol. 139, pp. 8748–8763). PMLR.
- [19] Koch, S., Vaskevicius, N., Colosi, M., Hermosilla, P., & Ropinski, T. (2024). SGRec3D: Self-supervised 3D scene graph learning via object-level scene reconstruction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 3392–3402). <https://doi.org/10.1109/WACV57701.2024.00337>

- [20] Tourani, A., Ejaz, S., Bavle, H., Morilla-Cabello, D., Sanchez-Lopez, J. L., & Voos, H. (2025). vS-Graphs: Integrating visual SLAM and situational graphs through multi-level scene understanding (arXiv:2503.01783). arXiv. <https://doi.org/10.48550/arXiv.2503.01783>
- [21] Bavle, H., Sánchez-López, J. L., Shaheer, M., Civera, J., & Voos, H. (2023). S-Graphs+: Real-time localization and mapping leveraging hierarchical representations. *IEEE Robotics and Automation Letters*, 8(8), 4927–4934. <https://doi.org/10.1109/LRA.2023.3290512>
- [22] Gu, Q., Kuwajerwala, A., Morin, S., Jatavallabhula, K. M., Sen, B., Agarwal, A., Rivera, C., Paul, W., Ellis, K., Chellappa, R., Gan, C., de Melo, C. M., Tenenbaum, J. B., Torralba, A., Shkurti, F., & Paull, L. (2024). ConceptGraphs: Open-vocabulary 3D scene graphs for perception and planning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 5021–5028). <https://doi.org/10.1109/ICRA57147.2024.10610243>
- [23] Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2024). Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. In *Computer Vision – ECCV 2024* (Vol. 47, pp. 38–55). Springer. https://doi.org/10.1007/978-3-031-72970-6_3

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДОВ ОБНАРУЖЕНИЯ ВЫБРОСОВ В ДАННЫХ МОНИТОРИНГА НЕФТЯНЫХ СКВАЖИН НА ОСНОВЕ SVM И DBSCAN*

Михайлов И.С.^{1, а}, Мьо Хлайн Вин^{1, b}, Сидоров К.О.^{1, c}

¹ Национальный исследовательский университет «Московский энергетический институт» (НИУ МЭИ), Москва, Россия.

E-mail: ^a fr82@mail.ru, ^b myohlaingwin69287@gmail.com, ^c kirill.sidoroff2014@yandex.ru

Аннотация. Цель исследования – сопоставить результаты One-Class Support Vector Machine (One-Class SVM) и DBSCAN при обнаружении статистических аномалий в данных мониторинга нефтяной скважины и оценить возможность их совместного использования. В анализе использованы шесть параметров многофазного измерителя; после преобразования методом главных компонент (PCA) они представлены тремя главными компонентами, сохраняющими 89,86% общей дисперсии. На выборке из 8639 наблюдений One-Class SVM выявил 424 аномальные точки, DBSCAN – 188; 85 наблюдений обнаружены обоими методами. Для классификации режимов течения по уровню газосодержания SVM с RBF-ядром ($C = 100,0$, $\gamma = 1,0$) на стратифицированном разбиении 70/30 показал точность 98,26% при F1-мере 0,97–0,99. При 5-кратной стратифицированной кросс-валидации F1_масго составила $0,9853 \pm 0,0026$. Совместное рассмотрение результатов позволяет выделять согласованные двумя методами, обнаруженные только SVM и обнаруженные только DBSCAN статистические выбросы для последующей экспертной проверки.

Ключевые слова: обнаружение выбросов, метод опорных векторов, DBSCAN, метод главных компонент, мониторинг нефтяных скважин, машинное обучение, классификация режимов течения, аномалии.

Введение

Цель настоящего исследования – сравнительный анализ результатов двух методов обнаружения аномалий (One-Class SVM и DBSCAN) на данных мониторинга нефтяной скважины и разработка практических рекомендаций по их совместному применению. Совместные и отдельные срабатывания используются для ранжирования статистических выбросов по приоритету экспертной проверки.

Для достижения цели поставлены следующие задачи:

1. сформировать признаковое пространство из шести параметров, используемых в дальнейшем анализе;
2. применить метод главных компонент для снижения размерности и определить число компонент, сохраняющих 89,86% общей дисперсии;

Для цитирования:

Михайлов И.С., Мьо Хлайн Вин, Сидоров К.О. Сравнительное исследование методов обнаружения выбросов в данных мониторинга нефтяных скважин на основе SVM и DBSCAN. *Интеллектуальный транспорт*. 2026. № 3(39). С. 53–68. DOI: 10.24412/3033-6007-2026-339-53-68.

* Исследование было проведено при поддержке РНФ (проект № 24-11-00285).

3. выполнить обнаружение статистических аномалий методами One-Class SVM и DBSCAN на преобразованных данных;
4. сопоставить результаты двух методов и разделить обнаруженные точки по сочетанию срабатываний алгоритмов;
5. обучить и оценить SVM-классификатор режимов течения по уровням газосодержания на разбиении 70/30 и при 5-кратной стратифицированной кросс-валидации;
6. сформулировать практические рекомендации по совместному применению методов в системах мониторинга.

Статья организована следующим образом. В разделе 1 представлена предметная область и мотивация исследования. В разделе 2 приведено описание исходных данных, в разделе 3 – применение PCA, в разделе 4 – построение и оценка SVM-классификатора режимов течения, в разделе 5 – обнаружение локальных аномалий методом DBSCAN. Раздел 6 посвящён сравнению One-Class SVM и DBSCAN, раздел 7 – классификации результатов по сочетанию срабатываний методов, раздел 8 – практическим рекомендациям, раздел 9 – выводам по экспериментальному исследованию. В заключении обобщены результаты и обозначены направления дальнейшей работы.

Современные системы мониторинга нефтяных скважин регистрируют множество взаимосвязанных параметров, однако традиционный пороговый контроль рассматривает их раздельно. В результате сложные сочетания признаков могут быть пропущены, а отклонения отдельных показателей – приводить к ложным срабатываниям. Дополнительное ограничение связано с дефицитом размеченных данных о нештатных ситуациях, поэтому для контроля представляют интерес методы машинного обучения без учителя, способные работать с многомерными потоками данных и интегрироваться в системы управления [1,2].

Подходы к обнаружению аномалий включают статистические, дистанционные, плотностные, опорно-векторные и древовидные методы. One-Class SVM формирует глобальную границу нормальных наблюдений, тогда как DBSCAN выявляет локальные точки в областях пониженной плотности, поэтому их совместное применение позволяет анализировать разные типы отклонений. Для устранения корреляционной избыточности и подготовки данных к визуализации используется PCA. Сравнительное исследование One-Class SVM и DBSCAN именно на данных мониторинга нефтяных скважин и методика интерпретации пересечения результатов представлены в литературе неполно [3,4].

Сравнение Isolation Forest и One-Class SVM при обнаружении аномалий в данных эксплуатации газопроводов представлено в работе [5]. Применение методов обнаружения выбросов без учителя для восстановления каротажных данных в осложнённых условиях бурения рассмотрено в работе [6].

Нейросетевые модели с объяснением вклада признаков методом SHAP [7] рассматриваются как альтернатива для дальнейшего сравнения; в эксперимент они не включались.

Исследование направлено на сопоставление глобального и локального методов обнаружения аномалий после снижения размерности и приоритизацию экспертной проверки выбросов при дефиците размеченных данных.

1. Предметная область

Пороговый контроль показаний датчиков нефтяных скважин не учитывает взаимосвязи параметров, что может приводить к ложным срабатываниям и пропуску аномалий [1]. Сложность получения размеченных данных о нештатных ситуациях обуславливает интерес к методам машинного обучения без учителя для автоматического выявления аномалий в режиме реального времени.

С практической стороны в данном исследовании рассматривается обнаружение статистических аномалий в показаниях многофазного измерителя обводнённости, установленного на устье нефтяной скважины. Данный измеритель регистрирует в режиме реального времени набор параметров, позволяющий оценивать профиль течения нефте-водо-газового потока и отслеживать его изменения.

2. Исходные данные

Многофазный измеритель обводнённости нефтяных скважин устанавливается на вертикальном участке нефтепровода. Измеритель оснащается двумя ультразвуковыми сенсорами, расположенными в центре и на периферии нефте-водо-газового потока. Измеритель регистрирует значения скорости ультразвука в центре и на периферии потока, газосодержание в центре и на периферии потока, температуру потока, обводнённость в центре и на периферии потока. На основании расчётов измеритель вычисляет интегральное значение обводнённости и газосодержания потока. Данные записываются в архив измерителя один раз в десять секунд.

Процедура предобработки включала:

1. удаление дублирующей строки заголовка;
2. приведение числовых столбцов к типу `float` с использованием `pd.to_numeric(errors='coerce')`;
3. удаление строк с полностью отсутствующими значениями;
4. удаление строк с пропусками в выбранных признаках.

Интегральное газосодержание используется для формирования целевой переменной и не включается во входное признаковое пространство. Классы низкого, среднего и высокого газосодержания задаются порогами 0,3 и 0,7. Их объёмы составляют соответственно 2184, 2602 и 3853 наблюдения. В анализируемую выборку вошли 8639 наблюдений. Это соответствует сумме категорий, представленных на рисунках 4 и 7. Исследование выполнено на данных многофазного измерителя. Переносимость результатов на другие скважины и оборудование требует отдельной проверки.

Удаление пропусков само по себе не позволяет отличить сбой датчика от технологического отклонения или дрейфа калибровки. В настоящем исследовании эти причины статистических выбросов не разграничиваются.

3. Метод главных компонент (PCA)

Метод главных компонент (Principal Component Analysis, PCA) применяется в данном исследовании для снижения размерности данных мониторинга нефтяных скважин. Исходные данные содержат шесть взаимосвязанных параметров (скорость звука в центре и на периферии, газосодержание в центре и на периферии, температура, интегральная обводнённость), многие из которых коррелируют между собой, что создаёт избыточность информации и затрудняет визуализацию [3, 4]. PCA позволяет преобразовать исходные признаки в новые переменные – главные компоненты, которые являются линейными комбинациями исходных признаков и упорядочены по убыванию дисперсии.

Для определения числа главных компонент применяется оценка объяснённой и кумулятивной дисперсии. Перед применением PCA каждый признак стандартизируется:

$$x_{\text{scaled}} = \frac{x - \mu}{\sigma}, \quad (1)$$

где x – исходное значение признака; μ – его среднее значение; σ – стандартное отклонение.

Первая главная компонента (PC1) объясняет 52,15% общей дисперсии и отражает преимущественно газосодержание и температуру потока: положительные нагрузки для газосодержания в центре (0,510), температуры (0,460) и газосодержания на периферии (0,248) сочетаются с отрицательными нагрузками для скорости звука в центре (–0,490) и интегральной обводнённости (–0,461). Вторая компонента (PC2) объясняет 27,51% дисперсии и связана с распределением параметров по сечению потока (нагрузки Звук_П = 0,706, Газ_П = 0,481, Вода_Инт = 0,409). Третья компонента (PC3) объясняет 10,20% дисперсии и отражает преимущественно периферийное газосодержание (Газ_П = 0,831) с отрицательными нагрузками температуры (–0,371) и скорости звука на периферии (–0,350). Совокупно первые три компоненты сохраняют 89,86% исходной информации, что позволяет использовать их для дальнейшего анализа вместо исходных шести признаков [3].

Эти три компоненты используются как входные признаки последующих моделей. Рисунки 2–4 показывают двумерную проекцию PC1-PC2 для наглядного сопоставления результатов, тогда как рисунки 5–7 используют трёхмерное пространство PC1-PC2-PC3 и тем самым учитывают все три выбранные компоненты.

4. Классификация с помощью метода опорных векторов SVM

Метод опорных векторов (Support Vector Machine, SVM) применяется в данном исследовании для классификации режимов течения по уровню газосодержания. В отличие от One-Class SVM, используемого для обнаружения аномалий, классический SVM решает задачу многоклассовой классификации, разделяя данные на три класса: низкое, среднее и высокое газосодержание. Результаты PCA и доли объяснённой дисперсии для главных компонент представлены в таблице 1.

Входными признаками для классификатора служат три главные компоненты (PC1, PC2, PC3), полученные в результате PCA-преобразования.

Таблица 1. Дисперсия, объясняемая главными компонентами

Главная компонента	Объясненная дисперсия (%)	Кумулятивная дисперсия (%)
PC1	52,15	52,15
PC2	27,51	79,66
PC3	10,20	89,86
PC4	6,14	96,00
PC5	2,73	98,73
PC6	1,27	100,0

Использование ядра RBF (Radial Basis Function) позволяет строить нелинейные разделяющие границы. В представленном эксперименте использованы значения параметра регуляризации $C = 100,0$ и параметра RBF-ядра $\gamma = 1,0$. Формула 2 задаёт решающую функцию SVM, применяемую после обучения модели.

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b, \quad (2)$$

где n – количество данных в обучающей выборке; α_i – множители Лагранжа, полученные во время обучения (ненулевые только для опорных векторов); y_i – истинная метка класса (± 1) для i -го наблюдения; x_i – обучающие данные; $K(x_i, x)$ – функция ядра (например, линейная, RBF), которая вычисляет сходство между x_i и новой точкой x без явного вычисления многомерных координат.

В формуле 2 значение $f(x)$ является решающим значением классификатора для новой точки x и вычисляется по опорным векторам, их коэффициентам и выбранной функции ядра. Параметр $C = 100,0$ используется на этапе обучения как параметр регуляризации; представленные метрики относятся к указанной конфигурации модели.

Обучение SVM производится на 70% данных, оставшиеся 30% используются для тестирования. Стратификация выборки обеспечивает сохранение пропорций классов в обучающей и тестовой выборках.

Оценка качества классификации выполнена на тестовой части выборки. Полученная точность составила 98,26%, а значения F1-меры для классов низкого, среднего и высокого газосодержания находились в диапазоне 0,97–0,99.

Полученные показатели относятся к одному стратифицированному разбиению 70/30 и к указанной конфигурации SVM с тремя главными компонентами и RBF-ядром. Дополнительная оценка выполнена с помощью 5-кратной стратифицированной кросс-валидации (табл. 2).

Первоначально использованы параметры $C = 10,0$, $\gamma = \text{auto}$, что обеспечило точность 98,42% на тестовой выборке.

Для обоснованного выбора гиперпараметров SVC выполнен сеточный поиск (Grid Search) с 5-кратной стратифицированной кросс-валидацией. Сетка параметров охватывала $C \in \{0,1, 1,0, 10,0, 100,0\}$ и $\gamma \in \{\text{scale}, \text{auto}, 0,01, 0,1, 1,0\}$ при фиксированном RBF-ядре. Целевая метрика оптимизации – F1_macro. Лучшая комбинация: $C = 100,0$, $\gamma = 1,0$, ядро RBF, обеспечившая F1_macro = 0,9853 на кросс-валидации.

Таблица 2. Результаты 5-кратной кросс-валидации SVM-классификатора

Метрика	Среднее	Стандартное отклонение
Accuracy	0,9867	0,0025
Precision_macro	0,9853	0,0024
Recall_macro	0,9854	0,0029
F1_macro	0,9853	0,0026

Средние значения и стандартные отклонения характеризуют вариацию метрик по пяти разбиениям. Выбор гиперпараметров выполнялся по F1_macro, поэтому он не равнозначен максимизации точности на тестовой части. Эти оценки не заменяют матрицу ошибок и независимую проверку на новых периодах наблюдений. Для проверки отсутствия утечки данных необходимо также фиксировать, что масштабирование признаков и PCA обучаются только на обучающей части каждого разбиения.

5. Метод DBSCAN

Для обнаружения локальных аномалий в данных мониторинга нефтяных скважин применён алгоритм пространственной кластеризации на основе плотности DBSCAN. Основное преимущество данного метода заключается в способности выявлять кластеры произвольной формы и отсутствии необходимости заранее задавать количество кластеров. Ключевыми параметрами метода являются радиус окрестности ε и минимальное количество точек $minPts$ [8]. В ходе исследования использованы значения $\varepsilon = 0,5$ и $minPts = 5$. Входными данными служат три главные компоненты, полученные в результате преобразования методом главных компонент.

В результате применения алгоритма DBSCAN выявлено 188 аномальных наблюдений, что составляет 2,18% от общего объёма данных. Данный результат показывает, что DBSCAN позволяет выявлять локальные выбросы, физическая природа которых требует проверки по исходным измерениям и эксплуатационным данным. В отличие от метода опорных векторов, который ориентирован на глобальные аномалии, DBSCAN дополняет статистическое описание данных, обнаруживая точки, находящиеся в областях с низкой плотностью данных [9]. Само по себе число срабатываний не подтверждает повышение надежности обнаружения реальных нарушений.

Для трёхмерного пространства PC1-PC3 выбор $minPts = 5$ удовлетворяет эвристике $minPts \geq dim + 1 = 4$. Построен график расстояний до k -го ближайшего соседа ($k = minPts - 1 = 4$) для точек в пространстве главных компонент (рис. 1). Точка максимальной кривизны, определяемая по второй производной функции k -расстояний, соответствует значению $\varepsilon \approx 16,17$.

Полученное значение не совпадает с использованным $\varepsilon = 0,5$, поэтому данный график не может служить подтверждением оптимальности $\varepsilon = 0,5$. Условие для $minPts$ также не определяет единственное оптимальное значение. Настройка $\varepsilon = 0,5$, $minPts = 5$ рассматривается как исследуемая конфигурация.

Для оценки устойчивости результатов к выбору гиперпараметров проведён анализ чувствительности. При варьировании ε от 0,3 до 1,0 число выявленных выбросов изменялось в диапазоне [24, 558], при этом число

кластеров уменьшалось с 39 до 4.

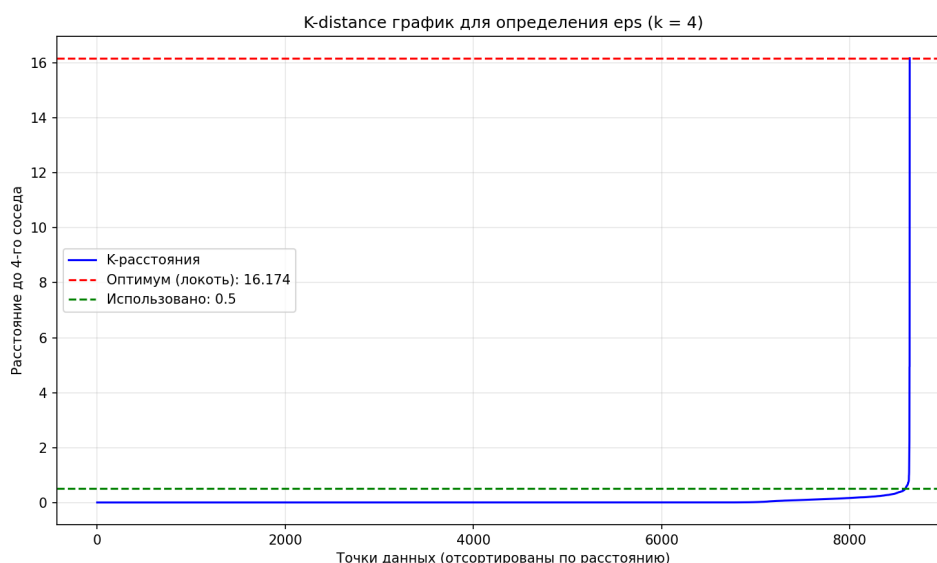


Рис. 1. График k -расстояний для DBSCAN ($k = 4$) и сравнение вычисленной точки максимальной кривизны с использованным $\varepsilon = 0,5$.

Таблица 3. Чувствительность DBSCAN к ε ($minPts = 5$)

ε	Выбросов	Доля, %	Кластеров
0,3	558	6,46	39
0,4	340	3,94	26
0,5	188	2,18	19
0,6	112	1,30	13
0,7	82	0,95	12
0,8	58	0,67	10
1,0	24	0,28	4

Таблица 4. Чувствительность DBSCAN к $minPts$ ($\varepsilon = 0,5$)

$minPts$	Выбросов	Доля, %	Кластеров
3	122	1,41	24
5	188	2,18	19
7	278	3,22	16
10	317	3,67	13
15	420	4,86	15
20	546	6,32	12

При уменьшении ε с 0,5 до 0,4 число выбросов возрастает со 188 до 340, а при увеличении до 0,6 сокращается до 112. Изменение $minPts$ с 5 до 3 или 7 приводит соответственно к 122 или 278 срабатываниям. Такая вариация не дает оснований считать результат устойчивым только на основе близости используемых настроек. При эксплуатации необходима проверка

качества срабатываний, а не только их числа.

6. Сравнение методов обнаружения аномалий SVM и DBSCAN

В ходе исследования проведён сравнительный анализ двух методов обнаружения аномалий – One-Class SVM и DBSCAN. Метод опорных векторов выявил 424 аномалии, тогда как DBSCAN обнаружил 188 аномалий. Пересечение результатов составило 85 образцов, или 16,13% от объединения срабатываний (527 уникальных наблюдений) [10]. Низкая степень пересечения (16,13%) допускает две интерпретации: методы обнаруживают различные типы аномалий; низкая согласованность может указывать на недостаточную надёжность хотя бы одного из методов в данной предметной области. Разделение срабатываний на категории не разрешает эту неоднозначность без независимой эксплуатационной валидации.

Для оценки устойчивости результатов One-Class SVM проведён анализ чувствительности к параметрам ν и γ . При варьировании $\nu \in \{0,01, 0,03, 0,05, 0,07, 0,10, 0,15\}$ число выбросов изменялось в диапазоне [88, 1233].

При $\nu = 0,05$ выявлено 424 наблюдения (4,91%). При этом значении ν в дополнительном анализе сравнены варианты $\gamma \in \{\text{scale}, \text{auto}, 0,001, 0,01, 0,1, 1,0\}$; для scale и auto получено по 424 срабатывания. При варьировании γ диапазон выбросов составил [424, 473], что свидетельствует о существенно меньшей чувствительности результатов к параметру γ по сравнению с ν .

Таблица 5. Чувствительность One-Class SVM к ν

ν	Выбросов	Доля, %
0,01	88	1,02
0,03	295	3,41
0,05	424	4,91
0,07	579	6,70
0,10	933	10,80
0,15	1233	14,27

Таблица 6. Чувствительность One-Class SVM к γ ($\nu = 0,05$)

γ	Выбросов	Доля, %
scale	424	4,91
auto	424	4,91
0,001	456	5,28
0,01	439	5,08
0,1	437	5,06
1,0	473	5,48

6.1. Сравнение с базовым пороговым методом

Выполнено сравнение с правилом 3σ (табл. 7). Пороговый метод классифицирует наблюдение как аномальное, если хотя бы один из шести признаков выходит за пределы $\mu \pm 3\sigma$.

Таблица 7. Сравнение методов обнаружения аномалий

Метод	Обнаружено	Доля, %
Пороговый (правило 3σ)	88	1,02
One-Class SVM	424	4,91
DBSCAN	188	2,18
Пересечение ($SVM \cap DBSCAN$)	85	0,98

Согласованность порогового метода с One-Class SVM составила 71 образец (16,7% от результатов SVM), с DBSCAN – 55 образцов (29,3% от результатов DBSCAN).

Сравнение показывает различия в числе и составе срабатываний, но не устанавливает превосходство одного метода над другим. Большее число отмеченных наблюдений может включать как дополнительные полезные обнаружения, так и ложные тревоги.

7. Классификация выбросов

На основе сочетания срабатываний двух методов предложена следующая классификация результатов (табл. 8).

Таблица 8. Классификация результатов по сочетанию SVM и DBSCAN

Результат совместной классификации	Количество	Критерий отнесения
Согласованные выбросы, приоритетная проверка	85	Выявлены обоими методами
Потенциальные аномалии, экспертная оценка	339	Выявлены только One-Class SVM
Локальные выбросы, отдельная проверка	103	Выявлены только DBSCAN

8. Практические рекомендации

Результаты работы методов представлены на рисунках 2–7. На основе полученных результатов предложена следующая стратегия использования методов:

1. **Пересечение** (85 образцов) – точки, выявленные обоими методами. Их следует рассматривать как приоритетные кандидаты для экспертной и, при наличии эксплуатационных данных, технической проверки, а не как автоматически подтверждённые аварийные состояния. Сопоставление результатов двух методов представлено на рисунке 4 в проекции PC1-PC2 и на рисунке 7 в пространстве PC1-PC2-PC3.
2. **Только SVM** (339 образцов) – точки, выявленные только One-Class SVM. Для них рекомендуется отдельная экспертная оценка и сопоставление

с временным контекстом мониторинга. Результаты One-Class SVM представлены на рисунке 2 в проекции PC1-PC2 и на рисунке 5 в пространстве PC1-PC2-PC3.

3. **Только DBSCAN** (103 образца) – точки, выявленные только DBSCAN. Их следует рассматривать как локальные статистические выбросы или кандидаты на шум и не исключать автоматически без дополнительной проверки. Результаты DBSCAN представлены на рисунке 3 в проекции PC1-PC2 и на рисунке 6 в пространстве PC1-PC2-PC3.

Для последующего анализа категории «Пересечение», «Только SVM» и «Только DBSCAN» следует сохранять отдельно. Без независимой эксплуатационной разметки эти категории отражают сочетание срабатываний алгоритмов. Физическая интерпретация точки требует сопоставления координат PC1-PC3 с соответствующими исходными показаниями и временем регистрации. Анализ скорости звука, температуры и газосодержания в центре и на периферии потока совместно с эксплуатационными журналами должен предшествовать выводу о неисправности оборудования. Отсутствие независимой событийной разметки не позволяет связывать 85 согласованных срабатываний с технологическими нарушениями.

9. Выводы по результатам экспериментального исследования

Эксперимент показал, что PCA сократил размерность с шести до трёх главных компонент при сохранении 89,86% общей дисперсии. На выборке из 8639 наблюдений One-Class SVM выявил 424 аномальные точки, DBSCAN – 188; общими для двух методов оказались 85 наблюдений, что составляет 16,13% от объединения срабатываний двух методов (527 уникальных точек).

SVM-классификация достигла точности 98,26% при F1-мере 0,97–0,99. В практическом мониторинге пересечение методов следует считать приоритетными выбросами, результаты «только SVM» направлять на экспертную проверку, а точки «только DBSCAN» рассматривать как локальные статистические отклонения, требующие отдельной проверки. При 5-кратной стратифицированной кросс-валидации получены $accuracy = 0,9867 \pm 0,0025$ и $F1_macro = 0,9853 \pm 0,0026$. Пороговый метод по правилу 3σ выявил 88 наблюдений (1,02%).

Заключение

Применение PCA позволило сократить размерность данных с 6 до 3 компонент с сохранением 89,86% информации. One-Class SVM выявил 424 аномалии, DBSCAN – 188, из них 85 обнаружены обоими методами [11]. Выделены три категории: согласованные выбросы – 85 наблюдений, только SVM – 339, только DBSCAN – 103. Классификация режимов течения на основе SVM показала точность 98,26% при F1-мере 0,97–0,99 для трёх классов.

Научная новизна состоит в совместном анализе One-Class SVM и DBSCAN на данных мониторинга нефтяной скважины и исследовании чувствительности результатов к гиперпараметрам. Практическая значимость связана с возможностью применения подхода в системах мониторинга, раннего предупреждения и прогнозного обслуживания.

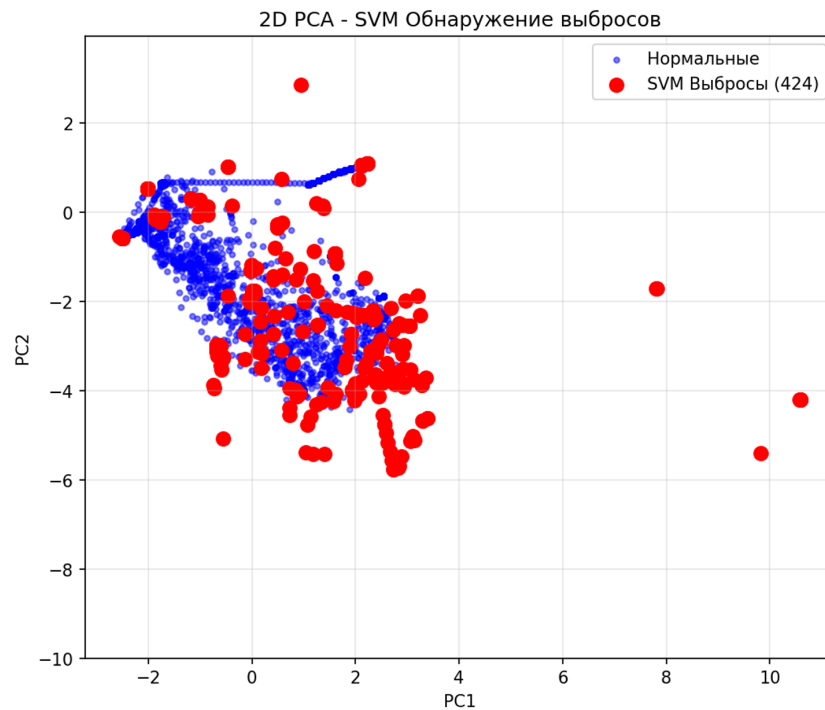


Рис. 2. Обнаружение выбросов методом One-Class SVM в пространстве PC1-PC2

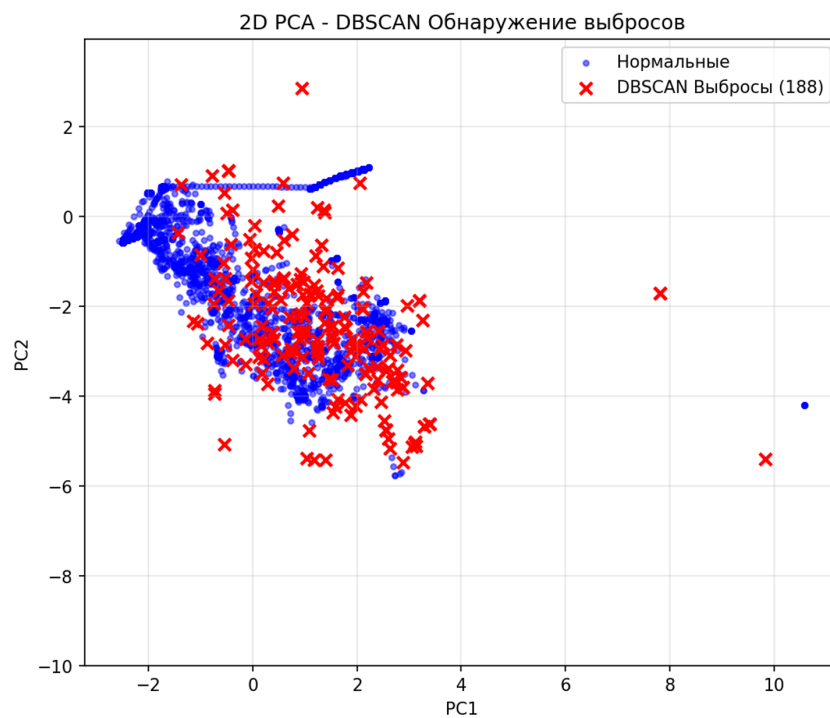


Рис. 3. Обнаружение выбросов методом DBSCAN в пространстве PC1-PC2

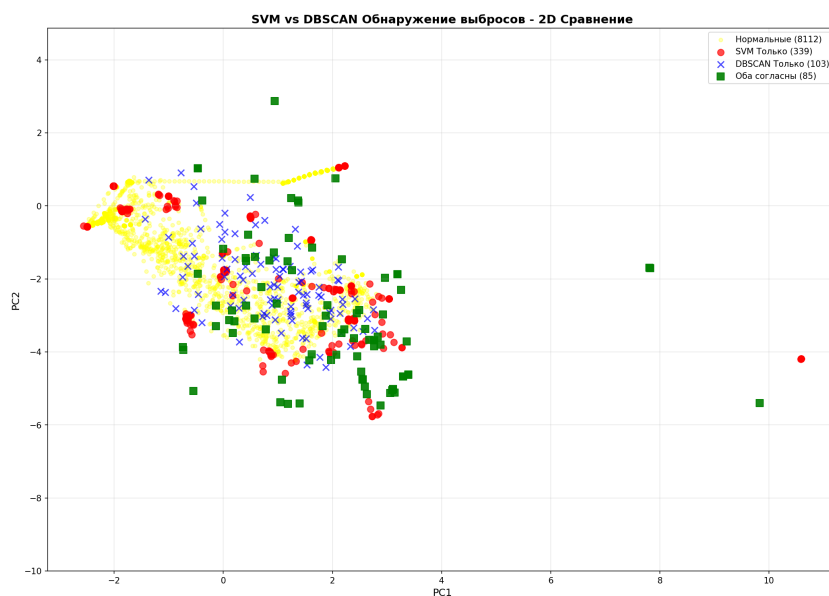


Рис. 4. Сравнение результатов обнаружения выбросов методами SVM и DBSCAN в пространстве PC1-PC2

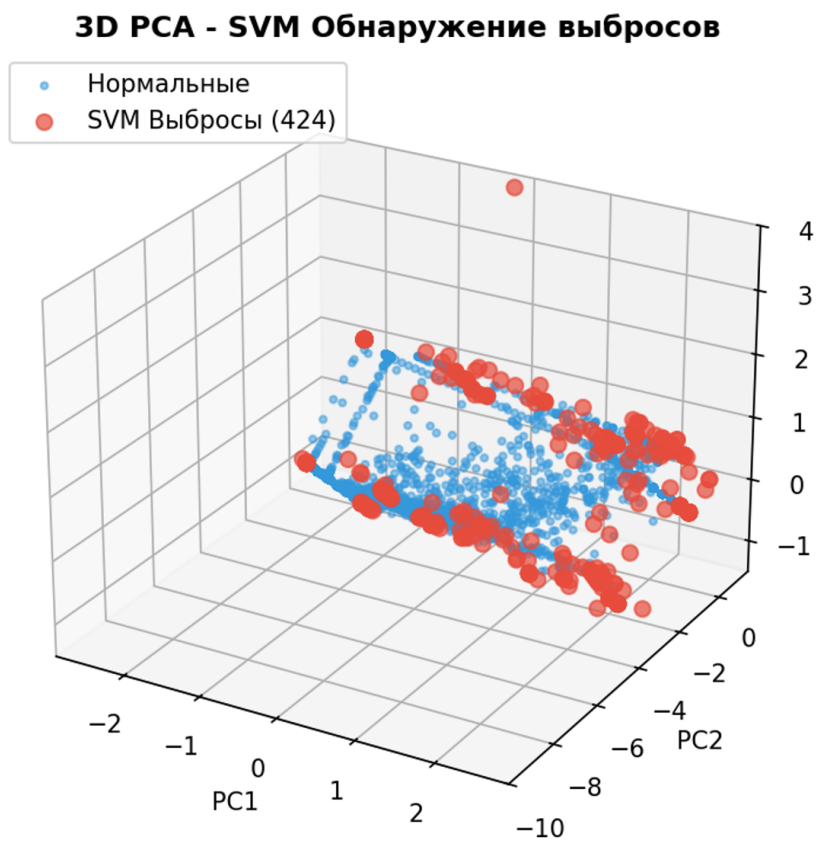


Рис. 5. Трёхмерная визуализация аномалий, обнаруженных методом One-Class SVM

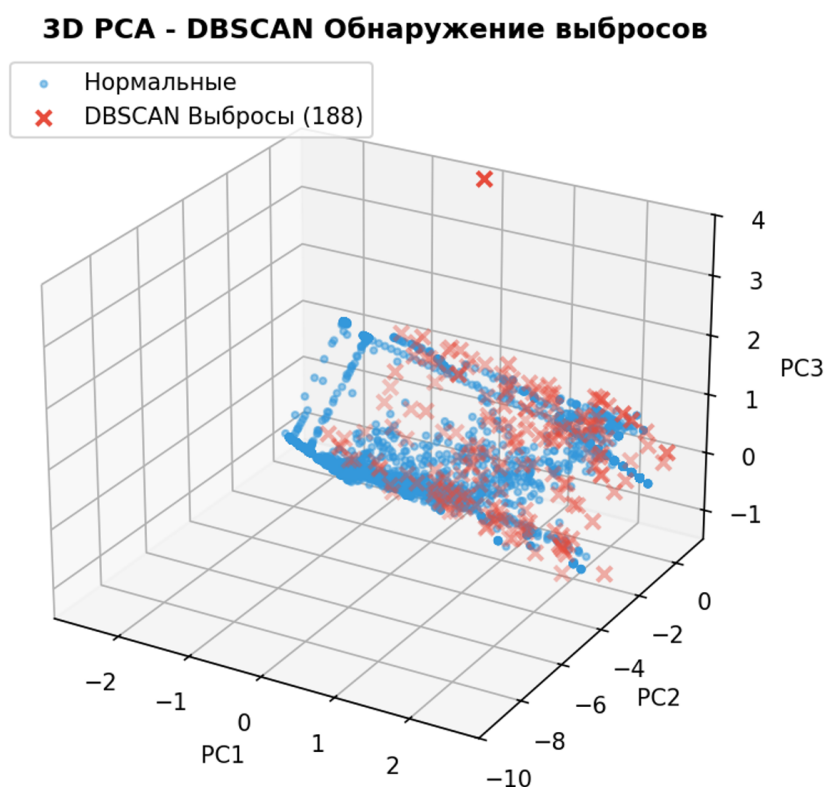


Рис. 6. Трёхмерная визуализация аномалий, обнаруженных методом DBSCAN

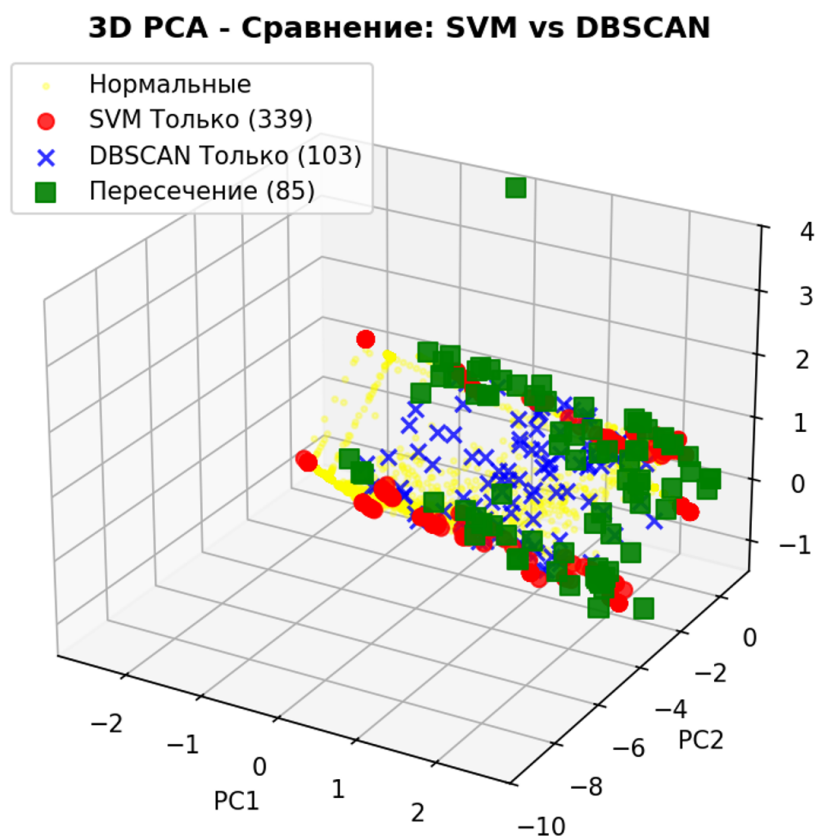


Рис. 7. Сравнение результатов обнаружения выбросов методами SVM и DBSCAN в трёхмерном пространстве главных компонент (PC1-PC2-PC3)

Дальнейшие исследования должны включать проверку подхода на других месторождениях, автоматическую настройку параметров и сопоставление выявленных выбросов с реальными инцидентами и данными технической диагностики.

Благодарности и финансирование

Исследование было проведено при поддержке РФФИ (проект № 24-11-00285).

The study was conducted with the support of the Russian Science Foundation (project No. 24-11-00285).

Список литературы

- [1] Ebrahimi, M. Anomaly detection in oil-producing wells: a comparative study of one-class classifiers in a multivariate time series dataset / M. Ebrahimi, E. Khamenechi, M. Ameri // *Journal of Petroleum Exploration and Production Technology*. – 2023. – Vol. 13. – P. 1845–1860.
- [2] Mohammadpour, A. Outlier detection and selection of representative fluid samples using machine learning: a case study of Iranian oil fields / A. Mohammadpour, D. Shahsavani, M. Riazi // *Journal of Petroleum Exploration and Production Technology*. – 2024. – Vol. 14, No. 10. – P. 2833–2852.
- [3] Liu, Y. Identification Method of Tubing Leakage Based on PCA-SVM in Gas Well / Y. Liu, H. Wang, Z. Chen // *Proceedings of the 2nd International Conference on Signal Processing and Intelligent Computing (SPIC)*. – 2024. – P. 1–6.
- [4] Wang, L. Identification method for true and false anomalies of oil pipeline leakage in an old industrial oil production well site / L. Wang, X. Zhang, J. Liu // *Oil & Gas Storage and Transportation*. – 2023. – Vol. 42, No. 5. – P. 567–575.
- [5] Comparison of Isolation Forest and One Class SVM in Anomaly Detection of Gas Pipeline Operation // *Proceedings of the 2023 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. – IEEE, 2024. – (IEEE Xplore, 13 February 2024).
- [6] Reconstruction of well logging data using unsupervised machine learning based outlier detection techniques (UML ODTs) under adverse drilling conditions / J.-R. Chen, R.-Z. Yang, T.-T. Li [et al.] // *Applied Geophysics*. – 2025. – DOI 10.1007/s11770-025-1178-z.
- [7] Lundberg, S. M. A Unified Approach to Interpreting Model Predictions / S. M. Lundberg, S.-I. Lee // *Advances in Neural Information Processing Systems*. – 2017. – Vol. 30. – P. 4765–4774.
- [8] Ahmadi, M. Data-Driven Modeling of Oil Saturation from Dielectric Logs Using Ensemble Regression, Dimensionality Reduction and Anomaly Detection Machine Learning Algorithms / M. Ahmadi, M. Bagheri, A. Sadeghi // *SPE Annual Technical Conference and Exhibition*. – 2023. – Paper SPE-215078-MS.
- [9] Patel, D. Oil and gas flow anomaly detection on offshore naturally flowing wells using deep neural networks / D. Patel, M. Shah, K. Patel // *Journal of Petroleum Science and Engineering*. – 2024. – Vol. 234. – Art. 109876.
- [10] Hassan, N. Unsupervised machine learning model for predicting anomalies in subsurface safety valves and application in offshore wells during oil production / N. Hassan, A. El-Sayed, S. Ibrahim // *Journal of Petroleum Exploration and Production Technology*. – 2023. – Vol. 13. – P. 2211–2225.
- [11] Zhao, Q. Varying-scale HCA-DBSCAN-based anomaly detection method for multi-dimensional energy data in steel industry / Q. Zhao, M. Li, Y. Sun // *Information Sciences*. – 2023. – Vol. 645. – Art. 119479.

COMPARING SVM AND DBSCAN FOR OUTLIER DETECTION IN OIL WELL MONITORING DATA*

Mikhailov I.S.^{1,a}, Myo Hlaing Win^{1,b}, Sidorov K.O.^{1,c}

¹ National Research University «Moscow Power Engineering Institute» (NRU MPEI), Moscow, Russia.

E-mail: ^a fr82@mail.ru, ^b myohlaingwin69287@gmail.com, ^c kirill.sidoroff2014@yandex.ru

Abstract. The aim of this study is to compare One-Class Support Vector Machine (One-Class SVM) and DBSCAN for detecting statistical anomalies in oil-well monitoring data and to assess their joint use. Six parameters from a multiphase water-cut meter were analyzed; Principal Component Analysis (PCA) reduced them to three principal components preserving 89.86% of the total variance. In a sample of 8639 observations, One-Class SVM identified 424 anomalous points and DBSCAN identified 188, with 85 observations detected by both methods. For flow-regime classification by gas-content level, SVM with an RBF kernel ($C = 100.0$, $\gamma = 1.0$) achieved 98.26% accuracy with an F1 score of 0.97–0.99 on a stratified 70/30 split. Five-fold stratified cross-validation yielded an F1_macro of 0.9853 ± 0.0026 . Joint interpretation separates points detected by both methods from SVM-only and DBSCAN-only detections for subsequent expert review.

Keywords: outlier detection, Support Vector Machine, DBSCAN, Principal Component Analysis, oil well monitoring, machine learning, flow regime classification, anomalies.

References

- [1] Ebrahimi, M., Khamsehchi, E., & Ameri, M. (2023). Anomaly detection in oil-producing wells: A comparative study of one-class classifiers in a multivariate time series dataset. *Journal of Petroleum Exploration and Production Technology*, 13, 1845–1860.
- [2] Mohammadpour, A., Shahsavani, D., & Riazi, M. (2024). Outlier detection and selection of representative fluid samples using machine learning: A case study of Iranian oil fields. *Journal of Petroleum Exploration and Production Technology*, 14(10), 2833–2852.
- [3] Liu, Y., Wang, H., & Chen, Z. (2024). Identification method of tubing leakage based on PCA-SVM in gas well. In *Proceedings of the 2nd International Conference on Signal Processing and Intelligent Computing (SPIC)* (pp. 1–6).
- [4] Wang, L., Zhang, X., & Liu, J. (2023). Identification method for true and false anomalies of oil pipeline leakage in an old industrial oil production well site. *Oil & Gas Storage and Transportation*, 42(5), 567–575.
- [5] Comparison of isolation forest and one class SVM in anomaly detection of gas pipeline operation. (2024). In *Proceedings of the 2023 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. IEEE.

For citation:

Mikhailov I.S., Myo Hlaing Win, Sidorov K.O. Comparing SVM and DBSCAN for Outlier Detection in Oil Well Monitoring Data. *Intelligent transport*, 2026, no. 3(39), pp. 53–68. DOI: 10.24412/3033-6007-2026-339-53-68. (in Russian)

*The study was conducted with the support of the Russian Science Foundation (project No. 24-11-00285).

- [6] Chen, J.-R., Yang, R.-Z., Li, T.-T., Xu, Y.-D., & Sun, Z.-P. (2025). Reconstruction of well logging data using unsupervised machine learning based outlier detection techniques (UML ODTs) under adverse drilling conditions. *Applied Geophysics*. <https://doi.org/10.1007/s11770-025-1178-z>
- [7] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- [8] Ahmadi, M., Bagheri, M., & Sadeghi, A. (2023). Data-driven modeling of oil saturation from dielectric logs using ensemble regression, dimensionality reduction and anomaly detection machine learning algorithms. In *SPE Annual Technical Conference and Exhibition* (Paper SPE-215078-MS).
- [9] Patel, D., Shah, M., & Patel, K. (2024). Oil and gas flow anomaly detection on offshore naturally flowing wells using deep neural networks. *Journal of Petroleum Science and Engineering*, 234, Article 109876.
- [10] Hassan, N., El-Sayed, A., & Ibrahim, S. (2023). Unsupervised machine learning model for predicting anomalies in subsurface safety valves and application in offshore wells during oil production. *Journal of Petroleum Exploration and Production Technology*, 13, 2211–2225.
- [11] Zhao, Q., Li, M., & Sun, Y. (2023). Varying-scale HCA-DBSCAN-based anomaly detection method for multi-dimensional energy data in steel industry. *Information Sciences*, 645, Article 119479.

ИНТЕРПРЕТИРУЕМОЕ УПРАВЛЕНИЕ ИНИЦИАЦИЕЙ ТРЕКОВ С ЗАРАНЕЕ ЗАДАННЫМИ ГРАНИЦАМИ ВМЕШАТЕЛЬСТВА В СИСТЕМАХ ВОСПРИЯТИЯ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ

Трофимов Ю.В.^{1,а}, Аверкин А.Н.^{1,2,б}, Шевченко А.В.^{1,с}, Кузнецов Е.М.^{1,д},
Лебедев А.Д.^{1,е}

¹ Государственный университет «Дубна», Дубна, Россия.

² Федеральный исследовательский центр «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), Москва, Россия.

E-mail: ^а ura_trofim@bk.ru, ^б averkin2003@inbox.ru, ^с leviathan0909@gmail.com, ^д kem.22@uni-dubna.ru,

^е lad.24@uni-dubna.ru

Аннотация. В системах воздушного наблюдения единичная ошибка обнаружения может закрепиться в виде устойчивого ложного трека и повлиять на данные, передаваемые последующим компонентам. Цель работы – уменьшить число таких ошибок, заранее ограничив масштаб вмешательства и сохранив возможность проследить логику каждого решения. Предложен внешний метод кратковременной задержки эпизодов инициации треков с низкой поддержкой. Число задержек регулируется накопительным бюджетом, а для каждого эпизода сохраняются оценка поддержки, текущий баланс бюджета, причина вмешательства и конечный статус. Метод апробирован на семи видеопоследовательностях из общедоступного набора аэросъемки, содержащего малые объекты, частичные перекрытия и заметные изменения масштаба. В эксперименте сравнивались три алгоритма сопровождения и обучаемый метод на основе ансамбля решающих деревьев. Для ByteTrack задержка была применена к 2,62% эпизодов инициации, из оперативного выхода исключены 425 ложных наблюдений и 176 ложных инициаций, а изменение объединённого показателя точности и полноты осталось в пределах установленного допуска. Для OC-SORT и SORT снижение ложных наблюдений по принятому статистическому критерию не подтверждено. Проверка 29 385 начальных фрагментов видеозаписей не выявила нарушений заранее заданных ограничений. Средние дополнительные вычислительные затраты составили 0,901 миллисекунды на кадр. Результаты показывают, что кратковременная задержка может сократить число ложных наблюдений при контролируемом масштабе вмешательства.

Ключевые слова: интеллектуальные транспортные системы, беспилотные летательные аппараты, сопровождение объектов, инициация треков, ограниченное вмешательство, интерпретируемый искусственный интеллект, кратковременная задержка.

Введение

Беспилотные аппараты применяют для наблюдения за дорогами, инспекции инфраструктуры и контроля логистических объектов. Проблема с ложным

Для цитирования:

Трофимов Ю.В., Аверкин А.Н., Шевченко А.В., Кузнецов Е.М., Лебедев А.Д. Интерпретируемое управление инициацией треков с заранее заданными границами вмешательства в системах восприятия беспилотных летательных аппаратов. *Интеллектуальный транспорт*. 2026. № 3(39). С. 69–84. DOI: 10.24412/3033-6007-2026-339-69-84.

обнаружением возникает, когда алгоритм сопровождения присваивает ему идентификатор и начинает вести его как самостоятельный объект. После этого ошибочное обнаружение уже входит в состав трека, попадает в оперативный выход системы и может быть передано модулю планирования. Обычный способ уменьшить число ложных треков – это повысить порог уверенности или потребовать повторного появления объекта, однако цена такого решения меняется вместе со сценой. Высота съёмки, освещение, масштаб и плотность объектов сдвигают распределение оценок, поэтому правило, настроенное на одной видеозаписи, в другой может задержать заметно больше настоящих целей. В работе предлагается принцип, где дополнительный слой ненадолго задерживает только небольшую долю новых эпизодов, а все остальные сразу передаёт исходному алгоритму сопровождения. Если эпизод продолжает наблюдаться, но подтверждений всё ещё недостаточно, задержка снимается и дальнейшую обработку продолжает исходный алгоритм сопровождения. Накопительный бюджет вмешательств ограничивает число задержек. Его баланс задаёт верхнюю границу как для количества вмешательств, так и для возможного отличия оперативного выхода от результата исходного алгоритма. Одновременно сохраняется возможность проследить каждое решение, так как журнал хранит поддержку эпизода, состояние бюджета, причину задержки и конечный статус.

Цель исследования – выяснить, позволяет ли такое правило уменьшить число ложных наблюдений в оперативном выходе, не выходя за заранее установленный допуск по качеству. Метод проверяется с тремя алгоритмами сопровождения и сопоставляется со случайным лесом, обученным в той же схеме внешней проверки. Отдельно оцениваются соблюдение формальных ограничений, прослеживаемость решений и дополнительные вычислительные затраты.

1. Обзор существующих работ

Алгоритмы сопровождения связывают обнаружения одного и того же объекта на последовательных кадрах, формируя его траекторию. Они прогнозируют возможное положение объекта, затем сопоставляют прогноз с новыми обнаружениями. Классической основой такого прогноза служит фильтр Калмана [1]. В SORT он объединён с взаимно однозначным сопоставлением рамок на соседних кадрах [2]. Подход хорошо работает при плавном движении, но остаётся чувствительным к пропускам и ложным обнаружениям. В более новых методах устойчивость сопровождения повышается разными способами. ByteTrack повторно рассматривает обнаружения с невысокой уверенностью и тем самым восстанавливает часть прерванных траекторий [3]. OC-SORT сильнее опирается на фактически наблюдаемое движение, а StrongSORT дополняет геометрию признаками внешнего вида [4, 5].

В процедуре сопровождения применяются и обучаемые компоненты. DeepSORT различает близко расположенные объекты по признакам внешнего вида, Tracktor переносит рамку между кадрами средствами регрессии детектора, а случайный лес объединяет решения множества деревьев [6–8]. Обучаемые модели требуют размеченных данных для настройки под конкретную задачу; при переносе на другую камеру или сцену их качество может измениться, а адаптация к новым условиям может потребовать дополнитель-

ных размеченных примеров и повторного обучения. Отдельный подход к работе с неуверенными решениями используется в селективной классификации, где модель отказывается от ответа на неуверенных примерах [9]. В предлагаемой схеме вместо полного отказа используется кратковременная задержка нового эпизода. Если эпизод продолжает наблюдаться, но уверенности по-прежнему недостаточно, его дальнейшая обработка передаётся исходному алгоритму сопровождения. Фиксированный порог уверенности и требование нескольких повторных наблюдений также ограничивают раннюю передачу слабых эпизодов. При большом числе слабых обнаружений фиксированное правило может затронуть почти все новые эпизоды. Предлагаемый метод отличается способом ограничения вмешательства.

2. Постановка задачи и критерии оценки

Задача состоит в построении внешнего правила обработки новых эпизодов треков, которое позволяет уменьшить число ложных наблюдений в оперативном выходе, не допуская заранее неограниченного вмешательства в работу исходного алгоритма сопровождения. Для каждого нового эпизода правило должно определить, передать его наблюдения без задержки либо временно удержать их, причём число таких вмешательств и возможное отличие оперативного выхода от исходного результата ограничиваются заданным бюджетом. На каждом кадре детектор формирует обнаружения, а алгоритм сопровождения связывает их с существующими треками или присваивает им новые идентификаторы. Наблюдением считается рамка обнаруженного объекта на одном кадре вместе с идентификатором трека и оценкой уверенности. Непрерывная последовательность наблюдений с одним идентификатором образует эпизод трека. Если после длительного перерыва тот же идентификатор появляется снова, дополнительный слой рассматривает его как новый эпизод. Вмешательством называется временная задержка первых наблюдений такого эпизода перед их передачей последующим компонентам. В оценке различаются оперативный выход и итоговая карта. Оперативный выход содержит только наблюдения, доступные последующим компонентам в момент обработки каждого кадра. Итоговая карта формируется после завершения видеозаписи. Если задержанный эпизод подтверждён, его ранние наблюдения можно восстановить. Основная оценка проводится по оперативному выходу.

После применения дополнительного слоя все наблюдения, переданные в оперативный выход, заново сопоставляются с разметкой. Это необходимо, так как исключение или задержка одного наблюдения может изменить соответствия для остальных наблюдений. Повторное сопоставление исключает двойной учёт. Рамка связывается с размеченным объектом при отношении площади пересечения к площади объединения не менее 0,5. Исключённые области VisDrone [10] удаляются после сопоставления. Ложной инициацией считается эпизод, который в течение полного окна решения не соответствует ни одному размеченному объекту. Само окно – это время, отведённое на наблюдение нового эпизода. Для сравнения видеозаписей разной длины число ложных наблюдений в оперативном выходе пересчитывается на 100 кадров по формуле (1).

$$O_{100} = 100 \cdot \frac{N_{\text{false}}}{N_{\text{frames}}}. \quad (1)$$

Здесь N_{false} обозначает число наблюдений эпизодов ложной инициации, переданных в оперативный выход, а N_{frames} соответствует числу кадров видеозаписи.

Качество оперативного выхода оценивается мерой F_1 , которая объединяет точность и полноту. Точность показывает долю наблюдений оперативного выхода, относящихся к истинным объектам. Полнота отражает долю размеченных объектов, сохранённых системой. Сначала проверяется, остаётся ли изменение качества в пределах установленного допуска. Нижняя граница 95-процентного доверительного интервала изменения F_1 должна быть выше значения $-0,01$. Это условие записывается формулой (2).

$$H_1 : L_{95}(\Delta F_1) > -0,01. \quad (2)$$

После выполнения первого условия проверяется снижение показателя ложных наблюдений. Верхняя граница 95-процентного доверительного интервала изменения показателя O_{100} должна быть ниже нуля. Условие проверки приведено в формуле (3).

$$H_2 : U_{95}(\Delta O_{100}) < 0. \quad (3)$$

Благодаря порядку, где второе условие проверяется только после первого, успешным нельзя признать метод, который удаляет ложные наблюдения ценой заметной потери настоящих объектов. Доверительные интервалы строятся по многократным парным выборкам из семи видеопоследовательностей. Во всех расчётах используется один и тот же набор повторов. Вместе с интервалами приводится выборочное стандартное отклонение. Эпизоды, начавшиеся слишком близко к концу записи, исключаются из основной проверки, поскольку для них недоступно полное окно решения.

3. Метод интерпретируемой кратковременной задержки эпизодов инициации треков

Последовательность обработки нового эпизода показана на рисунке 1. Сначала оцениваются его поддержка и текущий баланс бюджета. При низкой поддержке и достаточном балансе первые наблюдения эпизода временно задерживаются на срок не более заданного числа кадров. После завершения этого интервала эпизоду присваивается конечный статус.

Поддержка состоит из двух частей. Абсолютная поддержка показывает, насколько оценка уверенности превышает минимальное рабочее значение. Относительная поддержка характеризует положение нового эпизода относительно эпизодов, ранее появившихся в той же видеозаписи. Рассматриваемый эпизод в историю не включается, поэтому при расчёте используются только сведения, уже доступные к текущему моменту. Итоговая поддержка определяется формулой (4).

$$u_t = (1 - \eta)u_t^{\text{abs}} + \eta u_t^{\text{rel}}. \quad (4)$$

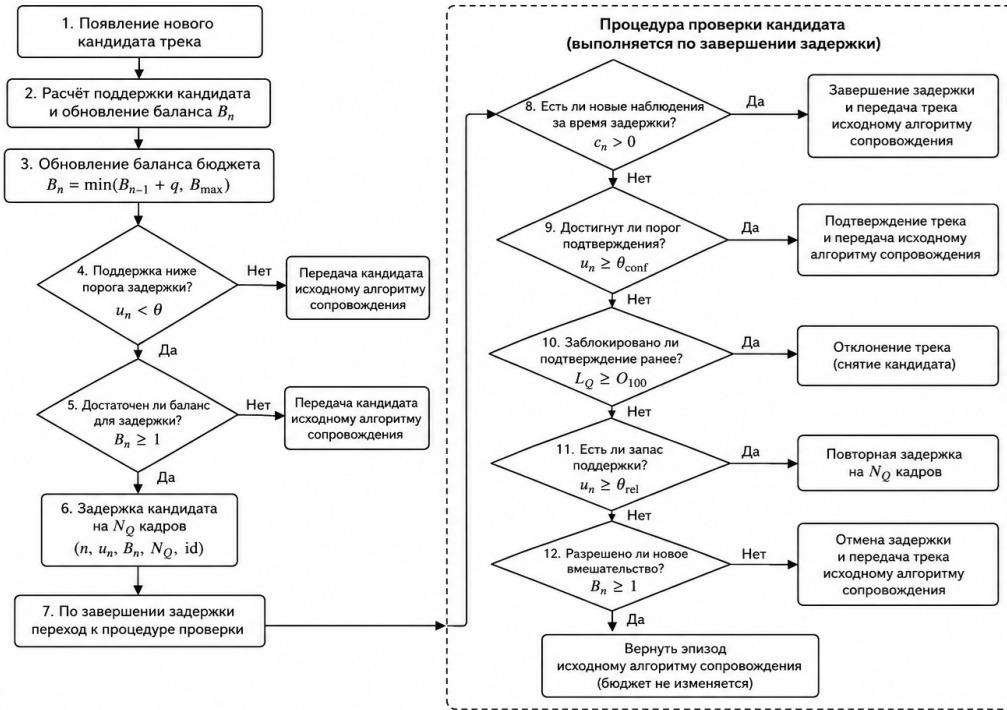


Рис. 1. Схема принятия решения при интерпретируемой кратковременной задержке эпизода инициации трека

Здесь u_t обозначает итоговую поддержку в момент времени t . Величины u_t^{abs} и u_t^{rel} соответствуют абсолютной и относительной поддержке. Параметр η определяет вклад относительной поддержки. Низкая поддержка лишь допускает задержку эпизода. Перед обработкой каждого эпизода инициации баланс пополняется на величину q_k , но не может превысить ёмкость C . Один случай задержки расходует одну единицу бюджета. Изменение баланса задаётся формулой (5).

$$\tilde{b}_k = \min(C, b_{k-1} + q_k), \quad b_k = \tilde{b}_k - a_k. \quad (5)$$

Величины b_{k-1} и b_k обозначают баланс до и после обработки эпизода. Параметр a_k равен единице, если эпизод временно задержан, и нулю в остальных случаях. Когда баланс достигает ёмкости бюджета, часть поступающего бюджета теряется. Эта потеря обозначается r_k . Суммирование изменений баланса даёт точное равенство и верхнюю границу числа вмешательств, приведённую в формуле (6).

$$N_Q(n) = b_0 + \sum_{k=1}^n q_k - \sum_{k=1}^n r_k - b_n \leq \lfloor b_0 + \sum_{k=1}^n q_k \rfloor \leq \lfloor b_0 + n q_{\max} \rfloor. \quad (6)$$

Здесь $N_Q(n)$ обозначает число случаев задержки среди первых n эпизодов, b_0 соответствует начальному балансу, а $q_{\max}$ задаёт наибольшее пополнение. Граница сохраняется при любой последовательности входных оценок. Фактическое число вмешательств может оказаться меньше, если эпизодов с низкой поддержкой недостаточно или часть бюджета теряется при заполненном накопителе. Эта граница ограничивает масштаб воздействия дополнительного слоя. Пока эпизод задержан, подтверждающие сведения накапливаются от

кадра к кадру. Повторные наблюдения увеличивают накопленную поддержку, старые наблюдения постепенно теряют вес, а пропуски уменьшают её значение. Правило накопления приведено в формуле (7).

$$H_t = \max \{0, \lambda^{\Delta t} H_{t-1} - \ln(1 - u_t + \varepsilon) - \mu(\Delta t - 1)\}, \quad P_t = 1 - \exp(-H_t). \quad (7)$$

Величина H_t обозначает накопленную поддержку. Параметр Δt соответствует числу кадров после предыдущего наблюдения, λ определяет скорость забывания, а μ задаёт штраф за пропуск. Значение P_t используется для подтверждения эпизода. Константа $\varepsilon = 10^{-6}$ предотвращает вычисление логарифма нуля. Когда накопленная поддержка достигает порога, эпизод подтверждается, а задержанные наблюдения возвращаются в итоговую карту. Если эпизод наблюдался не менее двух раз, но порог не достигнут, задержка снимается и дальнейшую обработку эпизода продолжает исходный алгоритм сопровождения. Окончательное отклонение допускается только при заранее заданном запрещающем признаке либо при отсутствии повторного наблюдения в полном окне решения. Если запись заканчивается раньше, эпизод помечается как цензурированный.

Один случай задержки временно скрывает ограниченное число наблюдений, поэтому число строк, которыми оперативный выход отличается от результата исходного алгоритма, также имеет верхнюю границу. Она определяется по формуле (8).

$$D(n) \leq L_Q N_Q(n) \leq L_Q \lfloor b_0 + \sum_{k=1}^n q_k \rfloor. \quad (8)$$

Здесь $D(n)$ обозначает число отличающихся строк, а L_Q соответствует наибольшему числу наблюдений одного эпизода, скрытых до завершения решения. Формула ограничивает возможный масштаб изменения выхода, но не определяет, являются ли задержанные наблюдения истинными или ложными. Параметры метода выбираются только по обучающим видеопоследовательностям. Сначала исключаются конфигурации, при которых изменение F_1 выходит за установленный допуск или ухудшается подтверждение истинных треков. Среди оставшихся выбирается вариант с наименьшим значением показателя ложных наблюдений. Если допустимой конфигурации нет, то дополнительная задержка не применяется и алгоритм работает в исходном режиме. Формальные свойства метода приведены в таблице 1.

Интерпретация отдельных решений является частью самого правила и не строится постфактум отдельной моделью. Для каждого нового эпизода сохраняются исходная уверенность, абсолютная и относительная поддержка, баланс бюджета до и после решения, причина немедленной передачи или задержки, накопленная поддержка и конечный статус. Все эти величины непосредственно участвуют в принятии решения. По журналу можно восстановить ход обработки эпизода и установить, почему его наблюдения были немедленно переданы в оперативный выход или задержаны, на каком основании эпизод был подтверждён либо отклонён и какой конечный статус ему присвоен. Запись также показывает, какой запас остаётся до формальной границы вмешательства. Такая интерпретация предназначена для инженерного разбора и обеспечивает прослеживаемость решения.

Таблица 1. Формальные свойства метода ограниченной задержки

Свойство	Формальная запись	Практический смысл
Ограниченное вмешательство	$N_Q(n) \leq \lfloor b_0 + \sum q_k \rfloor$	Число случаев задержки ограничено накопленным бюджетом
Ограничение изменения выхода	$D(n) \leq L_Q N_Q(n)$	Ограничено число временно скрытых наблюдений
Независимость от распределения оценок	Граница выполняется для любой последовательности оценок	Изменение оценок не нарушает установленный предел
Завершение решений	Каждый эпизод получает конечный статус	После обработки не остаётся незавершённых решений
Возврат неопределённого эпизода	Продолжающийся эпизод передаётся исходному алгоритму	Недостаточная поддержка не приравнивается к ложности
Сохранение исходного режима	При отсутствии допустимой конфигурации сохраняется исходный режим	Дополнительный слой не изменяет работу алгоритма без подтверждённой настройки

4. Организация эксперимента

Метод апробирован на семи видеопоследовательностях открытого набора VisDrone [10]. Набор содержит аэросъёмку с небольшими объектами, частичными перекрытиями и заметными изменениями масштаба. Во всех экспериментах использовались заранее сохранённые результаты детектора YOLOv8s, поэтому различия между режимами определялись правилами инициации треков. В качестве исходных алгоритмов сопровождения выбраны ByteTrack, OC-SORT и SORT. Все они получали один и тот же набор обнаружений. Всего обработано 2 846 кадров. Основные показатели рассчитывались по оперативному выходу – наблюдениям, доступным системе во время обработки записи. Эпизоды у конца видеозаписи исключались из основной проверки, если для них не хватало полного окна решения.

Для оценки использована схема последовательного исключения одной видеозаписи. На каждой итерации параметры выбирались по шести последовательностям, а седьмая применялась только для проверки, затем проверочная запись менялась. Таким образом, каждая последовательность один раз использовалась для итоговой оценки и не участвовала в соответствующей настройке метода. Случайный лес обучался заново внутри каждой итерации внешней проверки. Проверочная последовательность не использовалась ни при обучении, ни при выборе порога и настройке бюджета. Признаки нового эпизода не содержали сведений из будущих кадров. После применения каждого правила наблюдения оперативного выхода заново сопоставлялись с разметкой, что исключало искажение оценки после задержки отдельных наблюдений. Проверенные значения параметров и выбранные конфигурации для каждого алгоритма сопровождения приведены в таблице 2.

Таблица 2. Проверенные значения и выбранные параметры кратковременной задержки

Параметр	Проверенные значения	ByteTrack	OC-SORT и SORT
$q_{\max}$, предельное пополнение бюджета	0; 0,01; 0,03; 0,05	0,05 во всех проверках	0 в шести проверках; 0,05 в одной
C , ёмкость бюджета	2	2	2
T_Q , длительность задержки, кадров	1; 2	2	1
η , вес относительной поддержки	0,4; 0,6	0,6	0,4
Порог немедленной передачи	0,60; 0,65; 0,80	0,60	0,65 в шести проверках; 0,80 в одной
λ , коэффициент забывания	0,75; 0,90	0,90	0,75
μ , штраф за пропуск	0,15	0,15	0,15
τ_C , порог подтверждения	0,50; 0,60; 0,65; 0,70	0,70	0,50 в шести проверках; 0,65 в одной

Для ByteTrack режим с включённой задержкой был выбран во всех семи итерациях внешней проверки. Для OC-SORT и SORT в шести случаях из семи по результатам настройки задержка не применялась. Это означает, что на большинстве обучающих разбиений улучшение для этих алгоритмов не удовлетворяло установленным критериям. Для случайного леса проверялись глубины деревьев 6 и 10, минимальное число обучающих примеров в конечной вершине 4 и 8, а также пороги решения 0,35, 0,50 и 0,65. Модель включала 80 деревьев, дисбаланс классов учитывался сбалансированными весами. В варианте с бюджетным ограничением во всех итерациях были выбраны порог 0,65, предельное пополнение бюджета 0,05 и ёмкость 2. Измерения времени выполнения проводились на процессоре AMD Ryzen 7 7840HS. Перед измерением выполнялись два предварительных запуска, результаты которых не учитывались. Затем время выполнения оценивалось по 30 повторным измерениям. Учитывалось только время работы дополнительного правила на сохранённых выходах ByteTrack. Обнаружение объектов, исходное сопровождение, обучение случайного леса и сопоставление с разметкой в измерение не входили. На время измерений автоматическое освобождение неиспользуемой памяти и подробная запись служебных сообщений были отключены. Приведённые значения характеризуют вычислительные затраты дополнительного слоя, а не полное время работы системы восприятия.

5. Результаты

Основные показатели оперативного выхода приведены в таблице 3. Для ByteTrack кратковременная задержка уменьшила число ложных наблюдений с 46 368 до 45 943, а число ложных инициаций – с 10 031 до 9 855. Правило изменило обработку 392 из 14 967 новых эпизодов, то есть 2,62%.

Таблица 3. Основные показатели оперативного выхода

Алгоритм	Режим	F_1	Ложные наблюдения на 100 кадров	Ложные инициации на 100 кадров	Подтверждение истинных треков
ByteTrack	Исходный	0,403972	2017,96	457,03	0,6428
ByteTrack	Задержка	0,404400	1998,61	449,02	0,6418
OC-SORT	Исходный	0,392258	1130,66	164,63	0,5614
OC-SORT	Задержка	0,392276	1128,99	164,27	0,5595
SORT	Исходный	0,391946	1195,35	179,49	0,5632
SORT	Задержка	0,391942	1194,25	179,13	0,5613

Для OC-SORT и SORT изменения оказались значительно меньше. Режим с задержкой включался только в одной из семи итераций внешней проверки. В остальных случаях настройка оставляла исходный режим, поскольку на обучающих последовательностях безопасное улучшение не подтверждалось. Средние значения сами по себе не показывают, насколько результат устойчив при переходе между видеозаписями, поэтому изменения F_1 и показателя ложных наблюдений дополнительно проверялись по условиям (2) и (3). Результаты приведены в таблице 4.

Таблица 4. Проверка сохранения качества и снижения числа ложных наблюдений

Алгоритм	Показатель	Среднее изменение	Стандартное отклонение	95%-й доверительный интервал	Результат
ByteTrack	H_1, F_1	+0,000428	0,000341	[+0,000220; +0,000682]	Условие выполнено
ByteTrack	H_2, O_{100}	-19,35	10,08	[-26,047; -12,279]	Условие выполнено
OC-SORT	H_1, F_1	+0,000018	0,000047	[+0,000000; +0,000054]	Условие выполнено
OC-SORT	H_2, O_{100}	-1,67	4,43	[-5,024; 0,000]	Условие не выполнено
SORT	H_1, F_1	-0,000005	0,000012	[-0,000014; 0,000000]	Условие выполнено
SORT	H_2, O_{100}	-1,10	2,91	[-3,297; 0,000]	Условие не выполнено

Для ByteTrack выполнены оба условия: нижняя граница доверительного интервала изменения F_1 остаётся выше допустимого уровня, а верхняя граница интервала изменения показателя ложных наблюдений находится ниже нуля. Снижение числа ложных наблюдений тем самым подтверждено при сохранении качества в установленном диапазоне. Для OC-SORT и SORT условие сохранения качества также выполнено, однако верхняя граница интервала изменения показателя ложных наблюдений достигает нуля. Отдельно роль бюджетного ограничения оценивалась на сохранённых результатах ByteTrack. Кратковременная задержка сравнивалась с порогом уверенности, правилом подтверждения M из N , байесовским подтверждением и случайным лесом. Результаты приведены в таблице 5.

Таблица 5. Сравнение правил инициации на сохранённых результатах ByteTrack

Правило	Изменение F_1 и результат	Изменение O_{100} и результат	Доля вмешательства	Обучение
Порог уверенности	0,000000; условие выполнено	0,00; условие не выполнено	0%	Нет
Правило подтверждения М из N	-0,046263; условие не выполнено	-1386,84; не проверялось	Не рассчитывалась	Нет
Байесовское подтверждение	-0,187741; условие не выполнено	-1846,46; не проверялось	Не рассчитывалась	Нет
Случайный лес	-0,084383; условие не выполнено	-1084,26; не проверялось	Не рассчитывалась	Да
Случайный лес с бюджетом	-0,000346; условие выполнено	-85,83; условие выполнено	4,81%	Да
Кратковременная задержка	+0,000428; условие выполнено	-19,35; условие выполнено	2,60%	Нет

Правило подтверждения М из N, байесовское подтверждение и случайный лес без бюджета заметно уменьшили число ложных наблюдений, но одновременно снизили F_1 за пределами установленного допуска. Случайный лес с бюджетным ограничением выполнил оба условия и сильнее снизил показатель ложных наблюдений. Для его обучения требовались размеченные эпизоды, тогда как предлагаемое правило не содержит обучаемого классификатора. Средняя доля затронутых эпизодов составила 4,81%. Для кратковременной задержки среднее по итерациям внешней проверки было равно 2,60%, а доля 2,62% рассчитана по общему числу эпизодов во всех последовательностях.

Характеристики эпизодов ложной инициации оценивались на одном и том же наборе до и после применения правила. Если эпизод полностью исчезал из оперативного выхода, его длительность в оперативном выходе принималась равной нулю. Из-за длинного хвоста распределения и небольшой доли затронутых эпизодов изменения средних характеристик оказались невелики. При этом для ByteTrack из оперативного выхода были исключены 176 ложных инициаций и 425 ложных наблюдений. Распределение длительности и числа наблюдений в эпизодах до и после применения задержки приведено в таблице 6.

Формальные ограничения проверялись на каждом начальном фрагменте видеозаписей. Всего проанализировано 29 385 таких фрагментов. Результаты приведены в таблице 7.

Нарушений балансового равенства и границ, заданных формулами (6) и (8), не обнаружено. Наибольшее отношение фактического изменения выхода к бюджетному пределу составило 0,675, а к границе по числу случаев задержки в отдельных фрагментах достигало единицы. Эти результаты подтверждают соблюдение заданных ограничений. Вычислительные затраты измерялись на сохранённых выходах ByteTrack. Результаты приведены в таблице 8.

Среднее время работы кратковременной задержки составило 0,901 мс

Таблица 6. Характеристики эпизодов ложной инициации

Алгоритм и режим	Число ложных эпизодов	Средняя длительность, кадры	Медиана	P95	Среднее число наблюдений
ByteTrack без задержки	12 260	5,395	1	20	3,831
ByteTrack с задержкой	12 260	5,373	1	20	3,814
OC-SORT без задержки	5 210	6,308	3	22	5,307
OC-SORT с задержкой	5 210	6,307	3	22	5,306
SORT без задержки	5 479	6,109	3	21	5,233
SORT с задержкой	5 479	6,107	3	21	5,231

Таблица 7. Проверка границы изменения оперативного выхода

Алгоритм	Активные итерации внешней проверки	Наибольшее $D(n)$	Наибольшее отношение $D(n)/(L_Q N_Q(n))$	Отношение к бюджетной границе	Нарушения
ByteTrack	7	105	0,722	0,552	0
OC-SORT	1	27	1,000	0,675	0
SORT	1	27	1,000	0,659	0

Таблица 8. Время работы правил на сохранённых результатах ByteTrack

Правило	Число измерений	Среднее время, мс/кадр	Медиана, мс/кадр	95-й процентиль, мс/кадр
Порог уверенности	30	0,442	0,441	0,451
Правило подтверждения М из N	30	1,061	1,061	1,070
Байесовское подтверждение	30	0,625	0,623	0,642
Случайный лес	30	0,481	0,488	0,517
Случайный лес с бюджетом	30	0,552	0,556	0,592
Кратковременная задержка	30	0,901	0,899	0,925

на кадр, случайного леса с бюджетом – 0,552 мс. Предлагаемое правило хранит состояние эпизодов, временно сохраняет наблюдения и обновляет относительную поддержку. Для случайного леса использовался заранее сформированный набор признаков, поэтому сравнение отражает стоимость выполнения правил, а не полный цикл подготовки данных.

6. Обсуждение результатов и ограничения

Случайный лес с бюджетом снизил показатель ложных наблюдений сильнее предлагаемого правила и сохранил изменение F_1 в допустимом диапазоне, но затронул более высокую долю эпизодов. Если имеется актуальная размеченная выборка, соответствующая конкретной камере, высоте съёмки и типу сцены, и более высокая доля вмешательств допустима, обучаемый метод может быть предпочтительнее по величине эффекта. Кратковременная задержка рассчитана на условия, в которых обучение и регулярное обновление классификатора затруднены. Во время работы метод не использует разметку. Решение определяется поддержкой нового эпизода и текущим состоянием бюджета, а журнал сохраняет обе величины вместе с причиной конечного статуса. Каждое вмешательство можно разобрать без приближённой внешней модели, а общее число задержек и отличие оперативного выхода остаются ограниченными формулами (6) и (8). Сравнение со случайным лесом без бюджета показывает роль ограничения. Неограниченная модель исключила больше ложных наблюдений, но ухудшила F_1 за пределами установленного допуска. После введения бюджета оба условия были выполнены. Полученный результат показывает, что бюджетное ограничение может применяться как к обучаемому классификатору, так и к необучаемому правилу.

Для OC-SORT и SORT задержка применялась только в одной из семи итераций внешней проверки. В остальных случаях настройка оставляла исходный режим. Качество сохранилось, но устойчивое снижение числа ложных наблюдений не подтвердилось. Основным практический результат относится к ByteTrack. Имеющиеся данные не позволяют распространять его на все алгоритмы сопровождения.

Исследование ограничено семью последовательностями VisDrone и результатами одной модели обнаружения [10]. Эти записи содержат малые объекты, частичные перекрытия и изменения масштаба, но не охватывают всего разнообразия транспортных сцен. Переносимость полученных результатов на другие наборы данных, камеры, условия полёта и детекторы в настоящей работе не проверялась. В дальнейшем такую проверку целесообразно провести, например, на U2UData и V2U4Real [11,12]. Формальная граница задаёт максимальный масштаб изменения оперативного выхода при любой последовательности входных оценок. Перед практическим применением метод необходимо оценить на независимых данных и проверить, как ложные треки влияют на конкретный модуль планирования движения.

Дополнительное ограничение связано с граничными сценариями, включая исчерпание бюджета вмешательств, отсутствие повторного наблюдения эпизода и неполное окно решения в конце последовательности. Эти случаи обрабатываются правилами метода, однако их отдельная количественная оценка требует более крупной независимой выборки.

С учётом указанных ограничений количественные выводы настоящей

работы относятся к исследованным последовательностям VisDrone, используемому детектору YOLOv8s и рассмотренным алгоритмам сопровождения. Перспективы дальнейших исследований включают независимую проверку на U2UData, V2U4Real и других наборах данных, применение дополнительных детекторов и камер, отдельные стресс-тесты перечисленных граничных сценариев и оценку влияния задержки на решения реального модуля планирования движения.

Заключение

Предложен интерпретируемый метод кратковременной задержки эпизодов инициации треков. Накопительный бюджет заранее ограничивает число вмешательств, а журнал решения сохраняет поддержку эпизода, текущий баланс бюджета, причину задержки и конечный статус. Балансовое равенство задаёт верхнюю границу числа задержек и возможного изменения оперативного выхода. Проверка 29 385 начальных фрагментов не выявила нарушений этих ограничений.

Основной результат получен для ByteTrack. Число ложных наблюдений в оперативном выходе уменьшилось на 425, число ложных инициаций – на 176. Правило изменило обработку 392 из 14 967 новых эпизодов, что составило 2,62%. Изменение F_1 осталось в допустимом диапазоне, а снижение числа ложных наблюдений было подтверждено статистически. Для OC-SORT и SORT качество также сохранилось, однако устойчивого снижения числа ложных наблюдений не установлено. Случайный лес с бюджетом сильнее снизил число ложных наблюдений, но требовал размеченных обучающих эпизодов и затрагивал более высокую долю эпизодов инициации. Предлагаемая задержка не использует обучаемый классификатор и сохраняет заранее заданный предел воздействия.

Исходный код, конфигурации, сценарии расчёта и поля журнала интерпретации опубликованы в открытом репозитории¹. Следующий этап – независимая проверка на других данных и камерах, а также оценка метода в составе реального модуля планирования движения.

Благодарности

Исследование выполнено в рамках государственного задания Министерства науки и высшего образования Российской Федерации, тема № 124112200072-2.

Список литературы

- [1] Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems / R. E. Kalman // Journal of Basic Engineering. – 1960. – Vol. 82, No. 1. – P. 35–45. – DOI 10.1115/1.3662552.
- [2] Simple Online and Realtime Tracking / A. Bewley, Z. Ge, L. Ott [et al.] // Proceedings of the IEEE International Conference on Image Processing. – Piscataway, NJ : IEEE, 2016. – P. 3464–3468. – DOI 10.1109/ICIP.2016.7533003.

¹Defense4UAVSwarm [Электронный ресурс]. URL: <https://github.com/lebedeffson/Defense4UAVSwarm> (дата обращения: 17.07.2026).

- [3] ByteTrack: Multi-Object Tracking by Associating Every Detection Box / Y. Zhang, P. Sun, Y. Jiang [et al.] // European Conference on Computer Vision. – Cham : Springer, 2022. – P. 1–21. – DOI 10.1007/978-3-031-20047-2_1.
- [4] Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking / J. Cao, J. Pang, X. Weng [et al.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – Piscataway, NJ : IEEE, 2023. – P. 9686–9696. – DOI 10.1109/CVPR52729.2023.00934.
- [5] StrongSORT: Make DeepSORT Great Again / Y. Du, Z. Zhao, Y. Song [et al.] // IEEE Transactions on Multimedia. – 2023. – Vol. 25. – P. 8725–8737. – DOI 10.1109/TMM.2023.3240881.
- [6] Wojke, N. Simple Online and Realtime Tracking with a Deep Association Metric / N. Wojke, A. Bewley, D. Paulus // Proceedings of the IEEE International Conference on Image Processing. – Piscataway, NJ : IEEE, 2017. – P. 3645–3649. – DOI 10.1109/ICIP.2017.8296962.
- [7] Bergmann, P. Tracking Without Bells and Whistles / P. Bergmann, T. Meinhardt, L. Leal-Taixé // Proceedings of the IEEE/CVF International Conference on Computer Vision. – Piscataway, NJ : IEEE, 2019. – P. 941–951. – DOI 10.1109/ICCV.2019.00103.
- [8] Breiman, L. Random Forests / L. Breiman // Machine Learning. – 2001. – Vol. 45, No. 1. – P. 5–32. – DOI 10.1023/A:1010933404324.
- [9] Geifman, Y. Selective Classification for Deep Neural Networks / Y. Geifman, R. El-Yaniv // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4878–4887.
- [10] VisDrone-VDT2018: The Vision Meets Drone Video Detection and Tracking Challenge Results / P. Zhu, L. Wen, D. Du [et al.] // Computer Vision – ECCV 2018 Workshops. – Cham : Springer, 2019. – P. 496–518. – DOI 10.1007/978-3-030-11021-5_29.
- [11] U2UData: A Large-scale Cooperative Perception Dataset for Swarm UAVs Autonomous Flight / T. Feng, X. Wang, F. Han [et al.] // Proceedings of the 32nd ACM International Conference on Multimedia. – New York : ACM, 2024. – P. 7600–7608. – DOI 10.1145/3664647.3681151.
- [12] V2U4Real: A Real-world Large-scale Dataset for Vehicle-to-UAV Cooperative Perception / W. Li, H. Xiang, T. Wang [et al.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – Piscataway, NJ : IEEE, 2026. – P. 4728–4737.

INTERPRETABLE CONTROL OF TRACK INITIATION WITH PREDEFINED INTERVENTION BOUNDS IN UNMANNED AERIAL VEHICLE PERCEPTION SYSTEMS

Trofimov Yu. V.^{1, a}, Averkin A. N.^{1, 2, b}, Shevchenko A. V.^{1, c}, Kuznetsov E. M.^{1, d},
Lebedev A. D.^{1, e}

¹ Dubna State University, Dubna, Russia.

² Federal Research Center «Computer Science and Control» of the Russian Academy of Sciences, Moscow, Russia.

E-mail: ^a ura_trofim@bk.ru, ^b averkin2003@inbox.ru, ^c leviathan0909@gmail.com, ^d kem.22@uni-dubna.ru,

^e lad.24@uni-dubna.ru

Abstract. In aerial monitoring systems, a single detection error may develop into a persistent false track and affect data passed to downstream components. The study aims to reduce such errors while bounding the scale of intervention in advance and retaining a traceable account of each decision. An external method is proposed that briefly delays low-support track-initiation episodes. The number of delays is controlled by an accumulated budget, while the support estimate, current budget balance, reason for intervention, and final status are recorded for each episode. The method was evaluated on seven video sequences from a publicly available aerial-imagery dataset containing small objects, partial occlusions, and substantial scale changes. Three tracking algorithms and a supervised ensemble of decision trees were compared. For ByteTrack, the delay was applied to 2.62 percent of track-initiation episodes, removing 425 false observations and 176 false initiations from the operational output, while the change in the combined precision-recall measure remained within the predefined tolerance. For OC-SORT and SORT, the reduction in false observations was not confirmed under the adopted statistical criterion. An examination of 29,385 initial video fragments found no violations of the predefined limits. The additional processing time averaged 0.901 milliseconds per frame. The results show that a short delay can reduce the number of false observations while keeping the scale of intervention under control.

Keywords: *intelligent transport systems, unmanned aerial vehicles, object tracking, track initiation, bounded intervention, interpretable artificial intelligence, short-term delay.*

References

- [1] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35–45. <https://doi.org/10.1115/1.3662552>
- [2] Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 3464–3468). IEEE. <https://doi.org/10.1109/ICIP.2016.7533003>
- [3] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision* (pp. 1–21). Springer. https://doi.org/10.1007/978-3-031-20047-2_1

For citation:

Trofimov Yu. V., Averkin A. N., Shevchenko A. V., Kuznetsov E. M., Lebedev A. D. Interpretable Control of Track Initiation with Predefined Intervention Bounds in Unmanned Aerial Vehicle Perception Systems. *Intelligent transport*, 2026, no. 3(39), pp. 69–84. DOI: 10.24412/3033-6007-2026-339-69-84. (in Russian)

- [4] Cao, J., Pang, J., Weng, X., Khirodkar, R., & Kitani, K. (2023). Observation-centric SORT: Rethinking SORT for robust multi-object tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9686–9696). IEEE. <https://doi.org/10.1109/CVPR52729.2023.00934>
- [5] Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., & Meng, H. (2023). StrongSORT: Make DeepSORT great again. *IEEE Transactions on Multimedia*, 25, 8725–8737. <https://doi.org/10.1109/TMM.2023.3240881>
- [6] Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *Proceedings of the IEEE International Conference on Image Processing* (pp. 3645–3649). IEEE. <https://doi.org/10.1109/ICIP.2017.8296962>
- [7] Bergmann, P., Meinhardt, T., & Leal-Taixé, L. (2019). Tracking without bells and whistles. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 941–951). IEEE. <https://doi.org/10.1109/ICCV.2019.00103>
- [8] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [9] Geifman, Y., & El-Yaniv, R. (2017). Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4878–4887). Curran Associates.
- [10] Zhu, P., Wen, L., Du, D., et al. (2019). VisDrone-VDT2018: The vision meets drone video detection and tracking challenge results. In *Computer Vision – ECCV 2018 Workshops* (pp. 496–518). Springer. https://doi.org/10.1007/978-3-030-11021-5_29
- [11] Feng, T., Wang, X., Han, F., Zhang, L., & Zhu, W. (2024). U2UData: A large-scale cooperative perception dataset for swarm UAVs autonomous flight. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 7600–7608). ACM. <https://doi.org/10.1145/3664647.3681151>
- [12] Li, W., Xiang, H., Wang, T., Wu, S., Xia, Q., Wang, C., & Wen, C. (2026). V2U4Real: A real-world large-scale dataset for vehicle-to-UAV cooperative perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4728–4737). IEEE.

МЕТОДИКА ОЦЕНКИ ПОЛНОТЫ, ДОСТОВЕРНОСТИ И КАЧЕСТВА ДАННЫХ МОНИТОРИНГА ИНФРАСТРУКТУРЫ ИНТЕЛЛЕКТУАЛЬНЫХ ТРАНСПОРТНЫХ СИСТЕМ

Давыдов Р.Д.^{1, а}, Ахмадов Г.Д.^{1, b}, Жабриков М.В.^{1, c}

¹ Федеральное автономное учреждение «Российский дорожный научно-исследовательский институт» (ФАУ «РОСДОРНИИ»), Москва, Россия.

E-mail: ^a davydov@rosdornii.ru, ^b akhmadov@rosdornii.ru, ^c maxpkey@mail.ru

Аннотация. В статье представлена методика количественной оценки полноты, достоверности и качества первичных данных, собираемых в рамках мониторинга инфраструктуры интеллектуальных транспортных систем на автомобильных дорогах общего пользования Российской Федерации. Актуальность исследования обусловлена необходимостью формализованного контроля качества исходной информации, используемой для формирования сводных перечней оборудования ИТС и последующего анализа состояния цифровой дорожной инфраструктуры. Предложенная методика является инструментом проверки достоверности и полноты первичных данных, на основе которых такой анализ выполняется. Научная новизна работы заключается в разработке системы взаимосвязанных критериев и интегральных индексов, предназначенных для количественной оценки полноты и достоверности данных мониторинга ИТС, а также формирования сводного индекса качества информации. Методика включает расчет частных критериев качества данных, формирование интегрального индекса полноты данных, интегрального индекса достоверности данных и интегрального индекса качества информации. Предложенный подход обеспечивает воспроизводимость расчетов, сопоставимость результатов мониторинга и возможность последующей автоматизации процедур оценки качества данных мониторинга ИТС.

Ключевые слова: интеллектуальные транспортные системы, качество данных, полнота данных, достоверность данных, мониторинг ИТС.

Введение

Развитие ИТС на автомобильных дорогах общего пользования Российской Федерации сопровождается активным внедрением цифровых технологий мониторинга и управления дорожным движением [1]. В транспортную инфраструктуру интегрируются различные технические и программные решения, включая системы видеонаблюдения, детекторы транспортных потоков, метеорологические станции, динамические информационные табло и программные комплексы управления дорожным движением [2]. Расширение применения таких технологий приводит к формированию значительных массивов данных о состоянии и составе инфраструктуры ИТС [3, 4].

Для цитирования:

Давыдов Р.Д., Ахмадов Г.Д., Жабриков М.В. Методика оценки полноты, достоверности и качества данных мониторинга инфраструктуры интеллектуальных транспортных систем. *Интеллектуальный транспорт*. 2026. № 3(39). С. 85–99. DOI: 10.24412/3033-6007-2026-339-85-99.

В предыдущих исследованиях авторами были предложены методические подходы к организации регулярного мониторинга элементов ИТС, включающие формирование респондентской базы, разработку стандартизированных опросных форм и применение процедур обработки первичных данных [5,6].

В рамках мониторинга формируется структурированный массив данных, включающий сведения о типах и количестве установленного периферийного оборудования ИТС (детекторы транспорта, метеостанции, табло, камеры видеонаблюдения и др.), их технических характеристиках, датах ввода в эксплуатацию, принадлежности к категориям дорог, а также о владельцах и операторах инфраструктуры. Сбор данных осуществляется с использованием стандартизированных опросных форм, заполняемых респондентами – владельцами и операторами элементов ИТС на федеральных, региональных и местных автомобильных дорогах. Состав и структура опросных форм подробно описаны в [5,6].

Вместе с тем получение информации от большого числа респондентов неизбежно сопровождается риском возникновения неполноты данных, различий в уровне детализации представляемых сведений и возможных логических несоответствий. Даже после проведения процедур первичной обработки и форматно-логического контроля сохраняется необходимость оценки качества сформированного массива данных, поскольку достоверность аналитических выводов напрямую зависит от надежности исходной информации [7].

В этой связи важно разграничить две задачи: анализ состояния инфраструктуры ИТС (оценка количества и состава оборудования) и оценку качества первичных данных, на основе которых этот анализ проводится. Настоящее исследование посвящено второй задаче – разработке формализованного инструментария для проверки полноты и достоверности информации, поступающей от респондентов, до ее использования в аналитических целях. Такой подход позволяет выявлять проблемные области сбора данных и повышать надёжность итоговых выводов о состоянии инфраструктуры ИТС.

В научной и прикладной практике оценка качества данных рассматривается как важный элемент информационного обеспечения систем управления. Международные стандарты управления данными, включая комплекс стандартов ISO 8000¹, подчеркивают необходимость применения формализованных процедур оценки полноты, достоверности и согласованности информации. Международные исследования в области развития и стандартизации ИТС также указывают на необходимость формирования унифицированных подходов к управлению транспортными данными и обеспечению их сопоставимости на различных уровнях транспортной инфраструктуры [8]. Однако применительно к задачам мониторинга инфраструктуры ИТС такие методические подходы разработаны недостаточно.

В связи с этим возникает необходимость разработки воспроизводимого методического инструментария, позволяющего количественно оценивать полноту, достоверность и качество данных мониторинга ИТС. Использование системы интегральных индексов позволяет обеспечить сопоставимость результатов оценки, повысить прозрачность интерпретации данных и создать основу для автоматизации процедур контроля качества информационных массивов [9].

¹В частности, ISO 8000-1:2022. Data quality – Part 1: Overview. Geneva: International Organization for Standardization, 2022

Целью настоящего исследования является разработка методики количественной оценки полноты, достоверности и качества данных мониторинга ИТС на основе системы взаимосвязанных интегральных индексов.

1. Состав данных мониторинга ИТС

Исходный массив данных мониторинга элементов ИТС формируется на основе ответов респондентов на стандартизированные опросные формы. В состав данных входят следующие категории информации:

- идентификационные сведения об объекте (регион, наименование владельца/оператора);
- перечень установленных технических средств ИТС с указанием типа, модели, количества;
- временные параметры (год производства, год ввода в эксплуатацию);
- сведения о программном обеспечении (аналогично периферийному оборудованию).

Все параметры имеют формализованную структуру и допустимые форматы значений, что обеспечивает возможность автоматизированной обработки и последующей оценки качества данных.

2. Принципы построения системы оценки качества данных ИТС

Формирование методики оценки качества данных мониторинга ИТС основывается на совокупности методологических принципов, обеспечивающих объективность, сопоставимость и воспроизводимость результатов оценки. С учетом распределенного характера сбора информации и участия большого числа респондентов система оценки качества данных должна учитывать как полноту поступающей информации, так и степень ее достоверности².

Первым принципом является принцип объективности оценки. Оценка качества данных должна осуществляться на основе формализованных количественных критериев, позволяющих минимизировать влияние субъективных экспертных оценок. Использование математически определенных критериев обеспечивает единый подход к оценке данных, поступающих от различных участников мониторинга.

Вторым принципом является принцип воспроизводимости результатов оценки. Методика должна обеспечивать возможность повторного расчета критериев качества данных при использовании одинакового набора исходных данных. Это позволяет обеспечить прозрачность аналитических процедур и сопоставимость результатов мониторинга в различные периоды наблюдения.

Третьим принципом является принцип иерархичности системы критериев. Оценка качества данных осуществляется на нескольких уровнях, включая оценку отдельных критериев, формирование интегральных индексов полноты и достоверности, а также расчет сводного индекса качества информации. Такой подход позволяет учитывать различные аспекты качества

²Методические рекомендации по оценке достоверности, полноты и точности данных, используемых при расчёте показателей результативности и эффективности контрольно-надзорной деятельности: протокол заседания проектного комитета от 9 ноября 2017 г. № 73(13)

данных и формировать обобщенную количественную оценку их надежности.

Четвертым принципом является принцип сопоставимости данных мониторинга. Система критериев должна обеспечивать возможность сравнения результатов оценки качества данных между различными субъектами Российской Федерации, категориями автомобильных дорог и временными периодами мониторинга. Это позволяет выявлять проблемные области сбора информации и повышать эффективность управления процессом мониторинга.

Пятым принципом является принцип возможности автоматизации процедур оценки. Используемые критерии и алгоритмы расчета должны быть сформированы таким образом, чтобы их применение могло быть реализовано в автоматизированных информационных системах обработки данных. Это обеспечивает повышение оперативности анализа информации и снижение трудоемкости процедур контроля качества данных.

С учетом указанных принципов в рамках настоящего исследования предложена система интегральных индексов, позволяющая количественно оценивать различные аспекты качества данных мониторинга ИТС. В структуре методики выделяются критерии полноты данных и критерии достоверности информации, на основе которых формируется интегральный индекс качества информации. Общая структура предложенной системы оценки качества данных мониторинга ИТС наглядно представлена на рисунке 1.

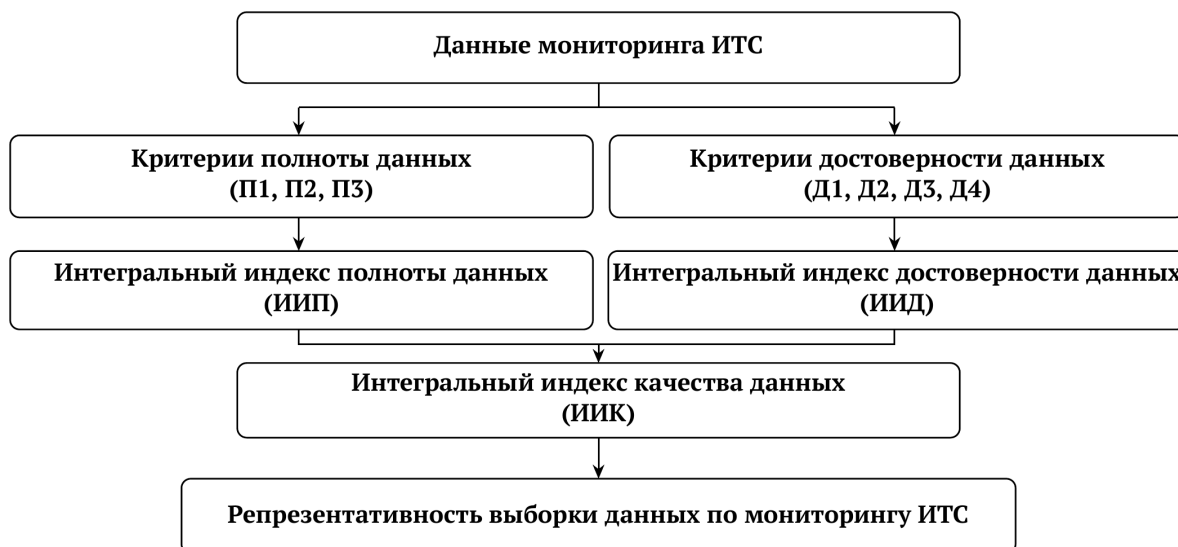


Рис. 1. Структура системы оценки качества данных мониторинга ИТС (составлен авторами)

Предлагаемая система оценки качества данных мониторинга ИТС имеет иерархическую структуру. На первом уровне рассчитываются частные критерии полноты и достоверности данных. На втором уровне на основе агрегирования частных критериев формируются интегральный индекс полноты данных и интегральный индекс достоверности данных. На третьем уровне указанные индексы объединяются в интегральный индекс качества информации, обеспечивающий получение сводной количественной оценки надежности данных мониторинга ИТС.

Для обоснования выбора критериев и демонстрации преимущественности разработанной методики с существующими подходами в таблице 1 приведено сопоставление используемых критериев с международными стандартами и

отраслевыми практиками.

Сопоставление выполнено авторами на основе анализа исходных источников в виде обобщенных характеристик.

Таблица 1. Сопоставление критериев качества данных с международными стандартами и отраслевыми практиками

Измерение качества данных	ISO/IEC 25012:2008 / ISO/IEC 25024:2015	FHWA (Traffic Data Quality Measurement, 2004)	Предложенный критерий
Полнота (Completeness)	Характеристика модели качества данных; для измерения установлены количественные меры	Степень, в которой данные не являются пропущенными и обладают достаточной широтой и глубиной для поставленной задачи	П1 (охват респондентов), П2 (охват категорий дорог), П3 (заполненность форм)
Достоверность (Assurasy)	Характеристика модели качества данных; измеряется через соответствие данных эталону	Степень, в которой данные являются безошибочными, удовлетворяют проверкам внутренней согласованности и не имеют дубликатов в пределах одной базы данных	Д1 (происхождение данных), Д2 (внутренняя непротиворечивость)
Согласованность (Consistency)	Характеристика модели качества данных; определяется как «отсутствие противоречий между данными»	Степень, в которой данные представлены в одном и том же формате	Д2 (внутренняя непротиворечивость), Д4 (согласованность с историческими данными)
Форматное соответствие (Format/Validity)	В рамках требований и оценки качества систем учитывается через соответствие структуре данных	Определяется как степень, в которой данные соответствуют установленному формату	Д3 (форматное соответствие)
Своевременность (Timeliness)	Характеристика модели качества данных; определяется как «степень, в которой данные являются достаточно актуальными для поставленной задачи»	Отражает промежуток времени между наступлением события и вводом информации в соответствующую базу данных	Д4 (согласованность с историческими данными)

Практическое применение предложенной методики предполагает последовательный переход от оценки отдельных характеристик данных к формированию сводной количественной оценки качества информации. На основе исходного массива данных мониторинга первоначально рассчитываются частные критерии полноты и достоверности данных, после чего выполняется формирование соответствующих интегральных индексов. На завершающем этапе интегральные индексы объединяются в сводный индекс качества информации, используемый для оценки надежности сформированной информационной базы мониторинга ИТС, выявления проблемных областей сбора

данных и повышения сопоставимости результатов мониторинга.

Порядок применения методики включает следующие этапы:

- 1) формирование исходного массива данных мониторинга ИТС;
- 2) расчет частных критериев полноты данных;
- 3) расчет частных критериев достоверности данных;
- 4) расчет интегрального индекса полноты данных;
- 5) расчет интегрального индекса достоверности данных;
- 6) расчет интегрального индекса качества информации;
- 7) интерпретация полученного значения по установленной шкале качества данных.

Подход к агрегированию критериев основан на методологии композиционных индексов [9], однако состав критериев и структура интегральных индексов адаптированы применительно к задачам мониторинга элементов ИТС³.

Таким образом, ссылка на методологию композиционных индексов используется как теоретическая основа агрегирования критериев, тогда как состав критериев и структура индексов полноты, достоверности и качества данных предложены применительно к задачам мониторинга ИТС.

3. Оценка полноты данных мониторинга ИТС

Полнота данных является одним из ключевых критериев качества информации, формируемой в рамках мониторинга ИТС. Недостаточная полнота данных может приводить к искажению аналитических результатов и снижению достоверности выводов о состоянии и развитии инфраструктуры ИТС⁴.

В рамках настоящего исследования полнота данных рассматривается как совокупная характеристика, отражающая степень охвата респондентов мониторинга, представление информации по различным категориям автомобильных дорог, а также степень заполненности опросных форм. Для количественной оценки полноты данных предложена система из трех критериев.

Первый критерий полноты (П1) характеризует долю респондентов, представивших информацию в рамках мониторинга. Данный критерий отражает уровень участия владельцев и операторов элементов ИТС в процессе сбора данных и рассчитывается как отношение числа респондентов, предоставивших ответы, к общему числу респондентов, включенных в мониторинг.

Второй критерий полноты (П2) отражает охват категорий автомобильных дорог, по которым представлена информация. В рамках мониторинга учитываются данные по автомобильным дорогам федерального, регионального и местного значения. Критерий определяется как доля категорий дорог, по которым представлены сведения об элементах ИТС.

Третий критерий полноты (П3) характеризует степень заполненности опросных форм и отражает долю заполненных информационных полей в

³UNECE. Generic Statistical Business Process Model (GSBPM). URL: <https://unece.org/statistics/gsbpm-v5.2> (дата обращения: 01.03.2026).

⁴ГОСТ Р ИСО 3534-1-2019. Статистические методы. Словарь и условные обозначения. Часть 1. Общие статистические термины и термины, используемые в теории вероятностей. М.: Стандартинформ, 2019

представленных респондентами формах мониторинга. Данный критерий позволяет учитывать уровень детализации представляемой информации и выявлять случаи частичного заполнения форм.

Для формирования обобщенной количественной оценки полноты данных используется интегральный индекс полноты данных (ИИП), рассчитываемый как взвешенная сумма указанных критериев:

$$\text{ИИП} = \alpha \cdot \text{П1} + \beta \cdot \text{П2} + \gamma \cdot \text{ПЗ}, \quad (1)$$

где α , β , γ – весовые коэффициенты критериев полноты данных.

Выбор весовых коэффициентов обусловлен различной значимостью критериев при формировании полноты данных мониторинга. Наибольшее значение имеет уровень участия респондентов, поскольку отсутствие информации от владельца или оператора инфраструктуры делает невозможным учет соответствующих элементов ИТС в формируемой базе данных мониторинга.

Критерий охвата категорий автомобильных дорог отражает степень представления данных по различным уровням дорожной сети. Наличие информации по федеральным, региональным и местным автомобильным дорогам позволяет обеспечить территориальную и функциональную полноту анализа инфраструктуры ИТС.

Критерий заполненности опросных форм характеризует степень детализации представляемой информации. Несмотря на важность данного критерия, его влияние на общую полноту данных является менее значительным по сравнению с участием респондентов и охватом категорий дорог, поскольку даже частично заполненные формы могут содержать сведения, пригодные для анализа.

С учетом указанной логики распределения значимости критериев в рамках методики в соответствии с существующими практиками и экспертной оценкой используются следующие значения весовых коэффициентов:

- $\alpha = 0,5$ – вес критерия участия респондентов;
- $\beta = 0,3$ – вес критерия охвата категорий автомобильных дорог;
- $\gamma = 0,2$ – вес критерия заполненности опросных форм.

Распределение весовых коэффициентов отражает сравнительную значимость критериев при обеспечении полноты данных мониторинга ИТС.

Использование интегрального индекса полноты позволяет получить единую количественную оценку степени полноты данных мониторинга ИТС и выявлять случаи недостаточного информационного охвата, требующие дополнительной проверки или уточнения представленных сведений.

4. Оценка достоверности данных мониторинга ИТС

Помимо полноты информации важной характеристикой качества данных мониторинга является их достоверность. Достоверность данных отражает степень соответствия представленных сведений фактическому состоянию элементов ИТС и корректность их представления в информационной базе мониторинга.

В условиях распределенного сбора информации, при котором данные предоставляются большим количеством респондентов, существует риск

возникновения логических несоответствий, ошибок ввода, а также различий в интерпретации отдельных параметров. В связи с этим возникает необходимость использования формализованных критериев, позволяющих количественно оценивать достоверность представляемых данных.

В рамках настоящего исследования для оценки достоверности данных мониторинга предложена система из четырех критериев.

Первый критерий (Д1) характеризует происхождение данных и отражает долю информации, полученной непосредственно от владельцев или операторов инфраструктуры ИТС. Данные, поступающие от первоисточников, обладают более высокой степенью надежности по сравнению со сведениями, полученными из вторичных источников или агрегированных отчетов.

Второй критерий (Д2) отражает внутреннюю непротиворечивость представленных данных. В рамках данного критерия проводится проверка логических взаимосвязей между параметрами, содержащимися в опросных формах. Например, год ввода оборудования в эксплуатацию не может быть меньше года его производства, а тип оборудования должен соответствовать заявленным функциональным характеристикам.

Третий критерий (Д3) характеризует форматное соответствие данных установленным требованиям. Данный критерий отражает степень соответствия представленных сведений структуре опросных форм, типам полей и допустимым форматам значений.

Четвертый критерий (Д4) отражает согласованность данных с результатами предыдущих этапов мониторинга. В рамках данного критерия проводится сопоставление текущих данных с ранее полученными с целью выявления аномальных или необоснованных изменений.

Для формирования обобщенной оценки достоверности данных используется интегральный индекс достоверности данных (ИИД), рассчитываемый как взвешенная сумма указанных критериев:

$$\text{ИИД} = \delta \cdot \text{Д1} + \varepsilon \cdot \text{Д2} + \zeta \cdot \text{Д3} + \eta \cdot \text{Д4}, \quad (2)$$

где δ , ε , ζ , η – весовые коэффициенты критериев достоверности данных.

Выбор весовых коэффициентов обусловлен различной значимостью критериев при оценке достоверности информации. Наибольшее значение имеет происхождение данных, поскольку сведения, поступающие непосредственно от владельцев и операторов инфраструктуры ИТС, обладают наибольшей степенью надежности.

Критерий внутренней непротиворечивости отражает логическую согласованность представленных данных и позволяет выявлять ошибки, возникающие при заполнении опросных форм. Форматное соответствие характеризует техническую корректность представления информации в установленной структуре данных. Критерий согласованности с историческими данными играет вспомогательную роль и используется для выявления аномальных изменений данных.

С учетом указанной иерархии значимости критериев в рамках методики в соответствии с существующими практиками и экспертной оценкой используются следующие значения весовых коэффициентов:

- $\delta = 0,4$ – вес критерия происхождения данных;
- $\varepsilon = 0,3$ – вес критерия внутренней непротиворечивости;

- $\zeta = 0,2$ – вес критерия форматного соответствия;
- $\eta = 0,1$ – вес критерия согласованности с историческими данными.

Распределение весовых коэффициентов отражает приоритетность происхождения данных и их логической непротиворечивости при оценке достоверности сведений мониторинга ИТС.

Использование интегрального индекса достоверности позволяет количественно оценивать степень надежности данных мониторинга ИТС и выявлять случаи, требующие дополнительной проверки или уточнения представленных сведений.

5. Интегральный индекс качества информации мониторинга ИТС

Для получения обобщенной количественной оценки качества данных мониторинга ИТС используется интегральный индекс качества информации. Данный индекс формируется на основе объединения индексов полноты и достоверности данных, отражающих различные аспекты качества формируемой информационной базы мониторинга.

Индекс полноты данных характеризует степень охвата респондентов, категорий автомобильных дорог и уровень заполненности опросных форм. Индекс достоверности данных отражает надежность представленной информации, включая происхождение данных, их внутреннюю непротиворечивость, форматное соответствие и согласованность с ранее полученными результатами мониторинга.

Для интегральной оценки качества информации используется интегральный индекс качества информации (ИИК), рассчитываемый как взвешенная сумма индекса полноты данных и индекса достоверности данных:

$$\text{ИИК} = \omega \cdot \text{ИИП} + \varphi \cdot \text{ИИД}, \quad (3)$$

где ИИП – интегральный индекс полноты данных; ИИД – интегральный индекс достоверности данных; ω , φ – весовые коэффициенты, отражающие значимость соответствующих компонентов оценки качества данных.

Таким образом, ИИК представляет собой сводный количественный индекс, интегрирующий характеристики полноты и достоверности данных мониторинга ИТС.

В рамках предложенной методики индексы полноты и достоверности рассматриваются как равнозначные компоненты качества данных мониторинга. Полнота информации определяет степень информационного охвата наблюдаемой системы, в то время как достоверность характеризует надежность и корректность представленных сведений. С учетом этого в расчетах используются равные значения весовых коэффициентов:

- $\omega = 0,5$ – вес индекса полноты данных;
- $\varphi = 0,5$ – вес индекса достоверности данных.

Для интерпретации полученных значений интегрального индекса качества информации используется следующая шкала оценки в соответствии с международными практиками применения композиционных индексов [9]:

- 0,90–1,00 – высокая степень качества данных;
- 0,75–0,89 – хорошее качество данных;
- 0,60–0,74 – удовлетворительное качество данных;
- менее 0,60 – неудовлетворительное качество данных.

Использование интегрального индекса качества информации позволяет формировать обобщенную количественную оценку надежности данных мониторинга ИТС и применять единый критерий при анализе результатов мониторинга. Полученные значения индекса могут использоваться для ранжирования источников информации, выявления проблемных областей сбора данных и повышения эффективности процедур контроля качества информационной базы мониторинга.

Предложенный индекс качества информации может применяться как при федеральном мониторинге инфраструктуры ИТС, так и при оценке качества данных отдельных региональных или ведомственных информационных систем.

6. Области определения и способ формирования критериев

Все частные критерии полноты (П1–П3) и достоверности (Д1–Д4), а также интегральные индексы (ИИП, ИИД, ИИК) являются нормированными величинами, принимающими значения в интервале $[0, 1]$. Значение 0 соответствует полному отсутствию учитываемого признака (например, ни один респондент не предоставил данные), значение 1 – максимально возможному уровню (например, все респонденты предоставили полностью заполненные формы).

Расчет частных критериев осуществляется на основе первичных данных, полученных из опросных форм мониторинга. Каждый критерий представляет собой отношение фактически зафиксированного значения к максимально возможному (эталонному) значению, определенному в соответствии со структурой мониторинга. Например, критерий П1 рассчитывается как отношение числа ответивших респондентов к общему числу респондентов, включенных в мониторинг. Аналогичным образом формируются остальные критерии.

Интегральные индексы ИИП, ИИД и ИИК являются взвешенными суммами соответствующих нормированных критериев и также находятся в интервале $[0, 1]$. Это обеспечивает единую шкалу интерпретации и сопоставимость результатов оценки качества данных на всех уровнях мониторинга.

7. Репрезентативность данных мониторинга ИТС

Помимо полноты и достоверности информации важной характеристикой качества данных мониторинга является репрезентативность выборки. Репрезентативность отражает степень соответствия полученной выборки данных структуре исследуемой совокупности объектов мониторинга и обеспечивает корректность последующих аналитических выводов.

В рамках мониторинга ИТС репрезентативность данных определяется уровнем охвата различных элементов дорожной инфраструктуры, территориальным распределением респондентов и представленностью различных типов оборудования и программного обеспечения ИТС. Недостаточная репре-

зентативность выборки может приводить к искажению результатов анализа и формированию некорректных выводов о состоянии инфраструктуры ИТС на уровне всей дорожной сети.

Оценка репрезентативности данных мониторинга осуществляется на основе анализа трех основных аспектов: географического охвата наблюдения, уровневой структуры дорожной сети и типового состава элементов ИТС. В соответствии с подходами, изложенными в [10], оценка репрезентативности при неполном охвате респондентов требует учета структуры совокупности и применения методов взвешивания для коррекции смещений выборки.

Географический аспект предполагает участие респондентов из различных субъектов Российской Федерации и позволяет учитывать территориальные особенности внедрения ИТС. Уровневый аспект связан с представлением данных по автомобильным дорогам различных категорий, включая федеральные, региональные и местные дороги. Типовой аспект отражает представленность различных типов периферийного оборудования и видов программного обеспечения, используемых в составе ИТС.

Совокупность указанных характеристик позволяет оценивать степень соответствия сформированной информационной базы мониторинга структуре исследуемой системы. При этом критерии полноты и достоверности данных выступают необходимым условием обеспечения репрезентативности выборки, поскольку неполные или недостоверные данные не могут рассматриваться как репрезентативные.

Разработанная система критериев качества данных мониторинга ИТС формирует методическую основу для последующей количественной оценки репрезентативности выборки. В рамках настоящего исследования репрезентативность рассматривается как условие, накладываемое на значения интегральных индексов полноты и достоверности. Выборка данных мониторинга считается репрезентативной, если одновременно выполняются условия по достижению пороговых значений как по полноте, так и по достоверности данных.

При выполнении условий интегральный индекс качества информации ИИК может быть использован как сводный критерий для ранжирования источников данных и сравнения результатов мониторинга по различным регионам. В случае невыполнения хотя бы одного из условий выборка признается недостаточно репрезентативной, что требует дополнительного сбора данных или корректировки респондентской базы.

Таким образом, связь ИИК с репрезентативностью носит пороговый характер: высокие значения ИИП и ИИД являются необходимым (но не достаточным) условием репрезентативности, поскольку также требуется учет территориального, уровневого и типового распределения объектов мониторинга. Последующий анализ этих аспектов предлагается проводить на основе процедур стратифицированного анализа, описание которых выходит за рамки настоящей работы и будет представлено в дальнейших исследованиях.

8. Обсуждение

Предложенная методика основана на методологии композиционных индексов, что обеспечивает воспроизводимость расчетов и прозрачность агрегирования критериев. По сравнению с существующими подходами к оценке качества данных (ISO 8000, методические рекомендации Росавтодора) разработанная система критериев адаптирована к специфике мониторинга ИТС и учитывает распределенный характер сбора информации от множества респондентов.

В качестве ограничений настоящего исследования следует отметить, что весовые коэффициенты критериев определены на основе экспертных оценок и существующих практик, однако их уточнение возможно по мере накопления эмпирических данных. Кроме того, предложенная формализация репрезентативности носит пороговый характер и требует дополнительной разработки процедур стратифицированного анализа. Указанные ограничения не снижают практической ценности методики, а формируют направления для дальнейших исследований.

Таким образом, работа представляет собой базовую методическую основу, которая может быть дополнена и уточнена при апробации на реальных массивах данных мониторинга ИТС.

Заключение

В результате проведенного исследования разработана методика количественной оценки качества данных мониторинга ИТС на автомобильных дорогах общего пользования Российской Федерации. Предложенный подход основан на использовании системы интегральных индексов, позволяющих комплексно оценивать полноту, достоверность и общее качество информации.

В рамках методики сформирована система из трех частных критериев оценки полноты данных (доля ответивших респондентов, охват категорий автомобильных дорог и степень заполненности опросных форм) и четырех частных критериев оценки достоверности (происхождение данных, внутренняя непротиворечивость, форматное соответствие, согласованность с историческими данными). На их основе предложены интегральный индекс полноты данных (ИИП), интегральный индекс достоверности данных (ИИД) и интегральный индекс качества информации (ИИК).

Разработанная методика является логическим продолжением методических подходов к организации мониторинга элементов ИТС, представленных в предыдущих исследованиях авторов [5,6], и может применяться для оценки качества сформированной информационной базы мониторинга.

Практическая реализация предложенной методики выполнена в рамках научно-исследовательской работы по формированию системы мониторинга элементов ИТС и созданию сводного перечня периферийного оборудования ИТС на автомобильных дорогах общего пользования Российской Федерации.

Направления дальнейших исследований:

- апробация предложенной методики на реальном массиве данных мониторинга;
- расчет интегральных индексов полноты, достоверности и качества

информации;

- уточнение весовых коэффициентов на основе эмпирических данных;
- разработка процедур стратифицированного анализа для углубленной оценки репрезентативности.

Список литературы

- [1] Мировой и российский опыт применения интеллектуальных транспортных систем / С. В. Егоров, П. В. Шационок, А. И. Ерпылева, Д. И. Жарков // Транспортное дело России. – 2022. – № 2. – С. 130–136. – DOI 10.52375/20728689_2022_2_130. – EDN QHNBQNH.
- [2] Солодкий, А. И. Развитие интеллектуальных транспортных систем в России: проблемы и пути их решения. Новый этап / А. И. Солодкий // Интеллект. Инновации. Инвестиции. – 2020. – № 6. – С. 10–19. – DOI 10.25198/2077-7175-2020-6-10. – EDN PDUSFR.
- [3] Жанказиев, С. В. Принципы формирования государственной системы сертификации элементов интеллектуальных транспортных систем в Российской Федерации / С. В. Жанказиев, А. И. Воробьев, М. В. Гаврилюк // Транспорт Российской Федерации. – 2021. – № 1–2(92–93). – С. 3–6. – EDN TMFCOL.
- [4] Развитие архитектуры интеллектуальных транспортных систем / Е. О. Андреев, С. В. Жанказиев, В. В. Зырянов, А. С. Павлов // T-Comm: Телекоммуникации и транспорт. – 2024. – Т. 18, № 1. – С. 38–43. – DOI 10.36724/2072-8735-2024-18-1-38-43. – EDN HNTJMK.
- [5] Методология анализа технических и программных элементов как основы эффективного управления национальной сетью интеллектуальных транспортных систем / Р. Д. Давыдов, Г. Д. Ахмадов, М. Ю. Белов, А. И. Панкратов // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). – 2025. – № 4(83). – С. 67–78. – EDN ZDIIZC.
- [6] Давыдов, Р. Д. Методический подход к организации регулярного мониторинга элементов интеллектуальных транспортных систем на автомобильных дорогах общего пользования / Р. Д. Давыдов, Г. Д. Ахмадов, М. В. Жабриков // Интеллектуальный транспорт. – 2026. – Т. 10, № 2(38). – С. 92–106. – DOI 10.24412/3033-6007-2026-238-92-106. – EDN VOMRIM.
- [7] From Data to Actions in Intelligent Transportation Systems: A Prescription of Functional Requirements for Model Actionability / I. Laña, J. J. Sanchez-Medina, E. I. Vlahogianni, J. Del Ser // Sensors. – 2021. – Vol. 21, no. 4. – Art. 1121. – DOI 10.3390/s21041121.
- [8] Duan, R. A comparative study on ITS (intelligent transport system) standardization policies in the U.S. and Europe / R. Duan // Heliyon. – 2023. – Vol. 9, no. 11. – Art. e21310. – DOI 10.1016/j.heliyon.2023.e21310.
- [9] On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness / S. Greco, A. Ishizaka, M. Tasiou, G. Torrisi // Social Indicators Research. – 2019. – Vol. 141. – P. 61–94. – DOI 10.1007/s11205-017-1832-9.
- [10] Särndal, C.-E. Estimation in Surveys with Nonresponse / C.-E. Särndal, S. Lundström. – Chichester : John Wiley & Sons, 2005. – 212 p.

METHODOLOGY FOR ASSESSING COMPLETENESS, RELIABILITY AND QUALITY OF MONITORING DATA FOR INTELLIGENT TRANSPORTATION SYSTEMS INFRASTRUCTURE

Davydov R. D.^{1, a}, Akhmadov G. D.^{1, b}, Zhabrikov M. V.^{1, c}

¹ Federal Autonomous Institution «Russian Road Research Institute» (FAI «ROSDORNII»), Moscow, Russia.

E-mail: ^a davydov@rosdornii.ru, ^b akhmadov@rosdornii.ru, ^c maxpkey@mail.ru

Abstract. The article presents a methodology for quantitative assessment of the completeness, reliability and quality of primary data collected within the framework of monitoring the infrastructure of intelligent transportation systems on public roads of the Russian Federation. The relevance of the study is determined by the need for formalized quality control of source information used for the formation of consolidated ITS equipment inventories and subsequent analysis of the state of digital road infrastructure. The proposed methodology is a tool for verifying the reliability and completeness of primary data, on the basis of which such an analysis is performed. The scientific novelty of the study lies in the development of a system of interconnected criteria and integral indices intended for quantitative assessment of the completeness and reliability of ITS monitoring data, as well as for the formation of a consolidated information quality index. The methodology includes the calculation of partial data quality criteria, the formation of an integral index of data completeness, an integral index of data reliability and an integral information quality index. The proposed approach ensures reproducibility of calculations, comparability of monitoring results and the possibility of subsequent automation of ITS data quality assessment procedures.

Keywords: *intelligent transportation systems, data quality, data completeness, data reliability, ITS monitoring.*

References

- [1] Egorov, S. V., Shatsionok, P. V., Erpyleva, A. I., & Zharkov, D. I. (2022). World and Russian experience in the application of intelligent transport systems. *Transportnoye Delo Rossii*, (2), 130–136. https://doi.org/10.52375/20728689_2022_2_130 (in Russian)
- [2] Solodkiy, A. I. (2020). Development of intelligent transport systems in Russia: Problems and solutions. A new stage. *Intellekt. Innovatsii. Investitsii*, (6), 10–19. <https://doi.org/10.25198/2077-7175-2020-6-10> (in Russian)
- [3] Zhankaziyevev, S. V., Vorobyev, A. I., & Gavriluk, M. V. (2021). Principles of forming a state certification system for elements of intelligent transport systems in the Russian Federation. *Transport Rossiyskoy Federatsii*, (1–2(92–93)), 3–6. (in Russian)

For citation:

Davydov R. D., Akhmadov G. D., Zhabrikov M. V. Methodology for Assessing Completeness, Reliability and Quality of Monitoring Data for Intelligent Transportation Systems Infrastructure. *Intelligent transport*, 2026, no. 3(39), pp. 85–99. DOI: 10.24412/3033-6007-2026-339-85-99. (in Russian)

- [4] Andreyev, E. O., Zhankaziyev, S. V., Zyryanov, V. V., & Pavlov, A. S. (2024). Development of the architecture of intelligent transport systems. *T-Comm: Telekommunikatsii i Transport*, 18(1), 38–43. <https://doi.org/10.36724/2072-8735-2024-18-1-38-43> (in Russian)
- [5] Davydov, R. D., Akhmadov, G. D., Belov, M. Yu., & Pankratov, A. I. (2025). Methodology for analyzing technical and software elements as a basis for effective management of the national network of intelligent transport systems. *Vestnik Moskovskogo Avtomobilno-Dorozhnogo Gosudarstvennogo Tekhnicheskogo Universiteta (MADI)*, (4(83)), 67–78. (in Russian)
- [6] Davydov, R. D., Akhmadov, G. D., & Zhabrikov, M. V. (2026). Methodological approach to organizing regular monitoring of intelligent transport system elements on public roads. *Intelligent Transport*, 10(2(38)), 92–106. <https://doi.org/10.24412/3033-6007-2026-238-92-106> (in Russian)
- [7] Laña, I., Sanchez-Medina, J. J., Vlahogianni, E. I., & Del Ser, J. (2021). From data to actions in intelligent transportation systems: A prescription of functional requirements for model actionability. *Sensors*, 21(4), Article 1121. <https://doi.org/10.3390/s21041121>
- [8] Duan, R. (2023). A comparative study on ITS (intelligent transport system) standardization policies in the U.S. and Europe. *Heliyon*, 9(11), Article e21310. <https://doi.org/10.1016/j.heliyon.2023.e21310>
- [9] Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141, 61–94. <https://doi.org/10.1007/s11205-017-1832-9>
- [10] Särndal, C.-E., & Lundström, S. (2005). *Estimation in surveys with nonresponse*. John Wiley & Sons.

УДК 656.13:004.89

DOI: 10.24412/3033-6007-2026-339-100-114

СОВМЕСТНАЯ РАЗРАБОТКА ДОКУМЕНТОВ ТРАНСПОРТНОГО ПЛАНИРОВАНИЯ И ИНТЕЛЛЕКТУАЛЬНЫХ ТРАНСПОРТНЫХ СИСТЕМ

Павлов Н.В.^{1,а}, Свистельников А.А.^{1,б}, Вальковская А.А.^{1,с}, Платова А.Р.^{1,д}

¹ Федеральное автономное учреждение «Российский дорожный научно-исследовательский институт» (ФАН «РОСДОРНИИ»), Москва, Россия.

E-mail: ^а pavlovnv@rosdornii.ru, ^б SvistelnikovAA@rosdornii.ru, ^с Valkovskaya@rosdornii.ru, ^д PlatovaAR@rosdornii.ru

Аннотация. В статье рассматривается проблема разрозненного применения инструментов транспортного планирования и интеллектуальных транспортных систем (ИТС) в городских агломерациях. Актуальность исследования обусловлена ростом уровня автомобилизации в Российской Федерации на 22% в 2013–2023 гг. и сохранением высокого уровня аварийности: в 2024 г. зарегистрировано 132 037 ДТП, в которых погибли 14 403 человека. Изолированное использование этих инструментов приводит к локальной оптимизации, когда улучшения на одном участке ухудшают ситуацию на смежных. Цель работы – анализ эффективности совместной разработки документов транспортного планирования и локальных проектов ИТС, выявление точек интеграции. Методология основана на системном и нормативно-правовом анализе, эмпирической базой стали данные Росстата, Минтранса России и результаты внедрения ИТС в 62 агломерациях. Формализованы четыре типа точек интеграции (информационная, целевая, оценочная, нормативная). Предложена система ключевых показателей эффективности по критериям безопасности дорожного движения, организации дорожного движения, комфорта и экологии. Установлены нормативные барьеры, включая отсутствие обязательной транспортной модели для начальных уровней зрелости ИТС. Обоснована итерационная схема «модель – проект – данные – модель». Сделан вывод о необходимости институциональных изменений для достижения синергетического эффекта.

Ключевые слова: транспортное планирование, интеллектуальные транспортные системы, городские агломерации, синергетический эффект, ключевые показатели эффективности, транспортное моделирование, нормативные барьеры, безопасность дорожного движения, экологическая эффективность.

Введение

Современное развитие транспортных систем городских агломераций сталкивается с рядом взаимосвязанных вызовов, требующих комплексных решений.

Первый вызов – устойчивый рост уровня автомобилизации. В Российской Федерации за период с 2013 по 2023 год автомобилизация увеличилась на 22%¹. В 2024 году количество транспортных средств, состоящих на го-

Для цитирования:

Павлов Н.В., Свистельников А.А., Вальковская А.А., Платова А.Р. Совместная разработка документов транспортного планирования и интеллектуальных транспортных систем. *Интеллектуальный транспорт*. 2026. № 3(39). С. 100–114. DOI: 10.24412/3033-6007-2026-339-100-114.

¹ Федеральная служба государственной статистики [Электронный ресурс]. URL: <https://rosstat.gov.ru/> (дата обращения: 27.04.2026).

сударственном учете, возросло ещё на 2,7%, достигнув 63,13 млн единиц. Темп прироста оказался наибольшим за последние семь лет [1].

Второй вызов – высокий уровень аварийности. В 2024 году зафиксировано снижение основных показателей дорожно-транспортной аварийности: количество ДТП сократилось на 0,3% (до 132 037), число погибших – на 0,7% (до 14 403), раненых – на 1% (до 164 754). Снижение всех трех показателей отмечено в 23 регионах. Однако в 33 субъектах Российской Федерации и на федеральной территории «Сириус» зафиксирован рост числа погибших [1].

Третий вызов – негативное воздействие транспорта на окружающую среду. Автомобильный транспорт является одним из основных источников выбросов углекислого газа, создающего парниковый эффект. На его долю приходится 23% общего антропогенного загрязнения атмосферного воздуха, причём эта доля продолжает расти [2].

Четвертый вызов связан с появлением и активным внедрением искусственного интеллекта (ИИ) в дорожно-транспортный комплекс Российской Федерации, что обуславливает необходимость оперативного формирования нормативно-технической базы для безопасной и эффективной интеграции ИИ в существующую транспортную инфраструктуру [3, 4].

Состояние проблемы

Несмотря на взаимодополняющий характер, инструменты стратегического транспортного планирования и операционного управления на основе ИТС зачастую применяются разрозненно, без формирования единого контура управления. Однако сами по себе ИТС не способны эффективно функционировать без соответствующего организационного обеспечения, а именно без создания многоуровневой системы центров управления транспортом, интегрирующих задачи транспортного планирования, организации движения, управления парковочным пространством и маршрутным транспортом общего пользования [5]. Как показывают исследования отечественного и международного опыта, именно синергетическое взаимодействие указанных инструментов является критически важным для достижения максимального эффекта [6].

Однако анализ отечественной и зарубежной литературы свидетельствует, что до настоящего времени сохраняется разрыв между стратегическим транспортным планированием и операционным управлением на базе ИТС. В большинстве исследований рассматривается лишь однонаправленное влияние – либо транспортное планирование как поставщик исходных данных для ИТС, либо ИТС как источник информации для калибровки моделей. Международные обзоры подтверждают результативность объединения данных, моделей и интеллектуальных средств управления, однако вопросы их закрепления в едином цикле разработки документов раскрыты недостаточно [9]. Системный анализ двунаправленной синергии с формализацией точек интеграции и оценкой эффективности совместной разработки документов представлен в научной литературе фрагментарно. Указанное противоречие между потребностью в комплексном управлении и разрозненностью инструментов образует научную проблему, для решения которой предпринято настоящее исследование. Его цель – анализ эффективности совместной разработки документов транспортного планирования и локальных проектов

ИТС, а также выявление ключевых точек их интеграции для устойчивого развития транспортных систем городских агломераций.

1. Методы исследования

В основу работы положен комплексный междисциплинарный подход, сочетающий методы системного анализа, нормативно-правового анализа и сравнительного анализа документов транспортного планирования.

Эмпирической базой выступили:

- открытые данные Федеральной службы государственной статистики и ФКУ «НЦ БДД МВД России» за 2013–2024 гг. [1]²;
- нормативные правовые акты и методические рекомендации Минтранса России [7]³;
- нормативно-технические документы, регулирующие общие положения и методы испытаний ИТС и технологии ИИ [3, 8, 9];
- результаты внедрения ИТС в 62 российских агломерациях в рамках федерального проекта «Общесистемные меры развития дорожного хозяйства»⁴.

Для оценки эффективности интеграции применен критериальный подход, основанный на системе ключевых показателей (безопасность дорожного движения, организация дорожного движения, комфорт, экология). Анализ синергетического эффекта выполнен с использованием логико-семантического моделирования связей между документами транспортного планирования и локальными проектами ИТС на разных стадиях их жизненного цикла.

Постановка задачи и предлагаемая последовательность совместной разработки документов транспортного планирования и локального проекта ИТС приведены на рисунке 1.

²Федеральная служба государственной статистики [Электронный ресурс]. URL: <https://rosstat.gov.ru/> (дата обращения: 27.04.2026).

³Распоряжение Министерства транспорта Российской Федерации от 27 апреля 2024 года № АК-95-р «Об утверждении Методических рекомендаций по разработке заявок (включая локальные проекты по созданию и модернизации интеллектуальных транспортных систем) субъектов Российской Федерации на получение субсидий из федерального бюджета бюджетам субъектов Российской Федерации в целях реализации мероприятия «Внедрены интеллектуальные транспортные системы, предусматривающие автоматизацию процессов управления дорожным движением в городских агломерациях, включающих города с населением свыше 300 тысяч человек» в рамках федерального проекта «Общесистемные меры развития дорожного хозяйства» государственной программы Российской Федерации «Развитие транспортной системы».

⁴Министерство транспорта Российской Федерации. «В российских регионах продолжают внедрять интеллектуальные транспортные системы» [Электронный ресурс]. URL: <https://mintrans.gov.ru/press-center/news/11607> (дата обращения: 27.04.2026).

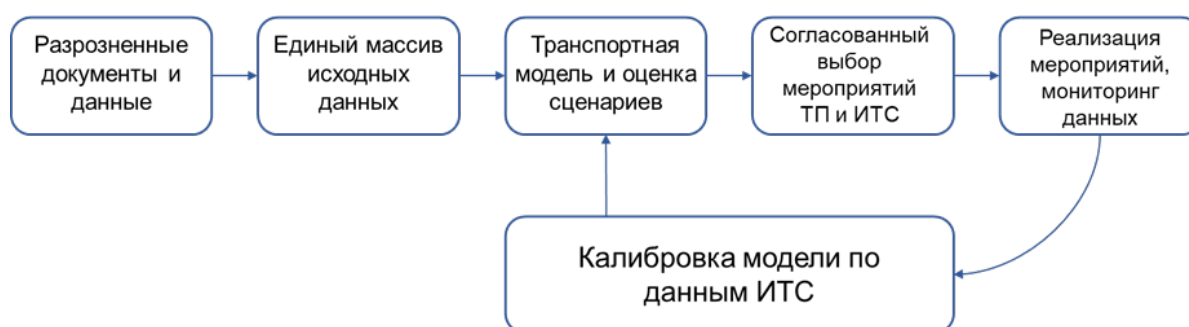


Рис. 1. Постановка задачи и предлагаемая последовательность совместной разработки документов транспортного планирования и локального проекта ИТС

2. Результаты и обсуждение

2.1. Транспортное планирование как фундамент для ИТС

Транспортное планирование представляет собой системный процесс, направленный на управление транспортным спросом и обеспечение устойчивого развития транспортной инфраструктуры. Его первичная задача – не только ликвидация узких мест на улично-дорожной сети (УДС) и мест концентрации ДТП, но и формирование сбалансированной транспортной сети, способной адаптироваться к перспективным изменениям.

Основой для разработки мероприятий в документах транспортного планирования (программа комплексного развития транспортной инфраструктуры (ПКРТИ), региональный комплексный план транспортного обслуживания населения (РКПТО), комплексная схема организации дорожного движения (КСОДД), региональный стандарт транспортного обслуживания населения (РСТО), документ планирования регулярных перевозок (ДПРП) и др.) и документах ИТС (локальные проекты интеллектуальных транспортных систем (ЛПИ ИТС)) являются комплексные транспортные обследования. К ним относятся:

- обследование интенсивности дорожного движения;
- обследование пассажиропотоков;
- социологические обследования подвижности населения.

Данные, полученные в ходе этих обследований, позволяют выявить проблемные зоны, построить матрицы корреспонденций и произвести моделирование существующей транспортной ситуации. Например, параметры интенсивности движения являются исходными данными для настройки алгоритмов светофорного регулирования – элемента подсистемы светофорного управления в составе ИТС.

Для проработки сценариев развития и оценки эффективности предлагаемых решений в транспортном планировании используется программное обеспечение для макро- и микромоделирования:

- **макромоделирование** применяется для анализа транспортной системы агломерации в целом, оценки последствий крупных градостроительных решений и изменения маршрутной сети;
- **микромоделирование** позволяет детально проработать сложные узлы

(перекрестки, развязки), оптимизировать режимы работы светофоров и организацию дорожного движения.

Следовательно, инструментом транспортного планирования является детальная транспортная модель, которая служит оптимальной основой для проектирования любых мероприятий, включая мероприятия ИТС, и позволяет оценить социально-экономическую эффективность всего комплекса решений, а также их влияние на транспортную систему. При этом транспортное планирование выполняет по отношению к ИТС не только информационную, но и нормативно-целевую функцию – формализованные в документах транспортного планирования целевые показатели становятся обязательными требованиями к функциональным характеристикам создаваемых подсистем ИТС. В противном случае возникает эффект, когда внедренные интеллектуальные решения улучшают ситуацию на одном перекрестке, но ухудшают ее на смежных участках.

2.2. Интеллектуальные транспортные системы как инструмент оперативного управления

ИТС представляют собой совокупность технических и программных средств, предназначенных для автоматизации управления транспортными потоками в реальном времени. В отличие от статических решений транспортного планирования (например, установка дорожных знаков), ИТС обеспечивают гибкое и адаптивное воздействие.

Среди подсистем ИТС можно выделить⁵:

- подсистемы косвенного управления – информируют участников движения о транспортных заторах, дорожной обстановке и предлагают альтернативные маршруты (например, через динамические информационные табло);
- подсистемы директивного управления – предписывают действия, не допускающие отклонений, включая автоматическое изменение режимов работы светофорных объектов, приоритизацию общественного транспорта на перекрестках.

Логика работы ИТС основана на цикле «сбор данных – анализ – прогноз – управляющее воздействие» [10].

Датчики и детекторы фиксируют параметры дорожного движения; аналитические модули системы на основе полученных данных строят прогноз развития ситуации. На основе прогноза формируется управляющее воздействие, направленное на предотвращение затора или минимизацию его последствий.

⁵ГОСТ Р 56294–2014, утвержденный Приказом Федерального агентства по техническому регулированию и метрологии от 11 декабря 2014 г. № 1966-ст.

2.3. Синергетический эффект интеграции транспортного планирования и ИТС

В контексте совместной разработки под точками интеграции предлагается понимать структурные элементы межсистемного взаимодействия, обеспечивающие координацию документов транспортного планирования (ПКРТИ, КСОДД, РКПТО) и локальных проектов ИТС на информационном, целевом, оценочном и нормативном уровнях. Их содержание и практический результат представлены в таблице 1.

Таблица 1. Точки интеграции документов транспортного планирования и локальных проектов ИТС

Тип интеграции	Содержание	Практический результат
Информационная	Единые массивы данных об интенсивности движения, пассажиропотоках и корреспонденциях.	Согласованная калибровка транспортной модели и настройка алгоритмов ИТС.
Целевая	Целевые показатели документов транспортного планирования используются как требования при внедрении ИТС.	Предотвращение локальной оптимизации в ущерб смежным участкам улично-дорожной сети.
Оценочная	Фактические данные ИТС используются для мониторинга достижения целевых показателей и последующей валидации и верификации транспортной модели.	Количественная оценка результата и своевременная корректировка решений.
Нормативная	Закрепление требований к применению транспортной модели и обмену данными в нормативно-технических документах.	Поддержание единого жизненного цикла планирования, реализации и мониторинга.

Взаимосвязь транспортного планирования и ИТС не является односторонней. Для агломерации a и периода наблюдения t синергетический эффект предлагается определять следующим образом:

$$\mathcal{E}_{\text{совм}}(a, t) = \mathcal{E}_{\text{ТП}}(a, t) + \mathcal{E}_{\text{ИТС}}(a, t) + \Delta(a, t), \quad (1)$$

где $\mathcal{E}_{\text{совм}}(a, t)$ – эффективность при совместной разработке документов транспортного планирования (ТП) и ИТС; $\mathcal{E}_{\text{ТП}}(a, t)$ – эффективность только мероприятий транспортного планирования; $\mathcal{E}_{\text{ИТС}}(a, t)$ – эффективность только интеллектуальных транспортных систем; $\Delta(a, t)$ – добавочный синергетический эффект, который возникает, только когда мероприятия ТП и ИТС применяются вместе, учитывая друг друга.

Влияние транспортного планирования на ИТС.

1. Внедрение ИТС – капиталоемкий процесс. Документы транспортного планирования предоставляют необходимые данные для технико-экономического обоснования проектов ИТС, демонстрируя ожидаемый эффект от комплексного внедрения мероприятий транспортной инфраструктуры, организации дорожного движения и ИТС;

2. Транспортное планирование позволяет выявить проблемы на улично-дорожной сети, где внедрение ИТС принесет наибольший социально-экономический эффект;
3. Наличие утвержденных документов транспортного планирования становится конкурентным преимуществом при подаче заявок на государственное финансирование⁶.

Влияние ИТС на транспортное планирование.

1. Данные, собираемые ИТС в режиме реального времени, позволяют уточнять и калибровать транспортные модели, используемые в документах транспортного планирования, повышая их точность и прогнозную эффективность;
2. ИТС предоставляют фактические данные о ключевых показателях эффективности после реализации тех или иных мероприятий, предложенных в документах транспортного планирования;
3. Транспортное планирование задает долгосрочную стратегию, а ИТС позволяют гибко адаптировать организацию дорожного движения к текущим условиям, включая дорожные инциденты [11].

2.4. Нормативные и организационные барьеры интеграции

Несмотря на доказанную синергию, практическая совместная разработка документов транспортного планирования и ИТС наталкивается на ряд системных ограничений:

1. Действующая редакция Методических рекомендаций по разработке заявок на получение межбюджетных субсидий⁷ не требует наличия утвержденной транспортной модели для агломераций, претендующих на первый и второй уровни «зрелости» ИТС. Это легитимизирует проектирование без оценки сетевых эффектов;
2. Организационно центры управления транспортом и центры транспортного планирования в большинстве регионов функционируют изолированно, без единого регламента обмена данными;
3. Отсутствует обязательная процедура обратной калибровки транспортных моделей по данным детекторов транспорта, что нарушает цикл «модель – проект – данные – модель». Преодоление указанных барьеров требует не технологических, а институциональных решений – внесения изменений в нормативную правовую и нормативно-техническую базу.

2.5. Рекомендуемый механизм калибровки транспортной модели

Калибровка (обратная калибровка) должна выполняться как регламентированная процедура. Текущий контроль полноты и качества данных детекторов целесообразно проводить ежемесячно, плановую калибровку – не реже одного

⁶Распоряжение Министерства транспорта Российской Федерации от 27 апреля 2024 г. № АК-95-р.

⁷Распоряжение Министерства транспорта Российской Федерации от 27 апреля 2024 г. № АК-95-р.

раза в квартал, а внеплановую – после существенных изменений схемы организации движения, маршрутной сети или транспортного спроса:

- на **первом** этапе проверяются геометрия сети, режимы регулирования и качество исходных данных;
- на **втором** корректируются параметры пропускной способности и скорости;
- на **третьем** уточняются матрицы корреспонденций методом их ограниченной оценки по данным детекторов с сохранением общего объема поездок.

В качестве ориентировочных критериев приемлемости могут использоваться значение показателя GEN (статистической меры согласованности расчетной и наблюдаемой интенсивности транспортного потока) менее 5 не менее чем для 85% сопоставляемых транспортных потоков и отклонение расчетного времени поездки от наблюдаемого не более чем на 15% не менее чем для 85% маршрутов [12]. После корректировки модель проверяется на независимом периоде наблюдений, при невыполнении критериев цикл повторяется. Конкретная периодичность и допустимые отклонения должны закрепляться в техническом задании с учетом масштаба модели и качества данных ИТС.

2.6. Ключевые показатели эффективности

Для объективной оценки синергетического эффекта от совместной разработки документов транспортного планирования и ИТС необходима система количественно измеримых показателей, охватывающих все значимые аспекты функционирования транспортной системы городской агломерации. В настоящем исследовании в качестве базовых выделены четыре интегральные группы: безопасность дорожного движения, организация дорожного движения, комфорт и экология, что отражает приоритеты государственной транспортной политики (снижение смертности, повышение мобильности, уменьшение нагрузки на среду)⁸ и позволяет комплексно оценивать как операционные решения ИТС, так и стратегические мероприятия транспортного планирования.

Безопасность дорожного движения

Основными индикаторами выступают абсолютное число ДТП, количество погибших и раненых, а также расчетные показатели – социальный риск и транспортный риск. Совместное применение инструментов транспортного планирования и ИТС позволяет воздействовать на безопасность как пассивными мерами (определение мест концентрации ДТП с последующим проектированием транспортной инфраструктуры) [13], так и активными (автоматическое снижение максимально разрешенной скорости при ухудшении погодных условий).

⁸Указ Президента РФ от 14.11.2025 № 841 «Об утверждении Стратегии повышения безопасности дорожного движения в Российской Федерации на период до 2030 года и на перспективу до 2036 года».

Организация дорожного движения

Организация дорожного движения направлена на снижение потерь времени (задержек) при обеспечении безопасности⁹. Постановление Правительства устанавливает систему параметров эффективности: средняя задержка транспортных средств, временной индекс, уровень обслуживания (А–F), показатель перегруженности дорог и буферный индекс¹⁰. Интеграция транспортного планирования и ИТС позволяет целенаправленно воздействовать на эти показатели. Документы транспортного планирования (КСОДД, ПКРТИ) задают целевые уровни обслуживания и предельные задержки для магистральной сети, а ИТС в реальном времени позволяют осуществлять управляющее воздействие на светофорные циклы, динамические ограничения скорости и маршрутизацию, удерживая фактические значения в заданных пределах.

Без двунаправленной связи эффект от внедрения ИТС остается локальным. Только совместная разработка документов транспортного планирования и ИТС, учитывающая все параметры эффективности организации дорожного движения, позволяет обеспечить устойчивое снижение перегруженности УДС и предсказуемость времени поездки в масштабах агломерации.

Комфорт участников движения

Комфорт – комплексный показатель, отражающий качество транспортного обслуживания и степень удовлетворенности поездкой, он включает:

- регулярность и предсказуемость времени поездки;
- задержки на перекрестках;
- плотность потока и буферный индекс;
- для общественного транспорта – наполнение салона и соблюдение расписания.

Операционально он оценивается по совокупности объективных показателей в соответствии с социальным стандартом транспортного обслуживания населения и опросных данных по оценке удовлетворенности участников движения по стандартизированной пятибалльной шкале.

Интеграция ИТС позволяет снижать нерегулярность поездок, что повышает комфорт.

Экологическая эффективность

Ключевым индикатором выступает объем выбросов парниковых газов (CO₂), а также оксидов азота (NO_x) и твердых частиц (PM). Прямое влияние оказывают скорость и режим движения, а интеграция ИТС (координация светофоров, управление скоростью и др.) позволяет сглаживать неравномерность транспортного потока.

Приведенная система показателей позволяет оценить совокупный результат совместной разработки документов транспортного планирования и

⁹Федеральный закон «Об организации дорожного движения в Российской Федерации и о внесении изменений в отдельные законодательные акты Российской Федерации» от 29.12.2017 № 443-ФЗ.

¹⁰Постановление Правительства РФ от 16.11.2018 № 1379 «Об утверждении Правил определения основных параметров дорожного движения и ведения их учета». – С. 9–10.

ИТС при условии единого порядка нормирования и агрегирования.

Для показателей, снижение которых является положительным результатом, нормированное изменение определяется как:

$$k_i = \frac{(X_0 - X_1)}{X_0}, \quad (2)$$

где k_i – относительное изменение показателя; X_0 – значение показателя в базовом периоде до реализации мероприятий; X_1 – значение показателя в расчетном периоде после реализации мероприятий.

Для показателей, рост которых является положительным результатом, нормированное изменение определяется как:

$$k_i = \frac{(X_1 - X_0)}{X_0}. \quad (3)$$

Внутри каждой группы рассчитывается среднее значение K_j , после чего формируется интегральный показатель эффективности:

$$E = \sum w_j K_j, \quad \sum w_j = 1, \quad (4)$$

где E – интегральный показатель эффективности совместной реализации мероприятий; w_j – весовой коэффициент j -й группы показателей; K_j – нормированное значение j -й группы показателей.

Рекомендуемые весовые коэффициенты групп показателей эффективности для различных профилей (профиль определяется исходя из анализа существующего положения и массивов данных) агломераций приведены в таблице 2.

Таблица 2. Рекомендуемые весовые коэффициенты групп показателей эффективности для различных профилей агломераций

Группа ключевых показателей эффективности	Сбалансированный профиль	Повышенная аварийность	Высокая перегруженность
Безопасность дорожного движения	0,30	0,40	0,25
Организация дорожного движения	0,30	0,25	0,40
Комфорт	0,20	0,15	0,20
Экология	0,20	0,20	0,15

Пример расчета: для сбалансированного профиля при улучшении показателей безопасности на 10%, организации движения на 12%, комфорта на 8% и экологической эффективности на 6% интегральный результат составит $E = 0,30 \cdot 0,10 + 0,30 \cdot 0,12 + 0,20 \cdot 0,08 + 0,20 \cdot 0,06 = 0,094$, или 9,4%. Пример иллюстрирует порядок расчета, фактические значения должны определяться по данным конкретной агломерации.

2.7. Партнерство технологий как фактор успешного функционирования дорожно-транспортного комплекса

Технологии ИИ, ИТС и математическое моделирование по отдельности уже доказали свою эффективность [9]:

- ИИ обеспечивает анализ больших массивов данных;
- ИТС позволяют управлять дорожно-транспортной инфраструктурой;
- моделирование дает возможность оценивать различные сценарии и прогнозировать эффективность проектных решений.

Однако максимальный эффект достигается только при их комплексной интеграции. Партнерство этих трёх компонентов создает непрерывный цикл «модель – проект – данные – модель», обеспечивая переход к проактивному, безопасному и эффективному управлению транспортными потоками. Государственная политика [14]¹¹ предполагает активное внедрение ИТС в городских агломерациях, однако действующая нормативная база определяет транспортную модель как обязательный элемент лишь для третьего уровня зрелости ИТС, допуская на начальных этапах проектирование без комплексного моделирования системных эффектов.

Такой подход приводит к инвестициям в ЛП ИТС без предварительной оценки их влияния на транспортные потоки в масштабах всей агломерации. Ключевой недостаток – игнорирование матриц корреспонденций, определяющих пространственное распределение транспортного спроса. В результате внедряемые решения могут улучшать локальные показатели, но одновременно вызывать затруднения на смежных участках сети. В качестве альтернативы предложена итерационная схема [7]: сначала создается транспортная модель для разработки оптимального проекта организации движения, затем, после реализации проекта, данные с датчиков ИТС используются для калибровки и актуализации модели, что замыкает цикл «модель – проект – данные – модель» и позволяет гибко настраивать проектные мероприятия под реальную дорожно-транспортную ситуацию.

Главное преимущество предлагаемого подхода – превращение транспортной модели в ключевой инструмент объективной оценки управленческих решений и инвестиций в ИТС. В долгосрочной перспективе создание и постоянная актуализация транспортной модели становятся обязательным условием для перехода к динамическому управлению движением в режиме реального времени, способному прогнозировать развитие ситуации и реагировать на изменение спроса.

Заключение

В результате проведенного исследования подтверждено, что транспортное планирование и ИТС являются не альтернативными, а взаимодополняющими инструментами управления транспортными системами городских агломераций. Их изолированное применение приводит к эффекту, когда улучшение

¹¹Распоряжение Министерства транспорта Российской Федерации от 27 апреля 2024 г. № АК-95-р; Указ Президента Российской Федерации от 7 мая 2024 г. № 309 «О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года».

на отдельном участке сопровождается ухудшением на смежных участках дорожной сети, что снижает общую эффективность принимаемых решений.

Одним из ключевых научных результатов работы стала формализация четырех типов точек интеграции между документами транспортного планирования (ПКРТИ, КСОДД, РКПТО) и ИТС (ЛП ИТС). К ним относятся информационная интеграция (общие данные транспортных обследований), целевая (плановые показатели эффективности), оценочная (мониторинг фактических результатов) и нормативная (взаимоувязывающие требования нормативно-технических документов). Выделение этих связей позволяет перейти от разрозненного управления к единому контуру стратегического и операционного уровней.

Для количественной оценки синергетического эффекта в статье предложена система ключевых показателей эффективности, объединяющая четыре приоритетных критерия:

- безопасность дорожного движения;
- организацию дорожного движения;
- комфорт;
- экологию.

В отличие от традиционных подходов, ориентированных преимущественно на безопасность или пропускную способность, данная система позволяет комплексно оценивать социально-экономические и экологические последствия совместной реализации мероприятий транспортного планирования и ИТС. Практическое применение системы требует калибровки весовых коэффициентов в зависимости от приоритетов конкретной агломерации.

В работе также установлены системные барьеры, препятствующие полноценной интеграции. Среди них – отсутствие обязательной транспортной модели для первых двух уровней «зрелости» ИТС согласно действующим методическим рекомендациям¹², изолированное функционирование центров управления транспортом и центров транспортного планирования в большинстве регионов, а также разрыв цикла обратной калибровки моделей по данным детекторов транспорта. Преодоление этих барьеров требует институциональных решений, включая внесение изменений в нормативную правовую и нормативно-техническую базу.

Практическая значимость работы заключается в обосновании итерационной схемы «модель – проект – данные – модель», которая замыкает контур управления на основе реальных данных детекторов транспорта. Внедрение данной схемы позволяет превратить транспортную модель в ключевой инструмент объективной оценки управленческих решений и инвестиций в ИТС, а в долгосрочной перспективе – обеспечить переход к динамическому управлению дорожным движением в режиме реального времени с прогнозированием развития транспортной ситуации.

¹²Распоряжение Министерства транспорта Российской Федерации от 27 апреля 2024 г. № АК-95-р.

Список литературы

- [1] Дорожно-транспортная аварийность в Российской Федерации в 2024 году : Информационно-аналитический обзор / К. С. Баканов, П. В. Ляхов, А. С. Айсанов [и др.]. – Москва : ФКУ «НИЦ БДД МВД России», 2025. – 148 с. – ISBN 978-5-6046954-5-6. – EDN ITIUJH.
- [2] Майборода, О. В. Анализ путей уменьшения выбросов CO₂ транспортными средствами / О. В. Майборода, А. Е. Чебышев, Н. В. Павлов // Автомобильные дороги. – 2023. – № 4(1097). – С. 108–112. – EDN RCKOGC.
- [3] Свистельников, А. А. Тенденции стандартизации: разработка и применение нормативных правовых и нормативно технических документов в сфере искусственного интеллекта в дорожно-транспортном комплексе / А. А. Свистельников, Н. В. Павлов, А. А. Вальковская // ИТ-Стандарт. – 2025. – № 4(45). – С. 50–57. – EDN DJFKDV.
- [4] Свистельников, А. А. Искусственный интеллект в дорожно-транспортном комплексе / А. А. Свистельников, Н. В. Павлов, А. А. Вальковская // Безопасные и качественные дороги. – 2025. – № 5(34). – С. 42–44.
- [5] Солодкий, А. И. Развитие интеллектуальных транспортных систем в России: проблемы и пути их решения. Новый этап / А. И. Солодкий // Интеллект. Инновации. Инвестиции. – 2020. – № 6. – С. 10–19. – DOI 10.25198/2077-7175-2020-6-10. – EDN PDUSFR.
- [6] Метелкин, П. В. Влияние ИТС на безопасность дорожного движения и развитие регионов / П. В. Метелкин // Системы безопасности. – 2025. – № 1(181). – С. 40–43.
- [7] Белов, М. Ю. Цифровизация управления транспортными потоками – партнерство технологий моделирования и реализации / М. Ю. Белов, А. А. Свистельников, Н. В. Павлов // Автомобильные дороги. – 2025. – № 12(1129). – С. 100–103. – EDN QJNEMD.
- [8] Павлов, Н. В. Эволюция архитектуры интеллектуальных транспортных систем в контексте цифровой трансформации транспортного комплекса Российской Федерации / Н. В. Павлов, А. В. Солдатенков, А. Р. Платова // ИТ-Стандарт. – 2026. – № 1(46). – С. 11–19. – EDN EBHDNV.
- [9] Mahrez, Z. Smart Urban Mobility: When Mobility Systems Meet Smart Data / Z. Mahrez, E. Sabir, E. Badidi [et al.] // IEEE Transactions on Intelligent Transportation Systems. – 2022. – Vol. 23, No. 7. – P. 6222–6239. – DOI 10.1109/TITS.2021.3084907.
- [10] Маркелов, В. М. Интеллектуальные транспортные системы как инструмент управления / В. М. Маркелов, И. В. Соловьев, В. Я. Цветков // Государственный советник. – 2014. – № 3(7). – С. 42–49. – EDN SQYTIT.
- [11] Жанказиев, С. В. Автоматизированный инцидент-менеджмент как инструмент снижения времени реагирования на дорожно-транспортные происшествия / С. В. Жанказиев, Д. Н. Кондык // Вестник Московского автомобильно-дорожного государственного технического университета (МАДИ). – 2026. – № 1(84). – С. 42–50. – EDN AZEYAN.
- [12] Department for Transport. TAG Unit M3.1: Highway Assignment Modelling. – London : Department for Transport, 2026. – 119 p. – URL: <https://www.gov.uk/government/publications/tag-unit-m3-1-highway-assignment-modelling> (дата обращения: 30.07.2026).
- [13] Принципы доступности, безопасности и комфорта в проектировании трамвайных узлов Москвы / Н. В. Павлов, А. Р. Платова, А. Е. Чекмарев, Д. Г. Плотников // Московский транспорт. Наука и проектирование. – 2026. – № 1(5). – С. 8–18. – EDN NKFAXV.
- [14] Жабриков, М. В. Прогноз развития ИТС России до 2030 года / М. В. Жабриков, Р. Д. Давыдов, Г. Д. Ахмадов // Автомобильные дороги. – 2026. – № 2(1131). – С. 94–101. – EDN YABAQU.

JOINT DEVELOPMENT OF TRANSPORT PLANNING DOCUMENTS AND INTELLIGENT TRANSPORT SYSTEMS

Pavlov N. V.^{1, a}, Svistelnikov A. A.^{1, b}, Valkovskaya A. A.^{1, c}, Platova A. R.^{1, d}

¹ Federal Autonomous Institution «Russian Road Research Institute» (FAI «ROSDORNII»), Moscow, Russia.

E-mail: ^a pavlovnv@rosdornii.ru, ^b SvistelnikovAA@rosdornii.ru, ^c Valkovskaya@rosdornii.ru, ^d PlatovaAR@rosdornii.ru

Abstract. The article examines the problem of the fragmented application of transport planning tools and Intelligent Transport Systems (ITS) in urban agglomerations. The relevance of the study is driven by a 22% increase in the motorization rate in the Russian Federation between 2013 and 2023, alongside persistently high accident rates – in 2024, 132,037 road traffic accidents were recorded, resulting in 14,403 fatalities. The isolated use of these tools leads to local optimization, where improvements in one section worsen the situation on adjacent sections. The aim of this work is to analyze the effectiveness of the joint development of transport planning documents and local ITS projects, and to identify integration points. The methodology is based on systemic and legal and regulatory analysis, with the empirical foundation comprising data from Rosstat, the Ministry of Transport of the Russian Federation, and the results of ITS implementation in 62 agglomerations. Four types of integration points are formalized (informational, target-based, evaluative, and regulatory). A system of key performance indicators is proposed, covering criteria for road safety, traffic management, comfort, and environmental impact. Regulatory barriers are identified, including the absence of a mandatory transport model for the initial maturity levels of ITS. An iterative scheme «model – project – data – model» is substantiated. The conclusion is drawn that institutional changes are necessary to achieve a synergistic effect.

Keywords: *transport planning, intelligent transport systems, urban agglomerations, synergistic effect, key performance indicators, transport modeling, regulatory barriers, road safety, environmental performance.*

References

- [1] Bakanov, K. S., Lyakhov, P. V., Aysanov, A. S., et al. (2025). *Road traffic accidents in the Russian Federation in 2024: Information and analytical review*. FKU “NTS BDD MVD Rossii”. (in Russian)
- [2] Mayboroda, O. V., Chebyshev, A. E., & Pavlov, N. V. (2023). Analysis of ways to reduce CO₂ emissions by vehicles. *Avtomobilnyye Dorogi*, 4(1097), 108–112. (in Russian)
- [3] Svistelnikov, A. A., Pavlov, N. V., & Valkovskaya, A. A. (2025). Standardization trends: Development and application of regulatory legal and normative technical documents in the field of artificial intelligence in the road transport complex. *IT-Standart*, 4(45), 50–57. (in Russian)
- [4] Svistelnikov, A. A., Pavlov, N. V., & Valkovskaya, A. A. (2025). Artificial intelligence in the road transport complex. *Bezopasnyye i Kachestvennyye Dorogi*, 5(34), 42–44. (in Russian)

For citation:

Pavlov N. V., Svistelnikov A. A., Valkovskaya A. A., Platova A. R. Joint Development of Transport Planning Documents and Intelligent Transport Systems. *Intelligent transport*, 2026, no. 3(39), pp. 100–114. DOI: 10.24412/3033-6007-2026-339-100-114. (in Russian)

- [5] Solodkiy, A. I. (2020). Development of intelligent transport systems in Russia: Problems and solutions. A new stage. *Intellekt. Innovatsii. Investitsii*, 6, 10–19. <https://doi.org/10.25198/2077-7175-2020-6-10> (in Russian)
- [6] Metelkin, P. V. (2025). Impact of ITS on road traffic safety and regional development. *Sistemy Bezopasnosti*, 1(181), 40–43. (in Russian)
- [7] Belov, M. Yu., Svistelnikov, A. A., & Pavlov, N. V. (2025). Digitalization of traffic flow management – Partnership of modeling and implementation technologies. *Avtomobilnyye Dorogi*, 12(1129), 100–103. (in Russian)
- [8] Pavlov, N. V., Soldatenkov, A. V., & Platova, A. R. (2026). Evolution of the architecture of intelligent transport systems in the context of digital transformation of the transport complex of the Russian Federation. *IT-Standart*, 1(46), 11–19. (in Russian)
- [9] Mahrez, Z., Sabir, E., Badidi, E., Saad, W., & Sadik, M. (2022). Smart urban mobility: When mobility systems meet smart data. *IEEE Transactions on Intelligent Transportation Systems*, 23(7), 6222–6239. <https://doi.org/10.1109/TITS.2021.3084907>
- [10] Markelov, V. M., Solovyev, I. V., & Tsvetkov, V. Ya. (2014). Intelligent transport systems as a management tool. *Gosudarstvennyy Sovetnik*, 3(7), 42–49. (in Russian)
- [11] Zhankaziyev, S. V., & Kondyk, D. N. (2026). Automated incident management as a tool for reducing response time to road traffic accidents. *Vestnik Moskovskogo Avtomobilno-Dorozhnogo Gosudarstvennogo Tekhnicheskogo Universiteta (MADI)*, 1(84), 42–50. (in Russian)
- [12] Department for Transport. (2026). *TAG unit M3.1: Highway assignment modelling*. Retrieved July 30, 2026, from <https://www.gov.uk/government/publications/tag-unit-m3-1-highway-assignment-modelling>
- [13] Pavlov, N. V., Platova, A. R., Chekmarev, A. E., & Plotnikov, D. G. (2026). Principles of accessibility, safety and comfort in the design of tram junctions in Moscow. *Moskovskiy Transport. Nauka i Proyektirovaniye*, 1(5), 8–18. (in Russian)
- [14] Zhabrikov, M. V., Davydov, R. D., & Akhmadov, G. D. (2026). Forecast of ITS development in Russia until 2030. *Avtomobilnyye Dorogi*, 2(1131), 94–101. (in Russian)

СОВРЕМЕННОЕ СОСТОЯНИЕ СИСТЕМ ИНТЕЛЛЕКТУАЛИЗАЦИИ, ЦИФРОВИЗАЦИИ И РОБОТИЗАЦИИ ОТЕЧЕСТВЕННОГО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

Бочков А.В.^{1,а}, Долгий А.И.^{1,б}, Зажигалкин А.В.^{1,2,с}, Розенберг Е.Н.^{1,д},
Суханов А.В.^{1,е}, Хатламаджиян А.Е.^{1,ф}, Чернуха А.В.^{1,г}

¹ Акционерное общество «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте» (АО «НИИАС»), Москва, Россия

² Ассоциация «Транспортная наука», Москва, Россия

E-mail: ^а a.bochkov@vnias.ru, ^б a.dolgiy@vnias.ru, ^с zashigalkin@mail.ru, ^д S.Lavruhina@vnias.ru,
^е a.suhanov@vnias.ru, ^ф a.hatlamadzhian@vnias.ru, ^г a.chernuha@vnias.ru

Аннотация. В работе систематизированы результаты научно-практической конференции «PRO//Движение. Транспортные системы управления» (18–20 августа 2026 г., г. Нижний Новгород), посвященной цифровизации перевозочного процесса, искусственному интеллекту и роботизации железнодорожного транспорта. В конференции приняли участие представители руководства холдинга «Российские железные дороги», научного отраслевого комплекса, а также производственных предприятий – руководители, ученые и инженеры. Рассмотрены целевая модель управления грузопотоками, проект «Цифровая железнодорожная станция», технологии виртуальной сцепки, технического зрения, прогнозной диагностики и физического искусственного интеллекта. Особое внимание уделено роли научной деятельности АО «НИИАС» в формировании единой архитектуры данных, моделей и исполнительных систем, а также вопросам функциональной безопасности, экономической эффективности и тиражирования разработок. Обсуждены перспективы развития искусственного интеллекта в железнодорожной отрасли.

Ключевые слова: железнодорожный транспорт, цифровая станция, управление движением, искусственный интеллект, техническое зрение, робототехника, цифровой двойник, функциональная безопасность.

Введение: контекст и значение конференции

Состоявшаяся в Нижнем Новгороде международная научно-практическая конференция и выставка «PRO//Движение. Транспортные системы управления»¹ стала ключевым отраслевым событием года, собрав более 600 участников – представителей руководства ОАО «РЖД», ведущих научно-исследовательских институтов, производителей оборудования и разработчиков систем управления движением.

В выступлении первого заместителя генерального директора ОАО «РЖД»

Для цитирования:

Бочков А.В., Долгий А.И., Зажигалкин А.В., Розенберг Е.Н., Суханов А.В., Хатламаджиян А.Е., Чернуха А.В. Современное состояние систем интеллектуализации, цифровизации и роботизации отечественного железнодорожного транспорта. *Интеллектуальный транспорт*. 2026. № 3(39). С. 115–158. DOI: 10.24412/3033-6007-2026-339-115-158.

¹<https://tcs-forum.ru/> [Дата последнего обращения: 3.09.2026]

Сергея Алексеевича Кобзева к приоритетным задачам отрасли были отнесены повышение скоростей движения, точности перевозок и уровня комфорта пассажиров. Решение указанных задач требует совершенствования методов управления перевозочным процессом и координации деятельности производственных и научно-исследовательских организаций.

В рамках конференции работало восемь тематических сессий, охватывающих все аспекты цифровой трансформации железнодорожного транспорта: от фундаментальных вопросов управления движением до применения искусственного интеллекта и роботизации. С учетом рубрик сетевого научно-практического журнала «Интеллектуальный транспорт» наибольший научный интерес представляют сессии, на которых обсуждались стратегические направления развития систем управления, интеллектуальные технологии, цифровое управление перевозочным процессом и роботизация производственных процессов. Современные системы интервального регулирования, беспилотные системы и вопросы функциональной безопасности, затронутые на отдельных сессиях конференции, заслуживают самостоятельной обзорной работы. Настоящий обзор систематизирует ключевые тезисы этих сессий.

Материалы конференции целесообразно рассматривать в разрезе трех взаимосвязанных элементов единого контура управления перевозочным процессом. Первый элемент образуют средства получения достоверной цифровой информации: напольные и бортовые устройства, системы счета осей, диагностики, технического зрения и мониторинга. Второй элемент включает модели планирования и поддержки решений: автоматизированную систему ведения графика исполненного движения нового поколения (АС ГИД НП), «Альфа-Модель», цифровые двойники и интеллектуальные алгоритмы прогнозирования. Третий элемент связан с исполнением сформированных заданий: автоматическим ведением поездов, цифровой централизацией, роботизированными комплексами и автономными локомотивами. Именно замыкание этих элементов в единый контур отличает цифровую трансформацию от простой установки разрозненных технических средств [1–3].

Общим для всех сессий вопросом, отдельно рассмотренным на завершающей сессии конференции, являлось обеспечение доверия к автоматизированным решениям. Для железнодорожного транспорта недостаточно показать высокую точность алгоритма в лабораторных условиях. Необходимо доказать полноту и своевременность данных, объяснимость рекомендаций, устойчивость связи, функциональную безопасность интеллектуальной системы, возможность резервного управления и экономическую целесообразность на всем жизненном цикле. Поэтому в докладах постоянно пересекались инженерная, информационная, организационная и нормативная проблематика доверия к системам управления транспортом. Результативность внедрения определяется согласованием указанных составляющих в рамках единой архитектуры, а не характеристиками отдельного технического решения. Здесь стоит отдельно отметить работы под руководством Е. Розенберга, И. Шубинского и К. Бочкова, на которые неоднократно ссылались при обсуждении вопроса безопасности перевозочного процесса [4–7].

Настоящая статья построена следующим образом. В разделе 1 рассмотрены стратегические направления развития систем управления железнодорожным транспортом, обозначенные на пленарном заседании конференции. Раздел 2 обобщает результаты транспортной интеллектуализации, раздел 3 –

цифровизации, раздел 4 – роботизации. В завершающем разделе даны выводы и дальнейшие направления исследований и разработок в области управления отечественным железнодорожным транспортом.

1. Стратегические приоритеты развития систем управления движением

1.1. Общая характеристика пленарного заседания

Пленарное заседание обозначило стратегические приоритеты развития интеллектуализации и роботизации систем управления движением на ближайшие годы. В центре внимания оказались проекты, направленные на повышение пропускной и провозной способности железнодорожной сети, снижение эксплуатационных затрат и повышение безопасности перевозочного процесса. Ключевым тезисом, прозвучавшим в выступлениях, стал переход от фрагментарной автоматизации отдельных процессов к созданию единой цифровой экосистемы управления перевозками.

Помимо первого заместителя генерального директора ОАО «РЖД» в пленарной сессии приняли участие министр транспорта и автомобильных дорог Нижегородской области Л. Самухин, управляющий директор дивизиона «Железнодорожная автоматика и телемеханика» группы компаний (ГК) «Нацпроектстрой» К. Хромушкин, генеральный директор АО «НИИАС» А. Долгий, управляющий директор по развитию интеллектуальных систем управления АО «Трансмашхолдинг» (ТМХ) А. Романчиков, президент ГК «Промэлектроника» И. Тильк и председатель совета директоров Курского электроаппаратного завода (КЭАЗ) А. Канунников. Модератором сессии выступила ведущая РЖД ТВ М. Ожерельева.

1.2. Технологии управления движением на Восточном полигоне и ВСМ

Значительное внимание на пленарной сессии было уделено Восточному полигону – стратегически важному направлению, которое обеспечивает связь с восточными регионами страны и внешними рынками. Сергей Алексеевич Кобзев в своем докладе привел показатели интенсивности перевозок: грузонапряженность Забайкальской железной дороги составляет более 180 млн тонн в год, что является рекордным показателем в мировой практике. Каждые 7 минут по Транссибу идет грузовой поезд, что создает серьезную нагрузку на инфраструктуру.

Для решения задачи обслуживания инфраструктуры в условиях такой интенсивности движения была разработана технология виртуальной сцепки [8]. Эта технология базируется на ранее разработанных и апробированных системах автоведения грузовых поездов. В пассажирском движении автоведение применяется уже несколько десятков лет, но для грузового движения потребовалось сочетание не только «умного» локомотива, но и «умных» устройств автоматики и телемеханики, а также согласование движения нескольких поездов одновременно. Международные исследования связывают виртуальную сцепку и подвижный блок с изменением принципов интервального регулирования, требований к связи и оценки безопасности в отказных режимах [9].

Сергей Алексеевич Кобзев сообщил, что были проведены опытные

поездки с пятью поездами, и сейчас есть техническая возможность сократить интервал между поездами до минимума, обеспечив синхронное вождение. Эта технология уже работает в штатном режиме на Восточном полигоне. Более того, генеральный директор ОАО «РЖД» Олег Валентинович Белозеров анонсировал применение этой технологии в пассажирском движении на направлении Москва – Юг, где также необходимо сокращение интервалов попутного следования.

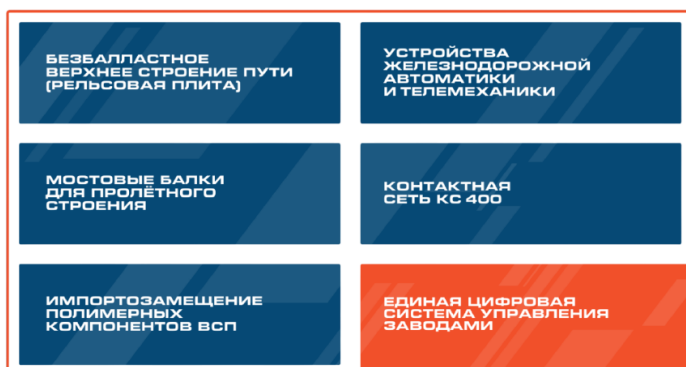
Особое место в дискуссии заняла тема строительства первой в России высокоскоростной магистрали (ВСМ) Москва – Санкт-Петербург. К. Хромушкин рассказал, что «Нацпроектстрой» является одним из ключевых поставщиков систем управления и автоматики для этого проекта (рис. 1). ВСМ задает новые отраслевые стандарты точности и надежности, а все технические решения, применяемые на магистрали, являются полностью отечественными.

РАБОТАЕМ НАД ПРОЕКТАМИ НАЦИОНАЛЬНОГО ЗНАЧЕНИЯ

НПС // ЖЕЛЕЗНОДОРОЖНАЯ
АВТОМАТИКА
И ТЕЛЕМЕХАНИКА

МОСКВА — САНКТ-ПЕТЕРБУРГ. ВСМ

Высокоскоростная железная дорога с движением до 400 км/ч
на участке в 676 км. Минимальное время в пути — 2 часа 15 минут



Дивизион ЖАТ ГК НПС | 2026

Рис. 1. Реализация проектов ВСМ дивизиона «Железнодорожная автоматика и телемеханика» ГК «Нацпроектстрой»

Проект ВСМ предусматривает разработку новых стандартов, материалов и систем организации движения поездов. Участие российских компаний создает условия для накопления компетенций и разработки решений, потенциально применимых на железнодорожных сетях колеи 1520 мм. Сравнение отечественных и зарубежных решений требует оценки по единым техническим и эксплуатационным критериям.

1.3. Новая парадигма управления перевозочным процессом

Генеральный директор АО «НИИАС» А. Долгий в своем выступлении представил видение новой целевой модели управления, которая должна прийти на смену существующим подходам. Современные транспортные системы управления – это не только системы исполнительных устройств, автоматики и диспетчеризации, как это было еще 5–10 лет назад. Сегодня это цифровые комплексы, включая верхнеуровневые системы планирования,

контроля и взаимодействия с различными транспортными системами других отраслей.

Сложившаяся система ключевых показателей эффективности (KPI), которая зачастую стимулирует станции к локальной оптимизации по принципу «передай от себя быстрее и больше», не подходит к современным реалиям. Такая модель может улучшать показатели одной станции, но ухудшать работу соседних и всего полигона в целом. Это приводит к многократным переработкам грузов, излишним простоям, кружности вагонов и, как следствие, росту себестоимости перевозок (рис. 2).

От режимов работы станции — к адаптивному интеллектуальному планированию. Изменение парадигмы управления перевозочным процессом.

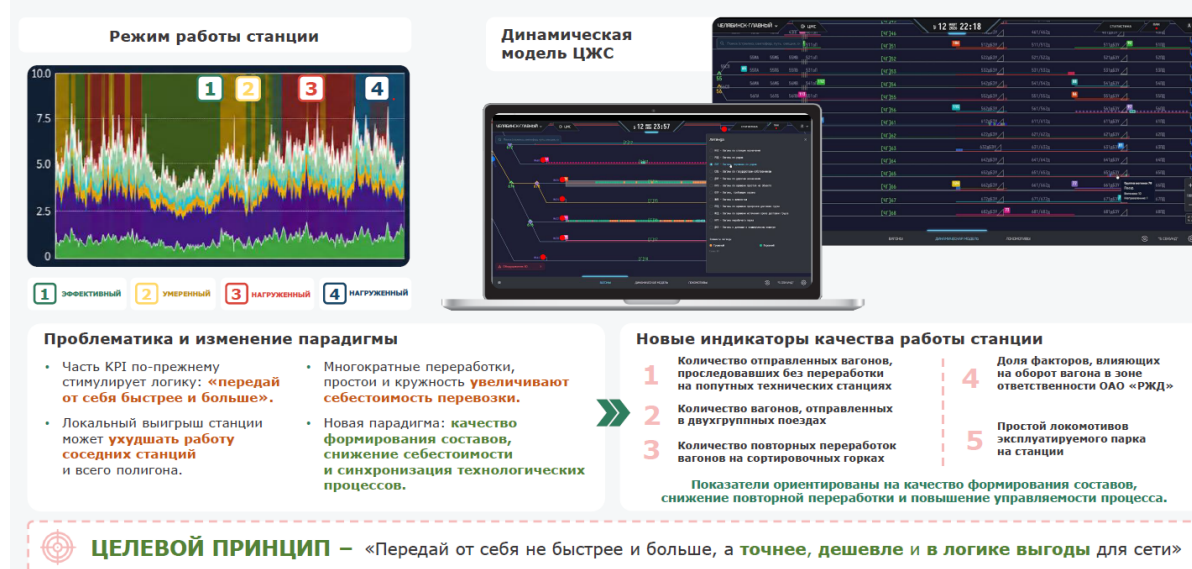


Рис. 2. Решение проблемы локальной оптимизации перевозочного процесса на станциях

В качестве альтернативы АО «НИИАС» предлагает новую систему индикаторов качества работы станции, ориентированную не на валовые объемы, а на качество формирования составов и снижение себестоимости [10–12]. Среди показателей – количество отправленных вагонов, проследовавших без переработки; количество вагонов, отправленных в двухгруппных поездах; количество повторных переработок вагонов; доля факторов, влияющих на оборот вагона в зоне ответственности ОАО «РЖД»; и простой локомотивов эксплуатируемого парка. Целевой принцип здесь звучит как «Передай от себя точнее и дешевле и в логике выгоды для сети», что знаменует переход от локальной оптимизации к глобальной эффективности всей транспортной системы. Методика формирования объективных показателей станции по данным «от колеса» представлена российскими исследователями в работе [12].

Для реализации этой модели предлагается переход к управлению грузопотоками со слежением за каждой элементарной вагоноотправкой. Это достигается через интеграцию АС ГИД НП и «Альфа-Модели» – инструмента прогнозирования и оптимизации (рис. 3). Данная модель предполагает сквозной цифровой контур, где данные непрерывно собираются о состоянии подвижного состава и инфраструктуры, затем на их основе осуществляется интеллектуальное планирование, а завершается цикл роботизированным

исполнением задач, сформированных системой.



Рис. 3. Реализация «Альфа-Модели» в увязке с АС ГИД НП

В докладе уделено внимание вопросу создания единой цифровой платформы управления перевозочным процессом, которая объединит более двух сотен автоматизированных систем, действующих на российских железных дорогах. Эта платформа должна устранить цифровые разрывы и различия в нормативно-справочной информации, создав непрерывный сквозной технологический процесс. Особое значение придается плановой модели движения – так называемому вариантному графику движения. Внедрение этой системы существенно расширило представление о природе транспортного процесса: по каждому поездоучастку сети каждый месяц создается порядка 30 тысяч различных вариантных графиков, включающих полное технологическое обеспечение перевозочного процесса. Научные подходы к устойчивости расписаний, распределению путевой мощности и оперативному разрешению конфликтов подробно представлены в работах [13–16].

1.4. Проект «Цифровая железнодорожная станция» как пилотный полигон новой парадигмы

Ключевым проектом, объединяющим все элементы новой модели управления, является концепция «Цифровая железнодорожная станция» (ЦЖС) на станции Челябинск-Главный [17]. Этот проект, разрабатываемый АО «НИИАС», стал пилотным полигоном для отработки сквозной цифровизации процессов работы станции. К моменту проведения конференции готовность ЦЖС оценивалась в 80%, а ключевая задача на 2026 год – проведение комплексных испытаний.

В рамках проекта на станции тестируются 19 из 24 технических решений. Уже сейчас маршруты на сортировочной станции прокладываются с использованием искусственного интеллекта, вагоны и локомотивы диагностируются «на ходу» с помощью технического зрения, а расцепку составов выполняет

робот (рис. 4). Данные в цифровой двойник станции поступают без ручного ввода напрямую с беспилотных локомотивов и систем мониторинга инфраструктуры. В международной практике возможности цифровых двойников для мониторинга, прогнозного обслуживания и анализа безопасности железнодорожных объектов рассматриваются в исследованиях [18–20].



Рис. 4. Концепция ЦЖС в рамках пилотного проекта на ст. Челябинск-Главный

Цифровая станция – это не просто ускорение обработки вагонов, но и снижение издержек, связанных с ошибками персонала. Как первым заместителем генерального директора ОАО «РЖД», так и генеральным директором АО «НИИАС» отмечено, что уровень ошибок и сбоев при цифровой автоматизации радикально ниже, чем на станциях с традиционной организацией труда. По оценкам специалистов, до 15% инцидентов на железнодорожном транспорте связаны с участием человека в технических операциях, а внедрение компонентов ЦЖС позволяет автоматизировать до двух третей операций, сокращая время обработки состава на 50% и повышая при этом перерабатывающую (пропускную) способность станции. Кумулятивный экономический эффект от тиражирования таких решений на сети оценивается в десятки млрд руб. ежегодно. Кроме того, цифровизация позволяет улучшить условия труда работников, освободив их от тяжелого физического труда на путях станции, особенно в наиболее опасных зонах, таких как горб сортировочной горки при расцепке вагонов.

Социальный аспект внедрения цифровых технологий также был затронут в выступлениях. Большая сортировочная станция – это тысячи работающих в смену людей в разных географических широтах и сложных погодных условиях. Тяжелый труд осмотрщиков вагонов, путейцев, составителей поездов, регулировщиков скорости – все это требует автоматизации не только для повышения эффективности, но и для сохранения здоровья и безопасности работников.

Отдельное внимание было уделено роботизированному комплексу для

расцепки вагонов при роспуске составов на сортировочной горке. Этот робот, разработанный АО «НИИАС» в кооперации с российскими роботостроителями, успешно прошел опытную эксплуатацию по одному пути сортировочной горки. За время испытаний робот выполнил более 9 тыс. расцепок вагонов в диапазоне скоростей до 10 км/ч (для сравнения: при ручной расцепке средняя скорость движения состава составляет 5 км/ч). Успешное завершение опытной эксплуатации позволяет перейти к оснащению всех путей станции и подготовке к приемочным испытаниям, а в дальнейшем – к тиражированию технологии на крупнейшие сортировочные станции сети, такие как Инская, Кинель, Входная и им. Максима Горького.

1.5. Транспортная автоматизация

В выступлении А. Романчикова автоматизация была показана как система взаимосвязанных инфраструктурных и бортовых компонентов (рис. 5). Инфраструктура формирует ограничения и разрешения на движение, а бортовой комплекс должен помочь машинисту соблюдать допустимую скорость, обеспечить торможение и контролировать состояние машиниста. Такой принцип особенно важен при сокращении межпоездных интервалов: рост пропускной способности допустим только при одновременном повышении точности определения местоположения, надежности обмена данными и быстродействия управляющих воздействий.

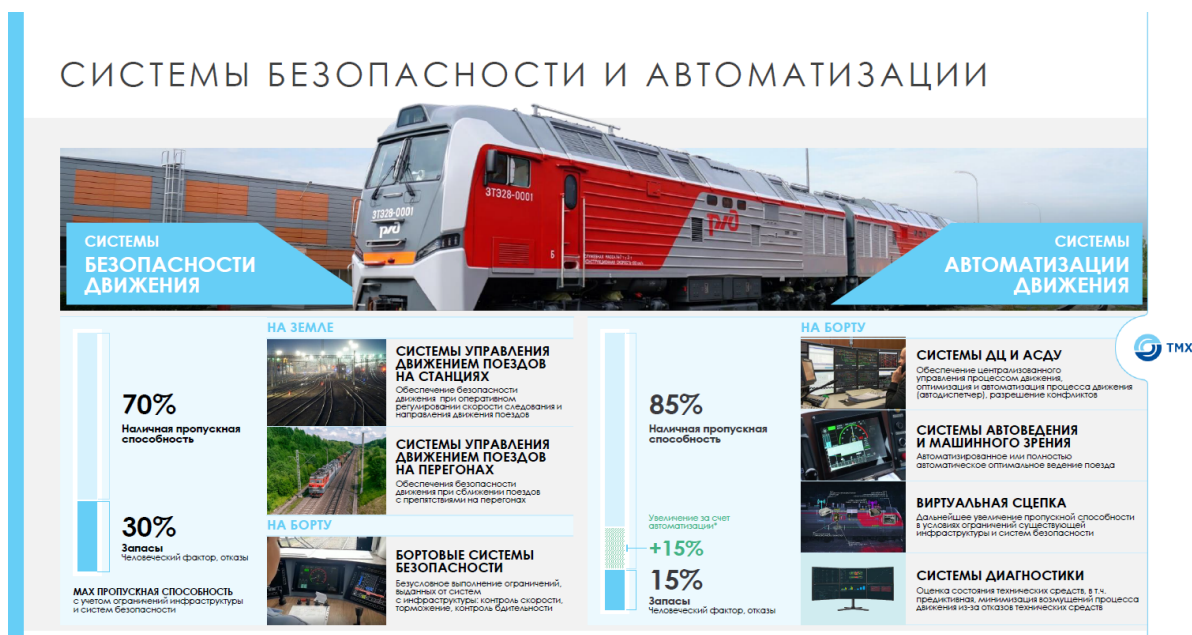


Рис. 5. Комплексная автоматизация на базе проектов «Трансмашхолдинг»

Совместно с АО «НИИАС» была разработана модернизированная отечественная бортовая система безопасности БОРТ-Р. В материалах отмечено прохождение приемочной комиссии и опытной эксплуатации на электровозах 2ЭС5К «Ермак» и электропоездах ЭП2ДМ. В июне 2026 года первый серийный комплект бортового комплекса предиктивной диагностики был установлен на электровоз 3ЭС5К. По представленным оценкам, комплекс способен выявлять до 43% отказов первой и второй категорий по диагностируемому оборудованию. В отличие от регламентного обслуживания, ориентированного

на усредненный ресурс, такая диагностика создает основу для ремонта по фактическому состоянию.



Рис. 6. Результаты деятельности ТМХ за 2025–2026 гг.

Наиболее зрелым примером интеграции назван комплекс виртуальной сцепки и автоведения (рис. 6). В презентационных материалах сообщалось, что в 2026 году применение виртуальной сцепки на Транссибе выросло на 16%, объем перевозок – на 7,6%, а количество тяжеловесных поездов – на 20,2%. Системами виртуальной сцепки было оборудовано 2435 электровозов. Эти показатели важны тем, что характеризуют уже не единичную демонстрацию, а масштабируемую технологию, влияние которой проявляется на уровне полигона.

Дальнейшая программа ТМХ включает применение близких принципов в проектах МЦД-4, беспилотного БАМа, цифровой железнодорожной станции и высокоавтоматизированных маневровых локомотивов ТЭМ23 с функцией «Автомашинист» совместно с АО «НИИАС».

1.6. Отечественная компонентная база и эксплуатационная устойчивость

Доклады КЭАЗ и ГК «Промэлектроника» дополнили обсуждение практическими вопросами технологического суверенитета. КЭАЗ представил комплекс энергоснабжения постового оборудования железнодорожной автоматики, телемеханики и связи: вводные устройства фидеров, комплектные распределительные устройства, системы бесперебойного питания, шкафы с изолирующими трансформаторами, средства технической диагностики и мониторинга. Заявленный уровень локализации достигает 96%, а оборудование рассчитано на эксплуатацию при температурах от -60 до $+60$ °С, вибрационных и ударных воздействиях, характерных для транспортной инфраструктуры.

Пример КЭАЗ показывает, что суверенность определяется наличием полного цикла проектирования и производства, отраслевой сертификации, ремонтной базы, запасных частей, типовых проектных решений и специа-

листов, способных сопровождать оборудование десятилетиями. Работа над альбомом типовых схем низковольтных комплектных устройств позволяет унифицировать проектирование и снизить риск появления уникальных, трудно обслуживаемых конфигураций на каждой станции.

Президент ГК «Промэлектроника» И. Тильк связал развитие автоматики с пятью ограничениями: дефицитом квалифицированного персонала, сокращением финансирования строительства и обслуживания средств железнодорожной автоматики и телемеханики (ЖАТ), уходом зарубежных поставщиков, быстрым изменением технологий и необходимостью осторожного применения радиосвязи и спутниковой навигации. В качестве ответа были представлены бесконтактное устройство контроля схода подвижного состава, комплекс технических средств автоматического закрепления, объектные контроллеры и решения для малодеятельных линий.

В докладе показана перспективная концепция «цифрового грузового вагона», которая предполагает объединение вагонных датчиков в поездную сеть (рис. 7).

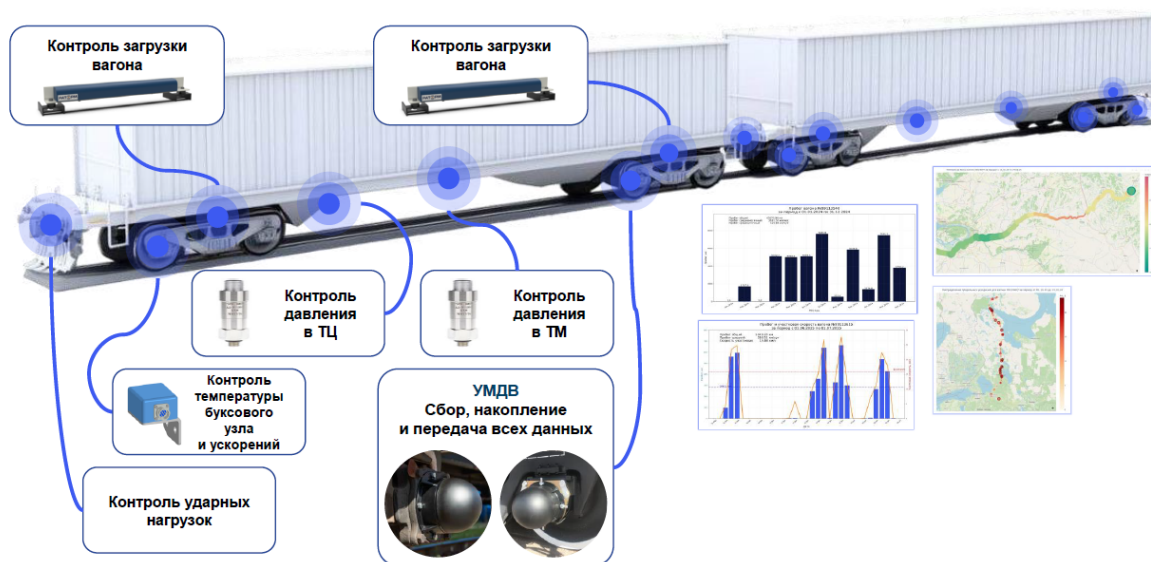


Рис. 7. Иллюстрация концепции цифрового вагона

Машинист и управляющая система получают сведения о состоянии тормозной магистрали, нагреве узлов, целостности состава и других параметрах в ходе движения, а не только при осмотре на станции. При реализации такой концепции вагон становится активным источником данных для цифрового полигона. Это принципиально меняет диагностику: решение о дальнейшем следовании, ограничении скорости или выводе вагона из состава может приниматься на основании непрерывной истории состояния.

2. Интеллектуализация железнодорожного транспорта

2.1. Искусственный интеллект как сквозная технология

В отличие от других сессий конференции, где акцент делался на аппаратных решениях, в центре внимания сессии по искусственному интеллекту (ИИ) оказались метаалгоритмы, нейросетевые модели и их интеграция в различные аспекты железнодорожной деятельности. Доклады показали, что искусственный интеллект на российских железных дорогах перешел из стадии научных экспериментов в фазу системного промышленного внедрения. Стоит отметить, что на международной арене современная классификация таких применений и вопросы нормативного регулирования обобщены в работах [2, 3, 21].

По оценке компании Baker McKenzie, к концу 2025 года только в сфере генеративного ИИ в мире активно развивалось более 10 000 проектов. Инвестиции в США на 2026 год запланированы на уровне около 745 млрд долларов только у четырех основных гигантов, в Китае – около 100 млрд долларов. Россия также увеличивает объемы инвестиций. Модератор сессии (А. Зажигалкин) подчеркнул, что искусственный интеллект становится реально работающей технологией, поскольку крупные финансовые компании инвестируют в это направление. Следовательно, искусственный интеллект сегодня – технология, которая уже приносит определенные эффекты и возврат инвестиций.

Вместе с тем А. Зажигалкин обозначил и риски, связанные с развитием ИИ: существенные риски юридического, этического и управленческого плана. Упомянут эффект сингулярности в области искусственного интеллекта, когда ИИ начнет самостоятельно управлять технологическими и иными процессами, оттесняя человека от соответствующих функций. При этом участники сессии сосредоточились на более прикладных вопросах: насколько ИИ пригоден, интересен и эффективен для реализации технологий в сфере железнодорожного транспорта.

2.2. Корпоративная система поддержки деятельности на основе LLM

Генеральный директор АО «ВНИИЖТ» С. Виноградов представил подход института к внедрению больших языковых моделей (LLM) в основные рабочие процессы. Корпоративная система на основе LLM получила наименование «СоТрудник» (рис. 8).

В докладе отмечены высокая скорость обработки типовых задач и сравнительно низкая стоимость вычислительных операций. Одновременно указаны ограничения универсальных языковых моделей: нестабильность результатов при смене предметной области, отсутствие знаний о внутренних процессах института и необходимость экспертной проверки сформированных материалов. Приведенные в докладе показатели производительности требуют верификации на репрезентативном наборе производственных задач.

Предложена архитектура корпоративной системы, включающая три компонента:

- **Память** – библиотека института с отчетами, статьями и стандартами с поддержкой семантического поиска;

- **Интерфейс** – унифицированный механизм доступа к корпоративным системам;
- **Правила** – методики и регламенты, определяющие порядок обработки запросов и ограничения моделей.

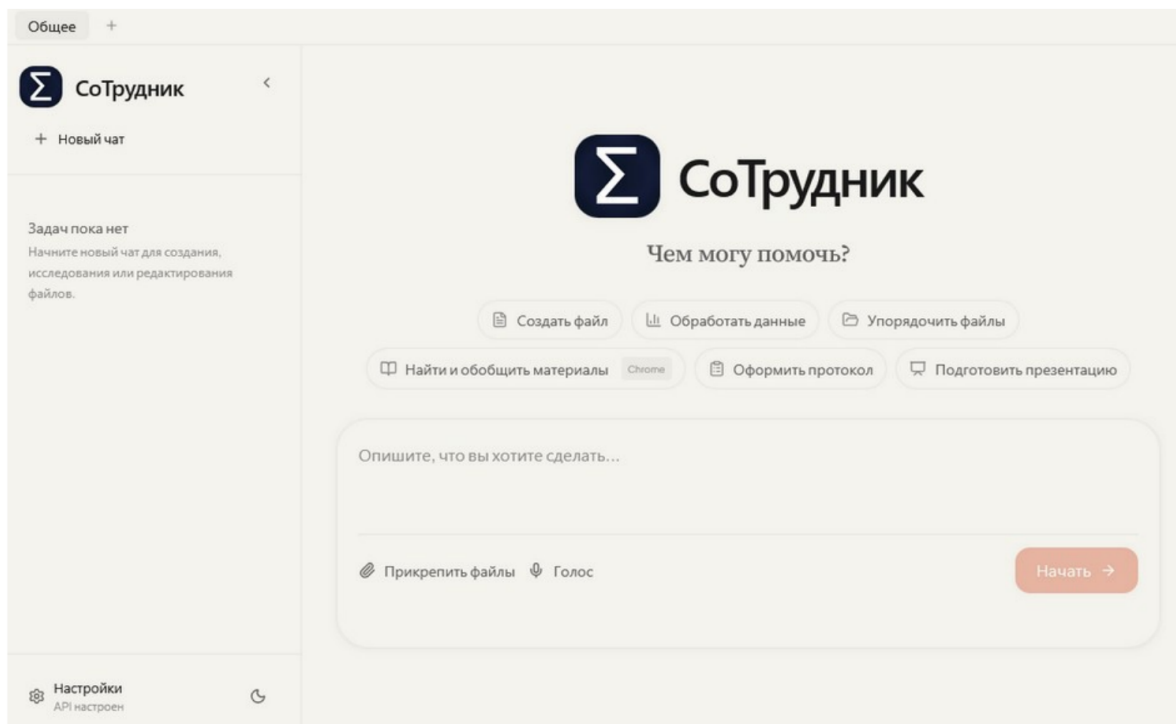


Рис. 8. Интерфейс «СоТрудник»

Разделение прикладного контура и базовой языковой модели позволяет обновлять модель без перестройки базы знаний, интерфейсов и регламентов. К заявленным областям применения системы относятся поиск в нормативных документах, обработка результатов измерений и испытаний, подготовка отчетных материалов, анализ изображений дефектов рельсов, а также разработка и корректировка программного кода в защищенном контуре. Результаты, сформированные системой, подлежат проверке профильным специалистом.

Также представлена система «СИГМА», реализующая многоагентный анализ сложных производственных задач (рис. 9). Для формирования заключения используются специализированные программные агенты, представляющие различные предметные области, включая локомотивное и путевое хозяйство, системы сигнализации, централизации и блокировки, психологию и технологию перевозочного процесса. Итоговое заключение имеет рекомендательный характер и утверждается ответственным специалистом.

В качестве следующего этапа внедрения заявлена интеграция платформы с процессами делопроизводства и проектного управления АО «ВНИИЖТ».

2.3. Физический искусственный интеллект и биоморфные роботы

В докладе заместителя генерального директора АО «НИИАС» А. Хатламджияна отмечено важное концептуальное различие между тремя типами ИИ (рис. 10).

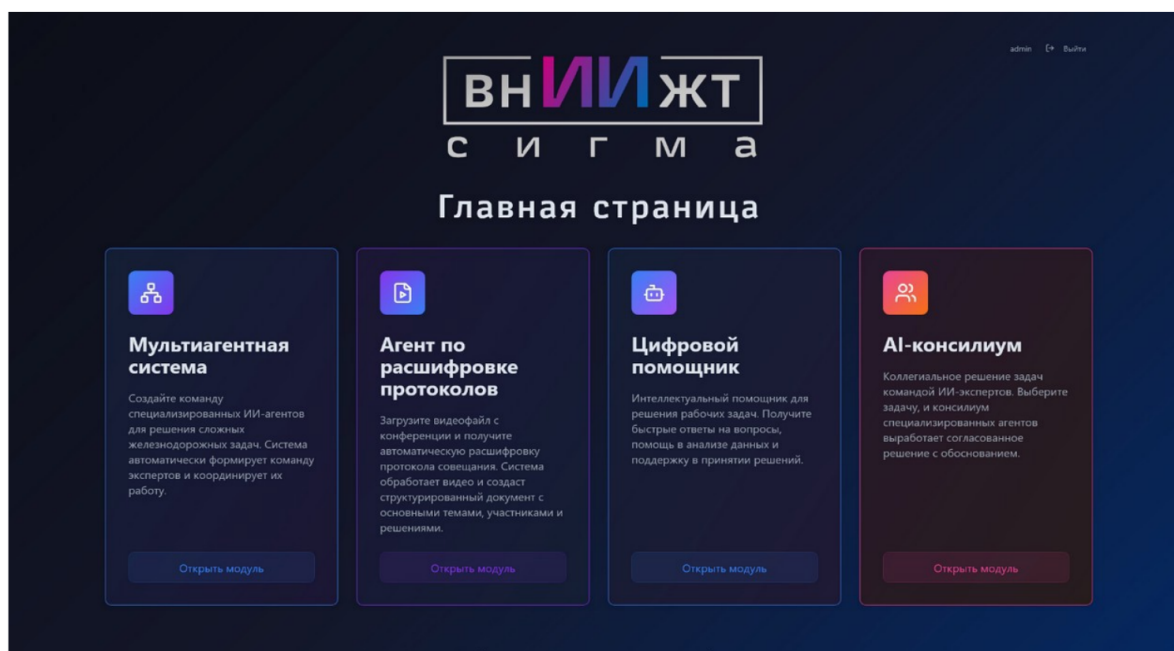


Рис. 9. Интерфейс «СИГМА»



Рис. 10. Три типа искусственного интеллекта для ж.-д. транспорта

Именно физический ИИ является следующим большим рубежом, поскольку он не просто «думает», а действует через мехатронные устройства, воспринимает мир через сенсоры, понимает физические свойства и контекст окружения и использует механизмы для взаимодействия с объектами. Ключевое преимущество подхода – создание независимого от аппаратной части интеллекта: один и тот же универсальный физический ИИ может управлять различными роботами, от точной сборки на конвейере до автономной логистики на складе.

В докладе приведены аналитические прогнозы: к 2050 году рынок роботов может достичь 5 триллионов долларов, а количество роботов-андроидов в мире вырастет до 1 миллиарда. Применительно к ОАО «РЖД» это означает около 70 тысяч роботов, что делает создание универсального «мозга» для них одной из ключевых технологических задач десятилетия.

В ОАО «РЖД» уже определены этапы создания физического ИИ и биоморфных роботов:

1. Разработка ядра физического ИИ;
2. Создание опытных образцов с использованием покупных антропоморфных роботов (Unitree, UBTech, Aldol);
3. Проведение испытаний в реальных условиях.

В качестве примеров применения биоморфных роботов приведено их использование в транспортной полиции, для контроля состояния грузовых вагонов при выходе из ремонта и для автоматизации контроля пассажирских вагонов на входе и выходе из ремонта (рис. 11).



Рис. 11. Существующая реализация физического ИИ для ж.-д. транспорта

Первоочередными площадками внедрения антропоморфных роботов с физическим ИИ определены вагоноремонтные предприятия, сортировочные станции, пункты технического обслуживания, терминально-складские комплексы и вокзалы.

Спектр заменяемых профессий широк: от инспекторов вагонов и состава

вителей поездов до слесарей, экипировщиков и работников информационных стоек. В докладе также представлена мировая практика создания центров обучения роботов (фабрик компетенций) в Китае, Германии, США и России (Сбер, Яндекс) (рис. 12) и обоснована необходимость создания отраслевого Центра обучения роботов ОАО «РЖД», который будет учить роботов быть именно «железнодорожниками» – безопасно, по регламенту и в реальных условиях эксплуатации.

Центры обучения роботов — мировой опыт и российская практика



Фабрики компетенций учат роботов быть «человеком в мире»

 Шанхай, Китай Первый центр совместного обучения гуманоидов, 5000 м², 100+ роботов.	 Пекин, Китай Крупнейший в Китае центр сбора данных для роботов, 10 000+ м².	 Мюнхен, Германия TUM RoboGym — крупнейший тренировочный центр Европы, инвестиции €17 млн.
 США Apptronik и Google DeepMind — Robot Park для сбора данных в логистике.	 Сбер, Москва Новый офис 8745 м² для 300+ специалистов; обучение робота Грина и модели Green-VLA.	 Яндекс, Москва VLA-модель физического ИИ: данные для роботизации собирают через VR-телеуправление и DAGger.

Отраслевой Центр обучения роботов ОАО «РЖД» учит быть «железнодорожником» - безопасно, по регламенту и в реальных условиях эксплуатации

Рис. 12. Мировой опыт обучения роботов

2.4. Применение методов ИИ в жизненном цикле систем ЖАТ: от технологического ПО до организации движения

В докладе М. Фомина, ИТ-директора дивизиона «Железнодорожная автоматика и телемеханика» ГК «Нацпроектстрой», выделены четыре области применения методов ИИ в жизненном цикле систем ЖАТ:

- цифровое проектирование инфраструктуры;
- диспетчеризация и управление движением;
- киберзащита критической инфраструктуры;
- управление техническим обслуживанием.

В первой области эффект связан с исключением ручного преобразования проектной документации, в последней – с накоплением данных об износе для планирования обслуживания по состоянию.

Основная часть доклада посвящена интеллектуальной системе управления перевозочным процессом (ИСУПП). Показана эволюция систем от контроля и управления объектами ЖАТ через диспетчерскую централизацию к уровню, включающему управление расписанием, автоматическое слежение за поездом, прогнозирование движения и разрешение конфликтов. В сопоставлении с зарубежными системами по пяти функциям – управление расписанием (TM), автоматическая установка маршрутов (ARS), разрешение конфликтов (CR),

автоматическое слежение (АТТ) и прогнозирование (ТФ) – заявлена полная реализация всех функций, тогда как у ряда аналогов реализация частичная.

Функциональность ИСУПП включает непрерывный мониторинг поездного положения, автоматическую установку маршрутов, контроль выполнения расписания, прогнозирование движения и предиктивное обнаружение конфликтов. Разрешение конфликтов выполняется методами машинного обучения с учетом технико-распорядительных актов станций, категорий поездов, приоритетов и текущих ограничений инфраструктуры; система работает в режиме автодиспетчера, периодически пересчитывая прогнозный график и формируя варианты устранения конфликтов. Постановка задач переупорядочения и перемаршрутизации поездов в реальном времени и би-критериального разрешения конфликтов изложена в [1,14,16]; отечественный опыт мультиагентного планирования и автоматической установки маршрутов на Московском центральном кольце (МЦК) описан в [22].

Второй блок доклада – применение больших языковых моделей при разработке технологического программного обеспечения систем ЖАТ на специализированных языках ЯЗОП и СТОЯЗ (рис. 13).

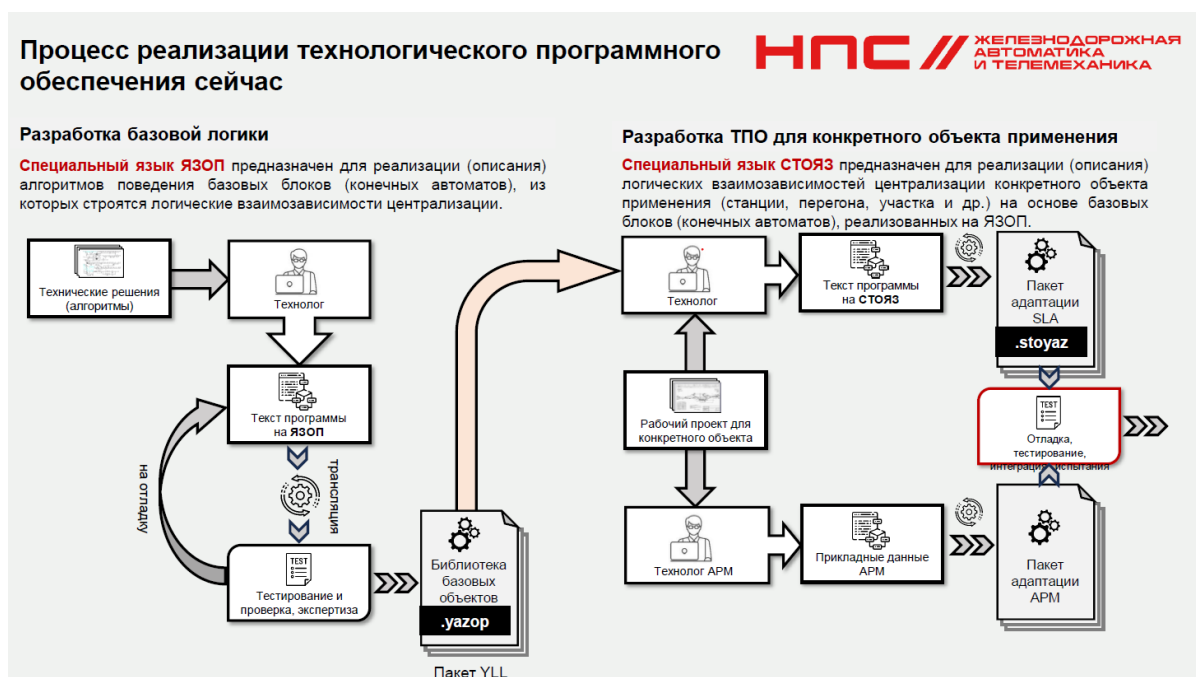


Рис. 13. Процесс реализации ПО ЖАТ на специализированных языках ЯЗОП и СТОЯЗ

Модели используются как ассистенты и как агенты, генерирующие и адаптирующие код под конкретный объект. В пилотном проекте трудоемкость разработки сокращена с 5 800 до 44 часов, а продуктовая ценность внедрения ИИ оценена в 30%. Возможности языковых моделей в задачах генерации кода исследованы в [23]. Для программного обеспечения, влияющего на безопасность движения, сгенерированный код в любом случае подлежит независимой верификации и валидации по требованиям признанных стандартов, поэтому эффект относится к сокращению трудоемкости разработки, а не к сокращению доказательной базы.

2.5. Информационная безопасность систем искусственного интеллекта

А. Малявкин, руководитель департамента услуг по информационной безопасности АО «Компания ТрансТелеКом», рассмотрел нормативную базу применения ИИ в Российской Федерации^{2,3}.

Из доклада следует, что нейросетевая модель является информационной системой, которую необходимо защищать и от которой необходимо защищаться. Выделены пять групп рисков применения больших языковых моделей: утечки данных (персональные данные, коммерческая тайна, токены, пароли); промпт-инъекции, прямые и не прямые [24]; злоупотребления агентов, включая генерацию недопустимого контента и отравление модели; «теневой ИИ» – неучтенные сервисы вне контроля подразделения информационной безопасности; несоответствие требованиям регуляторов. Классические средства защиты анализируют трафик и структурированные данные по сигнатурам, поэтому вход и выход модели без специализированного контроля остаются незащищенной зоной.

В докладе предложена трехуровневая модель защиты (рис. 14).



Рис. 14. Многоуровневая модель защиты систем на основе больших языковых моделей

В качестве средств названы шлюзы и межсетевые экраны для LLM, обеспечивающие фильтрацию промптов и аудит, поиск уязвимостей на стадиях обучения и эксплуатации, платформы AI Security и методика OWASP Top 10 for LLM Applications⁴. Рекомендуемая последовательность внедрения: аудит ИИ-систем (инвентаризация моделей, агентов и точек интеграции), определение ограничений использования ИИ, закрепление требований в политиках, моделях угроз и регламентах доступа, внедрение средств контроля. Отечественные подходы к киберзащите централизаций и к применению ИИ при

²Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (с Национальной стратегией развития искусственного интеллекта на период до 2030 года).

³Федеральный закон № 243-ФЗ «О поддержке развития технологий искусственного интеллекта».

⁴OWASP Top 10 for Large Language Model Applications, version 2025. OWASP Foundation. <https://owasp.org/www-project-top-10-for-large-language-model-applications/>

функциональном тестировании критической инфраструктуры представлены в [25, 26].

2.6. Техническое зрение и предиктивная диагностика инфраструктуры

Е. Павлов, первый заместитель генерального директора АО «ЭЛТЕЗА», в докладе «Применение нейромоделей и технического зрения на объектах инфраструктуры» обосновал переход от массового к гибкому и интеллектуальному производству, в котором скорость реакции на изменения становится основным конкурентным параметром. Представлена архитектура ИИ-ядра систем технического зрения на сверточных сетях, детекторах и трекерных архитектурах и приведены задачи для транспорта: обнаружение препятствий на переездах, контроль инцидентов на эскалаторах, платформах, в тоннелях и на мостах. Для производственных площадок описаны обнаружение дефектов на поверхности изделий, прогнозное обслуживание оборудования, контроль соблюдения требований охраны труда (каска, перчатки, очки, процедуры), обнаружение людей в слепых зонах техники и контроль заполняемости участков. Систематизация методов компьютерного зрения для железнодорожных задач дана в [27–32].

По одной из внедренных систем за 6 месяцев идентифицировано 180 тыс. транспортных средств, 14 тыс. человек и 1 тыс. животных, зафиксировано 23 факта несанкционированного проезда через закрытый переезд и 2 случая остановки на путях по неисправности; время подготовки отчета по инциденту сокращено до 1 минуты, оператор подключается только к тревожным событиям (рис. 15). Отличительные свойства решений: устойчивость к условиям съемки за счет мультисенсорного подхода и ансамбля моделей; исполнение на периферийных вычислителях, при котором данные не покидают инфраструктуру заказчика; интеграция без модернизации линий связи; аппаратное резервирование; открытость архитектуры для замены моделей – от сверточных сетей до VLA-агентов.

А. Лохач, начальник департамента информационно-аналитических систем АО «Фирма ТВЕМА», представил направления применения ИИ в диагностике инфраструктуры. Компания поставила более 400 мобильных и 5 000 ручных средств диагностики; ежегодно проверяется более 3,6 млн км пути в 43 странах. Показана эволюция от ручных методов к автоматизированной обработке с применением ИИ и больших данных, включая пилотные проекты с VK Tech. Примеры задач: обработка данных видеоконтроля (шпалы, крепления, накладки, болты, зазоры, поверхностные дефекты рельсов), данных путеизмерителя (автоматическое измерение параметров стрелочных переводов), данных неразрушающего контроля рельсов (автоматическое определение дефектов) и георадиолокации (загрязненность балластной призмы, состояние земляного полотна).

Основная часть доклада посвящена Единой корпоративной автоматизированной системе управления инфраструктурой – системе диагностики и мониторинга инфраструктуры (ЕК АСУИ СДМИ), в которой реализован прогноз бокового износа рельсов в кривых. На основе накопленных данных (1 ПБ за 6 лет, 1 134 сервера в Главном вычислительном центре ОАО «РЖД» (ГВЦ)) определяется дата достижения предельных значений износа, участки ранжируются по прогнозным величинам и формируются планы замены и

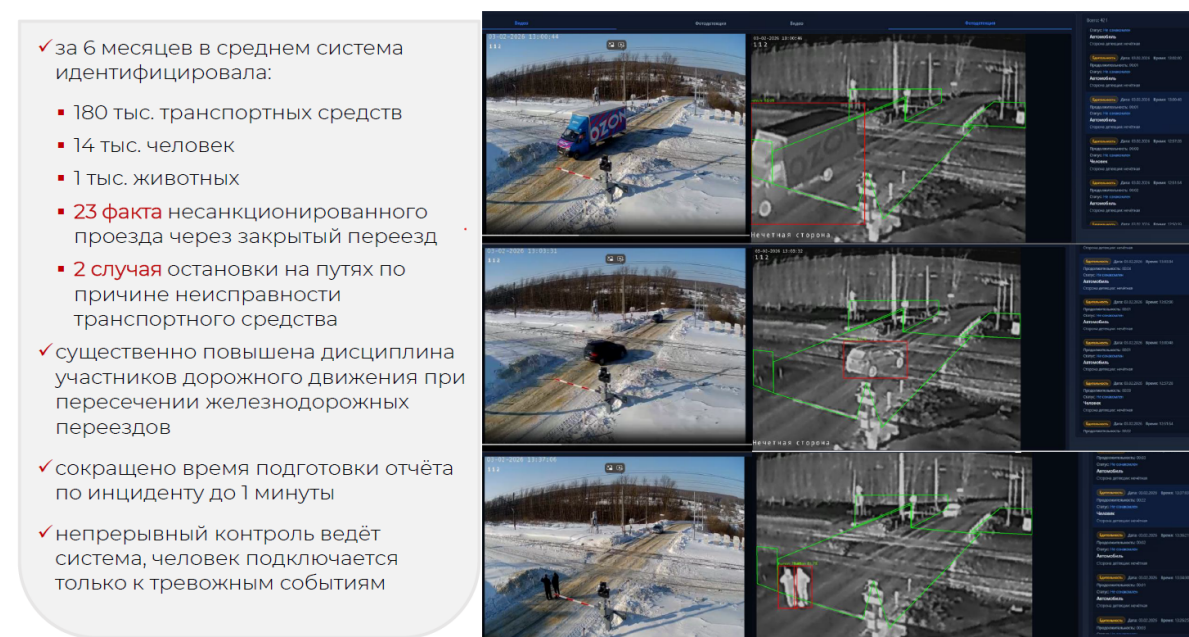


Рис. 15. Архитектура и результаты эксплуатации системы технического зрения на переезде (АО «ЭЛТЕЗА»)

переукладки рельсов. Эффект за прошедший год оценен в 512 км сэкономленных рельсов, или 14,4 млрд руб. Алгоритм прогноза и его место в предиктивном анализе состояния пути описаны в [33].

2.7. Компьютерное зрение для малолюдных технологий

А. Сулов, технический директор ООО «АВП Технология», представил комплекс бортовых решений на основе компьютерного зрения, направленных на снижение участия человека в опасных операциях. Система обнаружения препятствий для магистральных локомотивов в ходе испытаний подтвердила обнаружение, классификацию и локализацию объектов на расстоянии до 600 м; функции включают детектирование показаний светофоров и положения стрелочных переводов, формирование команд о наличии препятствий, автоматическую подачу звукового сигнала при обнаружении человека и вывод визуальной информации машинисту (рис. 16). Обзор бортовых систем обнаружения препятствий и оценки дистанции для железнодорожного транспорта приведен в [34]; компания является разработчиком систем автоведения, на которых построена технология виртуальной сцепки [8].

Смежные решения: система контроля положения стрелочных переводов, распознающая положение остряков, формирующая предупреждение и при критической угрозе выдающая команду на экстренное торможение; система предупреждения проезда запрещающих сигналов, детектирующая людей в запрещенных зонах с геопривязкой инцидентов и передачей сигнала в кабину. Система дистанционного видеоконтроля (СДВ) при движении вагонами вперед использует мобильный блок на хвостовом вагоне и передает изображение машинисту с предупреждением о приближении к препятствию, что позволяет отказаться от нахождения составителя на подножке последнего вагона. Система дистанционного управления локомотивом по радиоканалу обеспечивает управление тягой, торможением, реверсом и вспомогательным

Система обнаружения препятствий (СОП)



Рис. 16. Бортовая система обнаружения препятствий и дистанционного видеоконтроля (ООО «АВП Технологии»)

оборудованием с переносного или стационарного пульта и сокращает время нахождения персонала на путях.

2.8. Данные, обучение и жизненный цикл моделей

Обсуждение на сессии по искусственному интеллекту показало, что результат применения ИИ определяется не выбором конкретной архитектуры, а организацией жизненного цикла данных. Исходная информация формируется средствами автоматики, бортовыми регистраторами, диагностическими вагонами, камерами, датчиками инфраструктуры, диспетчерскими системами и ручными отчетами; источники различаются по частоте обновления, точности, форматам и правилам интерпретации. Механическое объединение таких данных приводит к тому, что модель обучается на закономерностях регистрации, а не технологического процесса. Проблема доступности, репрезентативности и повторного использования отраслевых наборов данных в мировой практике рассмотрена в [35].

Перед обучением требуется паспорт набора данных: происхождение записей, временной диапазон, метод разметки, полнота классов, известные искажения, условия эксплуатации и область применения. Для технического зрения необходимо включать зимние и переходные периоды, осадки, загрязнение оптики, контровое освещение, различные типы подвижного состава и редкие дефекты; для языковых моделей – отделять утвержденные нормативные документы от проектных версий и архивных инструкций; для моделей управления движением – сохранять контекст решения: состояние инфраструктуры, ограничения тягового электроснабжения, приоритеты поездов, наличие бригад и последствия воздействия [30, 36].

После внедрения модель подлежит контролю: изменение технологии, обновление оборудования или появление новых типов объектов вызывает

дрейф данных. Промышленная система ИИ должна регистрировать входы и решения, выявлять снижение качества, обеспечивать безопасный откат версии и повторную валидацию. В этом смысле модель является сопровождаемым технологическим объектом, а не готовым программным модулем.

2.9. Объяснимость и распределение ответственности

Участники сессии разделяли рекомендательную и управляющую функции ИИ. На начальном этапе интеллектуальные решения целесообразно внедрять как «второе мнение»: система обнаруживает риск, предлагает вариант и показывает факторы, повлиявшие на вывод, а решение принимает специалист. Такой режим позволяет накопить статистику, сопоставить рекомендации с действиями опытных работников и выделить классы ситуаций, в которых автоматизация надежнее человека.

При переходе к автоматическому исполнению ответственность распределяется между разработчиком модели, владельцем данных, интегратором, эксплуатирующим подразделением и должностным лицом, утвердившим область применения. Высокая средняя точность не является достаточным доказательством безопасности: редкая ошибка в критическом сценарии весит больше тысяч правильных ответов в штатной обстановке. Практическим следствием является гибридная архитектура: нейросетевая модель распознает ситуацию, прогнозирует ее развитие или формирует план верхнего уровня, а границы допустимых действий задаются детерминированными алгоритмами, нормативами и независимыми защитными контурами. Общие подходы к доказательству безопасности автономных систем и к применению нейронных сетей в ответственных приложениях изложены в [7]; требования к процессам разработки программного обеспечения железнодорожных применений установлены европейским стандартом⁵.

3. Цифровые решения для управления перевозочным процессом

3.1. Достоверность данных как условие проактивного управления

Четвертая сессия (модератор – А. Козловский, начальник комплекса цифрового моделирования АО «НИИАС») связала стратегические положения пленарного заседания с архитектурой управления перевозками. Первый доклад Е. Максимовой, директора Института управления и цифровых технологий РУТ (МИИТ), был посвящен достоверности данных при автоматизации управления движением. Наличие массива записей в автоматизированных системах не означает существования цифрового управления, поскольку качество решения, как правило, не может превышать качество исходной информации, а запоздалые данные, оставаясь формально правильными, теряют оперативную ценность. Особенности применения больших данных к исследованию грузопотоков рассмотрены в [37].

В традиционной схеме стоимость перевозочного процесса оценивается по расходам прошлых периодов; такие сведения необходимы для учета и ретроспективного анализа, но не характеризуют последствия конкретного оперативного решения. Из рассмотрения выпадают вторичные потери, воз-

⁵EN 50716:2023. Railway Applications – Requirements for software development. Brussels: CENELEC, 2023.

никающие позднее и в другом подразделении: задержка оборота вагона, нарушение подвода локомотива, конфликт с ремонтным «окном», производительный пробег, срыв согласованной доставки. Риск-ориентированный подход добавляет к стоимости произошедшего события стоимость и временной масштаб возможных последствий.

Е. Максимовой сформулированы пять требований к данным для прогнозирования: полнота, достоверность, своевременность, сопоставимость и прослеживаемость. Нарушение любого из них запускает цепочку: задержка регистрации → искаженная картина состояния → ошибка прогноза → неверная оценка риска → неэффективное корректирующее воздействие. Отдельно рассмотрена разница между фактическим временем события и временем его внесения в систему: если задержка сопоставима с резервом времени на управленческое решение, запись утрачивает значение для проактивного управления. Иллюстрацией служит пакетный ручной ввод, при котором последовательные операции регистрируются почти одновременно и причинно-временная последовательность переходов объекта между состояниями исчезает из цифровой информации. Исследование на пути необщего пользования с сопоставлением четырех независимых источников информации показало, что данные одной автоматизированной системы не должны считаться единственным основанием для прогноза.

Предложенная целевая схема (рис. 17) предусматривает автоматическое сопоставление источников на общей временной шкале и присвоение каждой записи уровня доверия: данные высокой достоверности поступают в модель автоматически, неоднозначные направляются на дополнительную проверку, записи с низким доверием требуют ручного подтверждения. Условие проактивности сформулировано количественно: задержка между событием и загрузкой подтвержденных данных должна быть меньше времени, имеющегося для предотвращения риска.

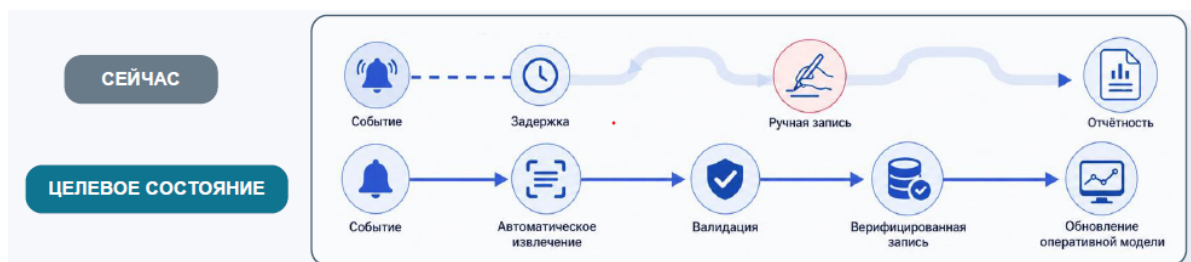


Рис. 17. Схема верификации цифровой информации

3.2. «Альфа-Модель»: гибридное прогнозирование на уровне вагоноотправки

И. Гургенидзе, заместитель генерального директора АО «НИИАС», в докладе «Интеллектуальное управление перевозками: гибридные модели на стыке ML и классических методов планирования на основе графиков движения и плана формирования» представила «Альфа-Модель» как парадигму управления, а не как отдельную информационную систему. Логика решения: система должна знать текущее состояние, построить прогноз, обнаружить нежелательный сценарий и изменить его до наступления последствий. Метод управления движением с использованием вариантных графиков, лежащий в основе плановой модели, защищен патентом [38].

Основой служит единая инфраструктурная модель. Сегодня различные системы используют собственные описания станций, перегонов, путей, профилей и ограничений, что затрудняет обмен и порождает разные трактовки одного объекта. В «Альфа-Модели» инфраструктура должна быть общей для нормативного планирования, фактического контроля, тяговых расчетов, моделирования и выдачи заданий; такой подход применен при проектировании ВСМ и должен быть распространен на действующую сеть. Фактическую информацию формируют АС ГИД НП, станционные системы и динамическая модель цифровой станции. В докладе использовано различие между системами, читающими «с листа» (план формирования, график) и «с колеса» (средства мониторинга) [11,12]; оркестратор, обрабатываемый на фактических данных, объединяет оба слоя и рассчитывает прогноз с учетом работы станций, категорий поездов, времени хода и межпоездных интервалов (рис. 18).

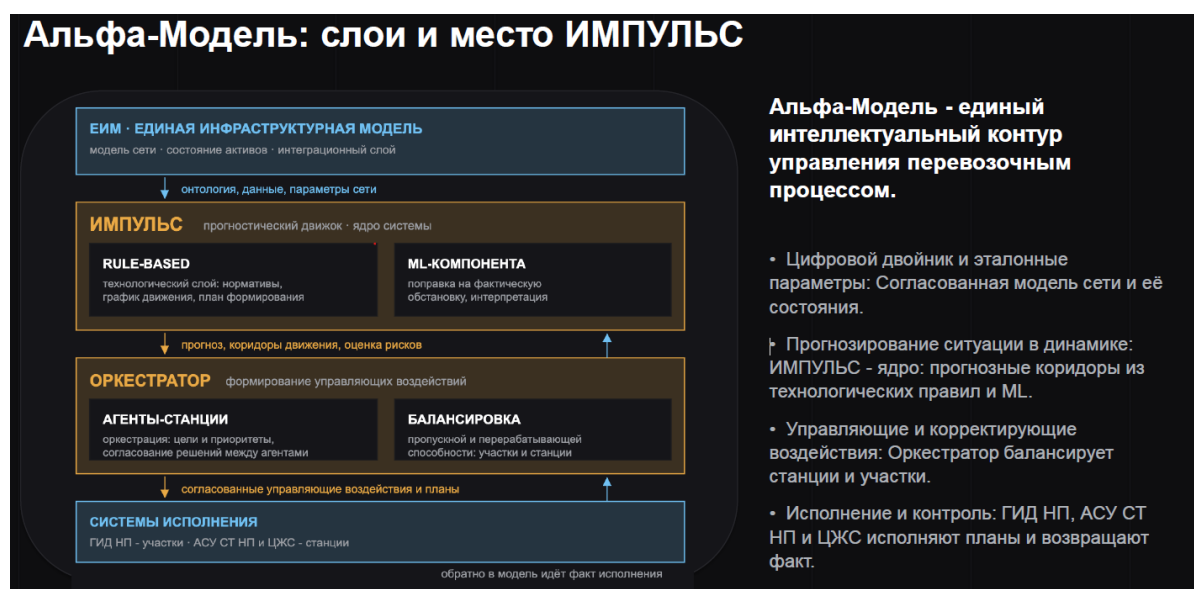


Рис. 18. Архитектура «Альфа-Модели»: единая инфраструктурная модель, фактические данные и контур упреждающих воздействий

Объектом прогноза является не только поезд, но и элементарная вагоноотправка: модель отвечает, где окажется конкретная отправка, в каком состоянии и в какой момент времени. Если прогнозируемый срок выходит за допустимый коридор, подбирается упреждающее воздействие на одном из трех уровней – изменение плана формирования, выбор маршрута вагоноотправки, регулирование движения по перегону. Расчет опирается на моделирование станций, варианты графики и движение поездов.

Рекомендации формируются в объяснимом виде: языковая модель используется для составления протокола, описывающего выявленный риск, учтенные ограничения, выбранное воздействие и ожидаемый результат, что позволяет диспетчеру проверить логику решения. Набор воздействий включает приоритетность пропуска и формирования поездов, перераспределение ресурсов и изменение планов; критерий оптимизации учитывает интересы государства, ОАО «РЖД», операторов, грузоотправителей и грузополучателей, то есть исключает улучшение одного показателя за счет неприемлемого ухудшения другого. Это согласуется с системой сетевых индикаторов качества работы станций, предложенной в докладе генерального директора АО

«НИИАС» А. Долгого (см. раздел 1).

3.3. Цифровизация ЖАТ

М. Фомин в докладе представил существенное изменение автоматизированных систем управления технологическими процессами – переход от циклического опроса зависимостей к событийно-ориентированной модели: расчет запускается при изменении состояния, а не повторяется с фиксированной частотой для всех объектов. На реальной станции число одновременных событий невелико, поэтому основная часть реакций выполняется за миллисекунды; наибольшая нагрузка возникает при запуске контроллеров и массовом обновлении состояния – в представленном в докладе примере полный пересчет занимал до 300 мс. В штатном режиме предел быстродействия все чаще определяется каналом связи, а не вычислителем.

Платформенность подтверждена при вынужденной замене элементной базы: после прекращения доступности выбранных микроконтроллеров проектирование вычислительного модуля и полный цикл испытаний выполнены за девять месяцев благодаря разделению на стандартный вычислительный модуль, унифицированную подсистему питания и прикладную часть. На нижнем уровне представлены универсальные платы управления стрелками и светофорами, модули дискретных и потенциальных входов, релейные интерфейсы и собственный пункт счета осей; один тип платы поддерживает несколько схем управления стрелочным приводом. Разрабатываемый компактный вычислительный модуль предназначен для переездов, разъездов и малых станций, где полноразмерный комплекс экономически избыточен, что обеспечивает масштабирование одной технологической базы от крупной станции до малодеятельного объекта.

Ю. Федоркин, главный инженер ООО «Компания „Стальэнерго“» (дивизион ЖАТ ГК «Нацпроектстрой»), представил решение по модернизации числовой кодовой автоблокировки без использования реле. На питающем конце рельсовой цепи группа релейных приборов заменяется двухчастотным генератором ГКЛС-К со стабилизацией временных параметров кода, который одновременно формирует сигнал 25 Гц для контроля рельсовой цепи и сигнал АЛС 75 Гц; функциональное резервирование позволяет второму каналу принять функции основного при его защитном отказе. На релейном конце блоки дешифрации и импульсные реле заменяются микропроцессорным приемником-дешифратором, который контролирует в том числе сход изолирующих стыков и передает диагностические сообщения в аппаратно-программный комплекс диспетчерского контроля; при внешней помехе устройство переходит в защитное состояние и возвращается к работе после нормализации (рис. 19).

На Восточном полигоне эксплуатируется около 3 700 приборов. Для защиты от грозовых и коммутационных перенапряжений применены конструктивные меры; заявленная электрическая прочность между корпусом и платой достигает 6 кВ. Система контролирует выработку ресурса варисторных модулей и при достижении 80% передает предупреждение для плановой замены, то есть обслуживание защитного элемента переводится на прогнозируемое состояние.

Следующий этап – передача данных между генератором, приемником

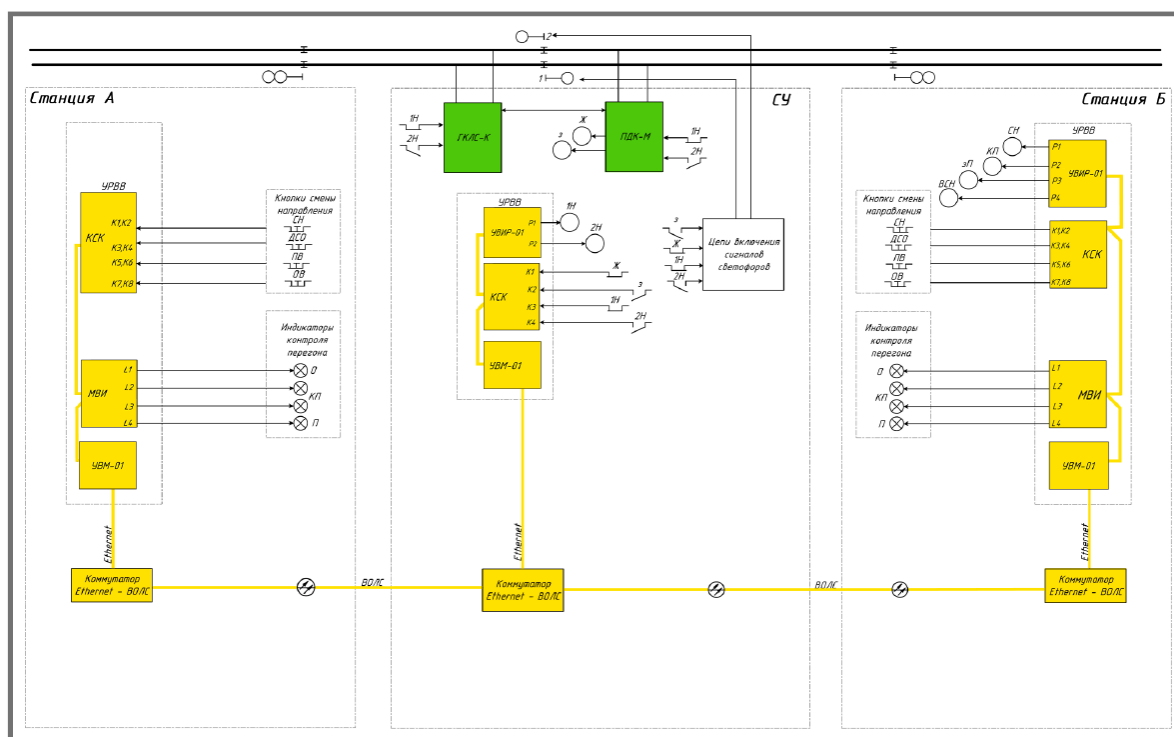


Рис. 19. Структура числовой кодовой автоблокировки от ООО «Компания „Стальэнерго“»

и станциями по волоконно-оптической сети для снижения зависимости от медного кабеля и повышения устойчивости к перенапряжениям. Статистика надежности оптической части на момент конференции не накоплена; включение опытного участка планировалось на 20 ноября 2026 г.

3.4. Замкнутый цикл управления городским рельсовым транспортом

А. Колесников, директор по управлению проектами компании «Лаборатория Сота», представил замкнутый цикл управления городским рельсовым транспортом из четырех этапов: анализ спроса и планирование, построение расписания, оперативная диспетчеризация, анализ исполнения с возвратом результатов в следующий цикл планирования. Приведен опыт единой диспетчерской системы Московского метрополитена (рис. 20) и системы мониторинга Центрального транспортного узла, контролирующей около тысячи единиц подвижного состава в сутки. Необходимость совместного рассмотрения расписания, путевого развития и технологических «окон» обоснована также в мировой практике [39]; отечественные архитектурные и информационные решения для интеллектуального управления городским рельсовым транспортом изложены в [40–43].

На этапе планирования интерактивные модели пересчитывают пассажиропотоки при закрытии участка, ремонте, строительстве крупного объекта или изменении маршрутной сети и формируют количественные задания: сколько пассажиров и по каким направлениям необходимо перевезти. При ручной разработке графика инженер тратит один-два рабочих дня на маршрут; представленный инструмент строит согласованное расписание для маршрутной сети города-миллионника за одну-пять минут, а оптимизация потребности в водителях занимает от десяти до пятидесяти минут, что позволяет



Рис. 20. Рабочее место диспетчера центра транспортного управления Московского метрополитена

пересчитывать график во время ремонтов.

В оперативном управлении рабочее место диспетчера обогащается нормативно-справочной информацией, состоянием инфраструктуры и подвижного состава, данными смежных служб. Автоматизированное перестроение графика действует как «второй пилот»: сначала рассчитывается способ наиболее быстрого выхода из сбоя и восстановления интервалов, затем – план до конца транспортных суток с учетом технического обслуживания и минимальных изменений ночной расстановки составов. Анализ фактически выполненных рейсов и транспортной доступности выполняется по тепловым картам; докладчик предостерегает от выводов без контекста, поскольку рост предложения может компенсировать массовое мероприятие. Для магистральной сети опыт применим при планировании пригородного сообщения, координации с наземным транспортом, организации тактового движения и развитии объединенных диспетчерских центров.

3.5. Система «Кама»: энергооптимальное ведение пакета поездов

М. Шеин, начальник отдела НИР-МПСУ ООО «НПО САУТ», представил интеллектуальную энергоэффективную систему управления движением поездов «Кама». Традиционно энергооптимальный режим рассчитывается для отдельного локомотива в среде моделирования, однако локальный оптимум одного поезда не является оптимальным для участка: ускорение, выбег или торможение изменяют нагрузку на контактную сеть и условия движения соседних поездов. Постановка задачи энергооптимального ведения пакета поездов изложена авторами системы в [44].

В централизованной схеме математические модели автоведения нескольких поездов работают на сервере и получают параметры, аналогичные бортовым системам, что позволяет сравнивать траектории, интервалы, режимы тяги и энергетические ограничения при сохранении требований функциональ-

ной безопасности. Согласованное управление снижает пиковую нагрузку на контактную сеть и выбирает более экономичный способ выполнения графика даже на коротких интервалах. Интерфейс объединяет привычное диспетчеру изображение ниток графика с результатами моделирования режимов ведения; структура системы приведена на рис. 21.

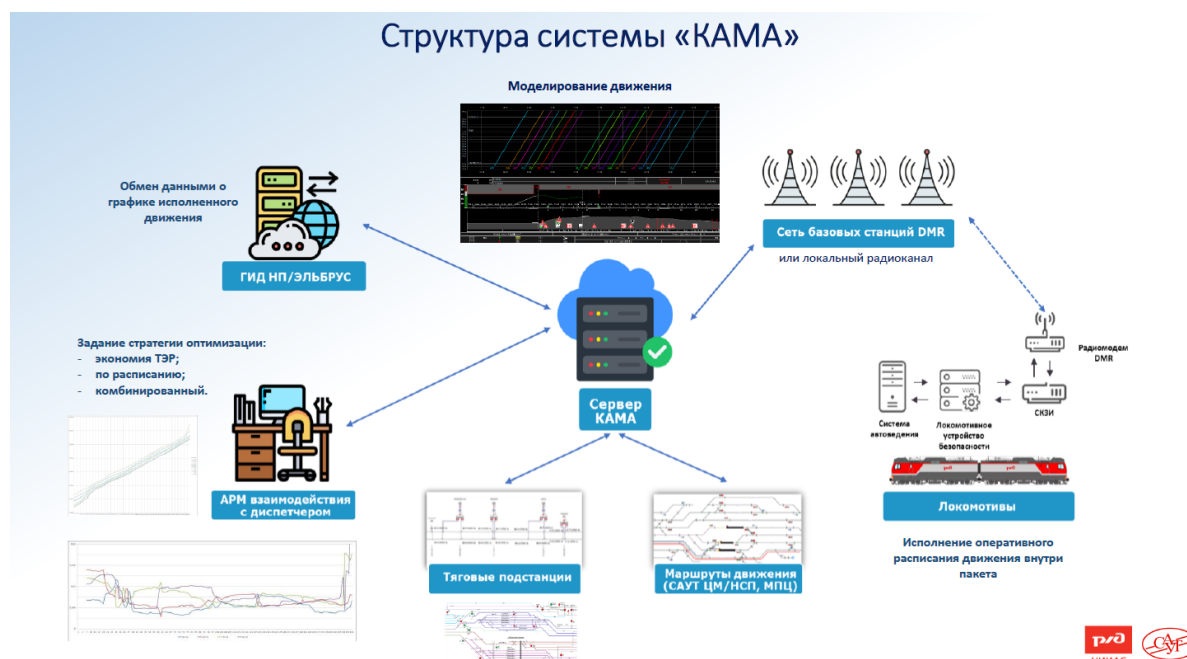


Рис. 21. Структура системы «Кама»

На момент доклада «Кама» представлена как проектный инструмент разработчика. Дальнейшая ценность решения зависит от интеграции трех контуров – управления движением, бортового автоведения и тягового электроснабжения; только совместная оптимизация позволяет ответить, что предпочтительнее в конкретной ситуации: пропустить больше поездов, быстрее восстановить график или снизить энергопотребление.

3.6. Линейные цепи по ВОЛС при модернизации действующих систем автоблокировки

М. Стародубцев, главный конструктор систем автоблокировки научно-производственного центра (НПЦ) «Промэлектроника», рассмотрел модернизацию линейных цепей действующих систем ЖАТ. Медножильные кабели просты в ремонте, но подвержены атмосферным и коммутационным перенапряжениям, электрохимической коррозии и деградации изоляции; волоконно-оптическая линия связи (ВОЛС) устраняет большинство этих факторов, однако прямая замена затруднена сохранением на сети большого объема релейной автоматики.

Предложено интерфейсное устройство, транслирующее состояния релейных цепей через оптический канал без изменения алгоритмов существующей автоматики. Компактный моноблок поддерживает четыре пары сигналов и дальность до 30 км; при добавлении волокна возможна кольцевая схема резервирования. Область применения – полуавтоматическая блокировка, переездная сигнализация, увязка с устройствами контроля схода подвижного

состава и модернизация числовой кодовой автоблокировки, протяженность которой на пространстве колеи 1520 мм оценивается примерно в 50 тыс. км. Экономический эффект связан с сокращением кабельной инфраструктуры и земляных работ.

На момент доклада решение готовилось к опытной эксплуатации, статистика отказов отсутствовала. Масштабирование должно опираться на сравнительные испытания надежности, ремонтпригодности и фактической стоимости жизненного цикла в сопоставлении с медными линиями.

4. Роботизация железнодорожного транспорта

4.1. Постановка вопроса: от отдельных внедрений к системной роботизации

Седьмая сессия проведена под модерацией Е. Дудорова, генерального директора ООО «Р-ТЕЛЕМАТИКА» и председателя правления консорциума робототехники и систем интеллектуального управления. По оценке модератора, за семь лет с начала совместной работы железнодорожных производителей, институтов (Концерн «Телематика», АО «НИИАС» и др.) и ОАО «РЖД» отношение к роботизации в отрасли изменилось от отсутствия понимания области применения до внедрения роботизированных комплексов; совместный аудит выявил более пятидесяти точек приложения роботизации. Запрос отрасли смещается от классических промышленных роботов к системам, замещающим человека в неструктурированной среде. Основным предметом обсуждения стал разрыв между технологической готовностью решений и зрелостью управленческой и нормативной среды, что раньше также было отмечено в работе [45].

4.2. Управленческая рамка роботизации: «Путь к экосистеме»

А. Зажигалкин, директор Ассоциации «Транспортная наука», в докладе «Роботизация РЖД. Путь к экосистеме» охарактеризовал текущее состояние как «системный разрыв». В холдинге «РЖД» накоплен портфель решений и реализованы десятки пилотных проектов – от программных роботов (RPA) на базе «Фабрики роботизации» ГВЦ, автоматизирующих документооборот и ИТ-поддержку, до физических комплексов: расцепщиков вагонов, диагностических систем, электропоездов с автоматическим управлением на МЦК и маневровых локомотивов. Однако эти результаты не образуют системного эффекта из-за отсутствия единой программно-управленческой рамки: концепция роботизации на период до 2035–2050 гг. разрабатывается, но не оформлена как корпоративная программа с целями, бюджетом и системой управления, а темп появления новых задач опережает готовность управленческого контура и нормативной среды (рис. 22). Онтологический подход к описанию цифровой железной дороги, на который опирается предлагаемая рамка, представлен в [46].

Выделены три управленческие проблемы: отсутствие единого квалифицированного заказчика и координатора (роль проектного офиса); необходимость «коридора пилотов» – упрощенного маршрута для малых проектов с ограниченным бюджетом и коротким циклом по аналогии с экспериментальными правовыми режимами; развитие кооперационной модели, поскольку в эко-

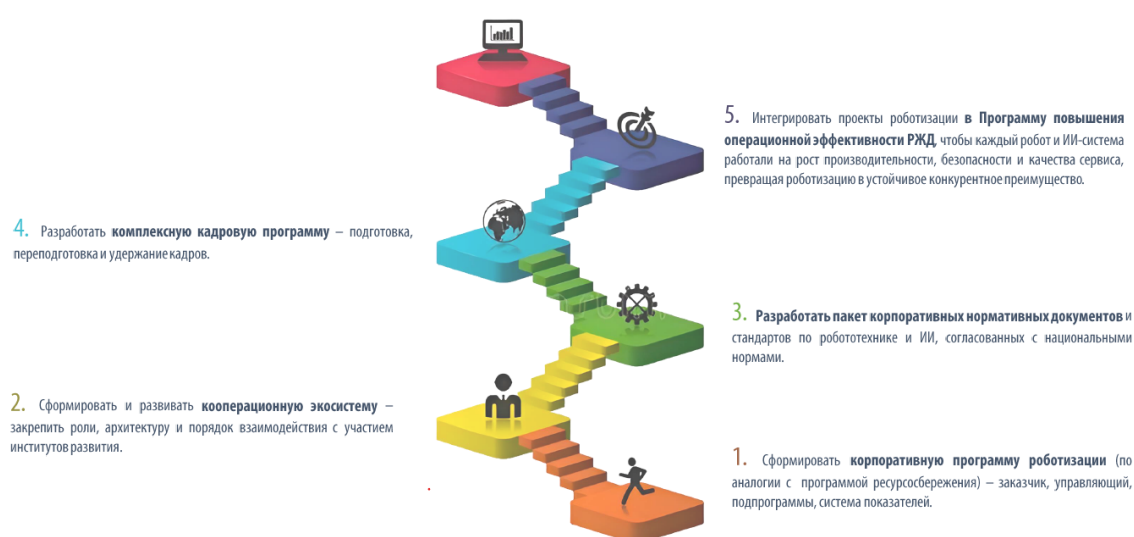


Рис. 22. Возможные практические шаги по преодолению «разрыва» в области роботизации

системе участвуют десятки организаций – от ГВЦ, АО «НИИАС» и ООО «Р-ТЕЛЕМАТИКА» до внешних разработчиков и институтов развития. При прогнозируемых 40–45 ключевых проектах число стейкхолдеров превысит сотню, что требует типовых ролей (заказчик, эксперт, управляющий, производитель, интегратор) и единых правил входа.

4.3. Инженерно-роботизированные комплексы «РОИН» на инфраструктуре колеи 1520 мм

Р. Коновалюк, заместитель генерального директора по развитию АО МГК «Интехрос», представил опыт применения инженерно-роботизированных комплексов «РОИН» (рис. 23).



Рис. 23. Линейка роботизированных комплексов «РОИН»

За 2025 г. комплексами «РОИН» заменено около 6 500 шпал, очищено более 73 км кюветов, выполнено более 350 км земляных работ и очищено от снега более 80 км. Совместно с Проектно-конструкторским бюро по инфраструктуре («ПКБ И») разработано 73 технологических документа, регламентирующих применение роботов в инфраструктурном комплексе, что позволило сформировать целевые показатели по росту доли механизации ручных работ по шпалам, рельсам и балласту и по производительности труда на ключевых операциях.

Направление развития – повышение степени автоматизации за счет систем технического зрения на вычислителе Jetson Orin NX, лидаров и стереокамер, распознающих элементы верхнего строения пути, то есть переход от дистанционно управляемой машины к автономной технической системе. В соответствии с поручением Президента РФ планируется десятикратное увеличение парка «РОИН» на сети. В перспективе – конструктивная модернизация Р300 и создание высокопроходимого робота для горной местности с изменяемым клиренсом до 2,5 м. Зарубежные прототипы роботов для обслуживания и диагностики пути описаны в [47–49].

4.4. Интеллектуальное ядро роботов: модели VLA и асинхронный инференс

П. Попов, заместитель директора Научно-исследовательского института транспорта РУТ (МИИТ) и руководитель Центра интеллектуальных и автономных транспортных систем, посвятил доклад моделям Vision-Language-Action (VLA), которые рассматриваются как следующая ступень после больших языковых моделей: робот получает возможность интерпретировать задание на естественном языке и адаптироваться к нестандартным ситуациям («когнитивная автоматизация»). Перенос знаний из визуально-языковых моделей в управление роботами и межплатформенное обучение на объединенных наборах траекторий представлены в [50–52]. В качестве практического примера показан прототип робота-тележки, созданный для чемпионата высоких технологий в Великом Новгороде и оснащенный лидаром, IP-камерой, измерительным модулем и системой аварийной остановки. Опыт докладчика в создании систем автоматического управления электропоездами на МЦК и в доказательстве безопасности автономных систем отражен в [7, 53].

Описана трехэтапная методология обучения VLA-моделей:

1. межплатформенное предварительное обучение на разнородных робототехнических данных (порядка 1 тыс. часов траекторий из открытых баз, например Open X-Embodiment [51]);
2. специализация на конкретной платформе (тип поезда, сенсорная система) с использованием собственных данных, собранных в реальных условиях;
3. дообучение на небольшом числе высококачественных демонстраций для освоения специфических навыков.

Выделены три препятствия для применения VLA на железнодорожном транспорте: инференсная задержка, поскольку вычислительная нагрузка требует мощных бортовых систем, а управление в реальном времени является критическим требованием; разрыв между обучением в симуляции и реальными условиями, требующий сбора данных на объектах и постоянного дообучения; проблема безопасности и сертификации, так как VLA являют-

ся «черными ящиками», формальная верификация которых затруднена. В качестве инструмента для снижения первого барьера предложено использовать архитектуру VLA-RAIL (Real-Time Asynchronous Inference Linker) [54] – открытый асинхронный связующий слой между VLA-моделями и разнородными роботами по принципу «подключи и работай», который объединяет очереди фрагментов действий (action chunks) и снижает рывки и паузы при исполнении (рис. 24).

Этапы обучения моделей VLA

Этап	Назначение	Данные для обучения
1. Межплатформенное предварительное обучение (Cross-embodiment pre-training)	Приобретение общих знаний о мире из разнообразных робототехнических данных	Около 1 тыс. ч записей траекторий движения из OXE
2. Предварительное обучение на целевой платформе (Single-embodiment pre-training)	Специализация на конкретной платформе (тип поезда, сенсоры)	Собственные данные, собранные на целевой платформе
3. Финальное обучение (Post-training)	Освоение конкретных сложных навыков	Небольшое число высококачественных демонстраций для специфических задач

Примечание. OXE (Open X-Embodiment) – это крупный открытый набор данных для обучения роботов.

Рис. 24. Этапы обучения VLA-моделей и схема асинхронного связующего слоя VLA-RAIL

Перспективные направления: интеграция VLM/VLA с ROS-фреймворками для голосового управления и ситуационной осведомленности; цифровые двойники и симуляторы на Unreal Engine для масштабируемого сбора данных и замкнутого цикла валидации; гибридные архитектуры безопасности, в которых VLA отвечает за планирование верхнего уровня, а сертифицируемые детерминированные алгоритмы – за низкоуровневый контроль [7].

4.5. Роботизированные системы АО «НИИАС»: от расцепки вагонов до диагностики

А. Хатламаджиян, заместитель генерального директора АО «НИИАС», в докладе о развитии роботизации в ОАО «РЖД» представил нормативный контекст на федеральном и корпоративном уровнях, текущий статус роботизации (рис. 25) и проект ее дорожной карты (рис. 26). На федеральном уровне это Единый план по достижению национальных целей до 2030 г., Концепция научно-технологического развития транспортного комплекса, проекты федеральных законов о робототехнике и автономных системах и развивающийся комплекс государственных стандартов. На корпоративном уровне – Стратегия научно-технического развития холдинга «РЖД», Концепция обеспечения технологического суверенитета и разрабатываемые

документы: Программа роботизации повторяющихся производственных процессов, Концепция разработки и внедрения робототехнических систем на инфраструктуре и Методика технико-экономической оценки проектов. Обзор проектов роботизации холдинга дан в публикации с участием докладчика [45].

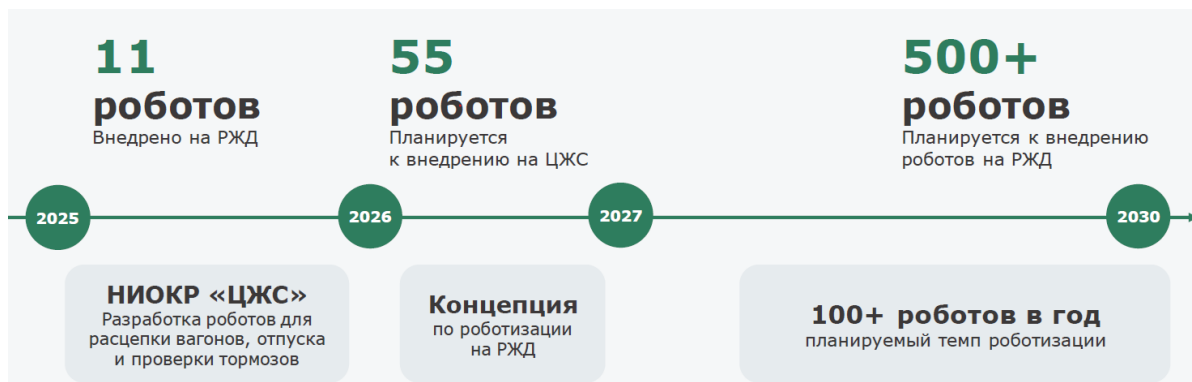


Рис. 25. Текущий статус роботизации ОАО «РЖД»

В рамках доклада были представлены практические результаты АО «НИИАС» и партнеров:

1. Робототехнический комплекс для расцепки вагонов при роспуске составов на сортировочной горке завершил опытную эксплуатацию на станции Челябинск-Главный с вероятностью успешной расцепки более 95%; впервые применен коллаборативный режим опытной эксплуатации, при котором в случае неуспешной попытки операцию выполнял человек, а эпизод использовался для дообучения [17,55];
2. Роботизированная система отпуска и проверки действия автоматических тормозов, включающая роботов для отпуска тормозов, проверки их действия и очистки от снега; завершается второй этап опытной эксплуатации, разрабатывается новый дизайн;
3. Применение биоморфных роботов для диагностики буксовых узлов и экипажной части: в автоматическом режиме контролируется более 40 типов параметров – элементы крепления (болты, крышки, замки), деформации конструкций (лестницы, отбойники, окна), соединительные элементы (кабели, разъемы, шланги) и расходные материалы (песок, фильтры, колодки).

Расчетный эффект диагностического робота: время контроля состава сокращается с трех до двух часов, производительность труда контролера повышается на 30%, простой поездов снижается на 75–80 часов в месяц (900–960 часов в год). В завершение представлен концептуальный облик антропоморфного робота для ОАО «РЖД» с гусеничным шасси для пересеченной местности, поворотным на 360° торсом, приводом для подъема тяжелых грузов, сменными захватами и поворотным шарниром для устойчивости на неровностях. Отмечено, что морфология человека не является обязательной: робот для железной дороги может иметь более мощные приводы и встроенный в кисти инструмент, что делает его эффективнее человека на конкретных операциях.



Рис. 26. Проект дорожной карты роботизации ОАО «РЖД»

4.6. Меры государственной поддержки разработки и внедрения робототехники

Е. Ярославский, руководитель департамента транспорта АНО «Инновационный инжиниринговый центр», представил действующие и перспективные меры государственной поддержки для разработчиков и заказчиков робототехники в интересах железнодорожной отрасли (рис. 27, 28).

Программа «доращивания» поставщиков корпораций: основные условия программы

Цель программы Поддержка проектов, связанных с разработкой (доработкой), созданием (расширением) производства новой продукции под требования российских корпораций Ожидаемый результат Прирост выручки технологической компании от реализации новой продукции (в течение общего срока реализации проекта) – в объеме не менее 5-кратного размера гранта Возврат средств гранта пропорционален объему достижения результатов проекта	Размер гранта От 25 до 250 млн руб. Срок финансирования Не более 3 лет Общий срок реализации проекта Не более 6 лет с начала финансирования Механизм финансирования - гранты (целевое, безвозмездное, безвозвратное) - техзадание на продукт утверждается корпорацией-заказчиком - постальное финансирование и приемка результатов Объем внебюджетного финансирования - Не менее 50% при выручке до 2 млрд руб. - Не менее 70% при выручке от 2 до 4 млрд руб. - Не менее 100% при выручке свыше 4 млрд руб.	Стадия готовности продукта 4 ≤ УПТ ≤ 7 Направления поддержки - ОКР - Изготовление, испытание и сертификация опытного образца - Создание производства (приобретение оборудования) - Изготовление пилотной партии - Испытания и сертификация пилотной партии Направления расходов за внебюджет - Аналогично расходам за бюджет Результаты интеллектуальной деятельности Принадлежат грантополучателю
---	---	--

Рис. 27. Программа поддержки «доращивания» поставщиков

В качестве примеров приведены три проекта в интересах ОАО «РЖД»: роботизированный комплекс «РОИН» МРГК-301ГШ – грант 81 млн руб. по программе доращивания, потребность до 2031 г. 260 единиц; многофункциональная путевая машина А650 – планируемый грант 153 млн руб. по «ТРАНСМАШ», потребность 10 единиц; тяжелый робот «РОИН» Р-701К для труднодоступных территорий – планируемый грант 208 млн руб. по «ТРАНСМАШ», потребность 30 единиц.

Отраслевая программа «ТРАНСМАШ»: основные условия программы

 Цель программы Поддержка проектов, предполагающих проведение научно-исследовательских работ по разработке (доработке) продукции под требования крупных корпораций в отрасли транспортного машиностроения  Ожидаемый результат Прирост выручки технологической компании от реализации инновационной продукции (в течение общего срока реализации проекта) – в объеме не менее 5-кратного размера гранта	 Размер гранта До 500 млн руб.  Срок финансирования АНО ЦПИИ – не более 3 лет Технокомпанией – не более 5 лет Общий срок реализации проекта Не более 7 лет с года заключения договора гранта  Стадия готовности продукта УГТ ≤ 6  Механизм финансирования - гранты (целевое, безвозмездное, безвозвратное) - техзадание на продукт утверждается корпорацией-заказчиком - постатпное финансирование и приемка результатов  Объём внебюджетного финансирования - Не менее 50% от размера гранта	 Направления поддержки - НИР и ОКР (зарплата, соисполнители, накладные) - Приобретение испытательного оборудования - Аренда оборудования, помещений, оснастки - Приобретение изделий сравнения - Изготовление, испытания и сертификация опытного образца - Регистрация РИД  Внебюджетное финансирование - Создание производства (приобретение, лизинг оборудования) - Изготовление, испытания и сертификация пилотной партии  Предусмотрена компенсация понесенных расходов с начала финансового года заключения договора гранта
---	--	--

Рис. 28. Отраслевая программа «ТРАНСМАШ»

Анонсирована программа льготного лизинга промышленной и сервисной робототехники в транспортной отрасли, разрабатываемая совместно с АО «Государственная транспортная лизинговая компания» (ГТЛК); пилотный запуск с участием двух-трех проектов планируется в 2026 г. Механизм позволяет заказчику внедрять роботизированные решения без единовременных капитальных вложений, а разработчику – получить гарантированный сбыт и масштабирование.

4.7. Переход от опытного образца к серийному внедрению

Обсуждение показало, что успешный прототип является лишь началом проекта. Для серийного внедрения операция должна стабильно выполняться при изменении погоды, освещения, состояния пути, типа подвижного состава и действий находящихся рядом работников: на сортировочной станции робот работает с загрязненными и изношенными деталями, снегом, наледью, вибрацией и вариативностью положения механизмов. Поэтому показатель успешности должен дополняться распределением отказов по условиям, временем восстановления, долей вмешательств человека и последствиями неуспешной попытки. Зарубежные прототипы железнодорожных роботов подтверждают эффективность мультисенсорного контроля, мобильных манипуляторов и специализированных платформ и одновременно важность полевых испытаний [47–49].

Коллаборативный режим испытаний робота-расцепщика представляет методический интерес: он обеспечивает непрерывность производственного процесса и собирает редкие сложные примеры. Для перехода к автономному режиму необходимо заранее определить безопасное состояние, порядок повторной попытки, условия вызова оператора и границы зоны совместной работы человека и машины; требования к коллаборативным режимам

промышленных роботов установлены в ISO/TS 15066:2016⁶. Имитационная оценка параметров мобильной платформы комплекса для расформирования составов выполнена специалистами АО «НИИАС» [55].

Тиражирование требует типизации не только робота, но и места его применения: для каждой станции описываются геометрия путей, диапазон скоростей, связь с горочной автоматикой, электроснабжение, каналы передачи данных, климатическое исполнение, требования к ограждению и обслуживанию. Чем больше различий компенсируется программной настройкой и сменным инструментом, тем ниже стоимость адаптации; чрезмерная универсальность, напротив, увеличивает массу, сложность и цену. Целесообразна платформенная линейка с общим шасси, системой безопасности и интерфейсами и специализированными рабочими органами.

4.8. Экономика роботизации и оценка полного эффекта

Сравнение стоимости робота с фондом оплаты труда одного работника недостаточно для железнодорожной отрасли. Роботизированный комплекс влияет на безопасность, простой подвижного состава, потребность в технологических перерывах, качество диагностики, расход материалов и устойчивость работы при кадровом дефиците; часть эффектов не дает немедленной экономии, но предотвращает потери или обеспечивает выполнение растущего объема работ без роста численности.

Методика технико-экономической оценки должна включать капитальные затраты, интеграцию, подготовку инфраструктуры, обучение, запасные части, вычислительные ресурсы, связь, периодическое подтверждение безопасности и модернизацию программного обеспечения; на стороне результата – сокращение трудоемкости, уменьшение простоев, предотвращенные травмы и отказы, рост пропускной способности, высвобождение персонала для квалифицированных операций и повышение достоверности цифровой информации.

Государственные программы закрывают часть риска разработки и создания производства, но не заменяют подтвержденный спрос; необходима ранняя фиксация потребности ОАО «РЖД», единых технических требований и графика закупок.

Заключение и перспективы развития отечественных систем управления железнодорожным транспортом

Проведенный анализ материалов международной научно-практической конференции «PRO//Движение. Транспортные системы управления» позволяет констатировать завершение этапа фрагментарной автоматизации отдельных технологических операций и переход отечественного железнодорожного транспорта к системной цифровой трансформации. В 2025–2026 годах получены результаты, имеющие принципиальное значение для дальнейшего развития отрасли.

На основе результатов конференции можно сформулировать следующие приоритетные задачи для ОАО «РЖД» на 2026–2030 годы:

1. **Завершение комплексных испытаний «Цифровой железнодорожной станции»** и подготовка к тиражированию на другие станции сети. Раз-

⁶ISO/TS 15066:2016. Robots and robotic devices – Collaborative robots. Geneva: ISO, 2016.

работка типовых проектов внедрения и создание центров компетенций для сопровождения цифровых станций.

2. **Разработка корпоративной программы роботизации** и методики оценки экономической эффективности проектов роботизации с учетом многокритериальных эффектов (социальных, экологических, стратегических).
3. **Активное участие в создании нормативной базы** для применения искусственного интеллекта и робототехники на железнодорожном транспорте, включая стандарты безопасности для систем с применением ИИ и регламенты взаимодействия «человек – робот».
4. **Активное развитие ключевых компетенций в области физического ИИ**, формирование набора приоритетных пилотных проектов по применению ИИ при организации управления движением, в том числе создание отраслевого Центра обучения роботов ОАО «РЖД» для подготовки роботов к работе в реальных условиях эксплуатации, формирование центра верификации ИИ-продуктов в холдинге «РЖД».
5. **Продолжение работ по созданию единой цифровой платформы** управления перевозочным процессом, объединяющей более двух сотен автоматизированных систем в сквозной технологический процесс.

Решение указанных задач обеспечит переход отечественного железнодорожного транспорта на качественно новый уровень управления, характеризующийся ростом пропускной и провозной способности сети, снижением эксплуатационных затрат, повышением безопасности перевозочного процесса и улучшением условий труда персонала. Технологический суверенитет, подтвержденный локализацией критических компонентов и независимостью от зарубежных поставщиков, создает конкурентные преимущества на пространстве колеи 1520 мм и формирует экспортный потенциал отечественных разработок в области систем управления железнодорожным транспортом.

Список литературы

- [1] Quaglietta, E. The ON-TIME real-time railway traffic management framework: a proof-of-concept using a scalable standardised data communication architecture / E. Quaglietta [et al.] // *Transportation Research Part C*. – 2016. – Vol. 63. – P. 23–50. – DOI 10.1016/j.trc.2015.11.014.
- [2] Tang, R. A literature review of Artificial Intelligence applications in railway systems / R. Tang [et al.] // *Transportation Research Part C*. – 2022. – Vol. 140. – Art. 103679. – DOI 10.1016/j.trc.2022.103679.
- [3] Bešinović, N. Artificial Intelligence in Railway Transport: Taxonomy, Regulations, and Applications / N. Bešinović [et al.] // *IEEE Transactions on Intelligent Transportation Systems*. – 2022. – Vol. 23, No. 9. – P. 14011–14024. – DOI 10.1109/TITS.2021.3131637.
- [4] Шубинский, И. Б. Общие положения обоснования функциональной безопасности интеллектуальных систем на железнодорожном транспорте / И. Б. Шубинский, Е. Н. Розенберг // *Надежность*. – 2023. – Т. 23, № 3. – С. 38–45. – DOI 10.21683/1729-2646-2023-23-3-38-45. – EDN HGNXNT.
- [5] Шубинский, И. Б. Надёжность, риски, безопасность систем управления на железнодорожном транспорте / И. Б. Шубинский, Е. Н. Розенберг, А. В. Бочков. – Вологда; Москва: ООО «Инфра-Инженерия», 2024. – 414 с. – ISBN 978-5-9729-1992-5. – EDN FMONU.
- [6] Бочков, К. А. Обзор архитектуры, особенностей и методов обеспечения безопасности в автоматизированных системах управления ответственными технологическими процессами

- железнодорожного транспорта / К. А. Бочков, С. Н. Харлап, Е. П. Литвинов // Проблемы безопасности на транспорте : Материалы XIV Международной научно-практической конференции, посвященной пятилетке качества. В 2-х частях, Гомель, 20–21 ноября 2025 года. – Гомель: Белорусский государственный университет транспорта, 2025. – С. 140–143. – EDN WJIMLJ.
- [7] Розенберг, Е. Н. Общие подходы к доказательству безопасности автономных систем / Е. Н. Розенберг, П. А. Попов, Д. В. Талалаев, А. М. Ольшанский, Н. А. Бояринова // Автоматика, связь, информатика. – 2022. – № 1. – С. 2–9.
- [8] Розенберг, Е. Н. Применение технологии виртуальной сцепки на Восточном полигоне / Е. Н. Розенберг, М. А. Дежков, В. Г. Новиков // Автоматика, связь, информатика. – 2025. – № 8. – С. 7–11. – DOI 10.62994/AT.2025.8.8.002.
- [9] Felez, J. Virtual Coupling in Railways: A Comprehensive Review / J. Felez, M. A. Vaquero-Serrano // Machines. – 2023. – Vol. 11, No. 5. – Art. 521. – DOI 10.3390/machines11050521.
- [10] Долгий, А. И. Нечетко-логический подход к автоматическому анализу показателей грузовой станции на основе данных «от колеса» / А. И. Долгий, А. В. Суханов // Интеллектуальные транспортные системы: материалы V Международной научно-практической конференции (Москва, 21 мая 2026 г.). – М.: РУТ (МИИТ), 2026. – С. 139–145. – DOI 10.30932/9785002709564-2026-139-145.
- [11] Долгий, А. И. Классификация режимов работы железнодорожной станции на основе реальных данных «от колеса» / А. И. Долгий, А. Е. Хатламаджиян, И. А. Ольгейзер, А. В. Суханов // Вестник Ростовского государственного университета путей сообщения. – 2025. – № 1(97). – С. 58–68. – DOI 10.46973/0201-727X_2025_1_58.
- [12] Хатламаджиян, А. Е. Формирование объективных показателей работы сортировочной станции на основе данных «от колеса» / А. Е. Хатламаджиян, И. А. Ольгейзер, А. В. Суханов, В. С. Иерусалимов // Автоматика на транспорте. – 2024. – Т. 10, № 3. – С. 254–268. – DOI 10.20295/2412-9186-2024-10-03-254-268.
- [13] Goverde, R. M. P. Railway timetable stability analysis using max-plus system theory / R. M. P. Goverde // Transportation Research Part B. – 2007. – Vol. 41, No. 2. – P. 179–201. – DOI 10.1016/j.trb.2006.02.003.
- [14] D'Ariano, A. Reordering and local rerouting strategies to manage train traffic in real time / A. D'Ariano, F. Corman, D. Pacciarelli, M. Pranzo // Transportation Science. – 2008. – Vol. 42, No. 4. – P. 405–419. – DOI 10.1287/trsc.1080.0247.
- [15] Lusby, R. M. Railway track allocation: models and methods / R. M. Lusby, J. Larsen, M. Ehrgott, D. Ryan // OR Spectrum. – 2011. – Vol. 33, No. 4. – P. 843–883. – DOI 10.1007/s00291-009-0189-0.
- [16] Corman, F. Bi-objective conflict detection and resolution in railway traffic management / F. Corman, A. D'Ariano, D. Pacciarelli, M. Pranzo // Transportation Research Part C. – 2012. – Vol. 20, No. 1. – P. 79–94. – DOI 10.1016/j.trc.2010.09.009.
- [17] Андреев, В. Е. Цифровая железнодорожная станция – от концепции к реальному внедрению / В. Е. Андреев, А. И. Долгий, В. В. Кудюкин, А. Е. Хатламаджиян, С. Ю. Гришаев, И. А. Ольгейзер // Автоматика, связь, информатика. – 2023. – № 9. – С. 2–6. – DOI 10.34649/AT.2023.9.9.001.
- [18] Kaewunruen, S. Digital Twins for Managing Railway Bridge Maintenance, Resilience, and Climate Change Adaptation / S. Kaewunruen, J. Sresakoolchai, Y.-H. Lin // Sensors. – 2023. – Vol. 23, No. 1. – Art. 252. – DOI 10.3390/s23010252.
- [19] van Dinter, R. Predictive maintenance using digital twins: A systematic literature review / R. van Dinter, B. Tekinerdogan, C. Catal // Information and Software Technology. – 2022. – Vol. 151. – Art. 107008. – DOI 10.1016/j.infsof.2022.107008.
- [20] Padovano, A. Improving safety management in railway stations through a simulation-based digital twin approach / A. Padovano, F. Longo, L. Manca, R. Grugni // Computers & Industrial Engineering. – 2024. – Vol. 187. – Art. 109839. – DOI 10.1016/j.cie.2023.109839.
- [21] Zhang, J. Artificial Intelligence Applied on Traffic Planning and Management for Rail Transport: A Review and Perspective / J. Zhang, J. Zhang // Discrete Dynamics in Nature and Society. – 2023. – Art. 1832501. – DOI 10.1155/2023/1832501.

- [22] Труды АО «НИИАС». Вып. 11, т. 3 / ред. совет: И. Н. Розенберг (пред.) [и др.]. – М.: Т8 Издательские Технологии, 2021. – С. 7–68.
- [23] Chen, M. Evaluating Large Language Models Trained on Code / M. Chen [et al.] // arXiv. – 2021. – arXiv:2107.03374. – DOI 10.48550/arXiv.2107.03374.
- [24] Greshake, K. Not What You've Signed Up For: Compromising Real-World LLM-Integrated Applications with Indirect Prompt Injection / K. Greshake, S. Abdelnabi, S. Mishra, C. Endres, T. Holz, M. Fritz // Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISeC '23). – New York: ACM, 2023. – P. 79–90. – DOI 10.1145/3605764.3623985.
- [25] Моисеев, В. В. Обеспечение кибербезопасности систем микропроцессорных и релейно-процессорных централизаций на сети ОАО «РЖД»: перспективы и практическое применение средств защиты / В. В. Моисеев // Автоматика на транспорте. – 2022. – Т. 8, № 3. – С. 266–275. – DOI 10.20295/2412-9186-2022-8-03-266-275.
- [26] Лецкий, Э. К. Интегрирование искусственного интеллекта в функциональное тестирование критической информационной инфраструктуры железнодорожного транспорта / Э. К. Лецкий, А. И. Гуров // Интеллектуальные технологии на транспорте. – 2025. – № 1. – С. 14–19. – DOI 10.20295/2413-2527-2025-141-14-19.
- [27] De Donato, L. A Survey on Audio-Video Based Defect Detection Through Deep Learning in Railway Maintenance / L. De Donato [et al.] // IEEE Access. – 2022. – Vol. 10. – P. 65376–65400. – DOI 10.1109/ACCESS.2022.3183102.
- [28] Alif, M. A. R. Advances in computer vision for comprehensive railway engineering: from track inspection to rolling stock and safety monitoring / M. A. R. Alif, G. Tucker, M. Hussain // Railway Engineering Science. – 2026. – DOI 10.1007/s40534-026-00434-7.
- [29] Olivier, B. A Review of Computer Vision for Railways / B. Olivier [et al.] // IEEE Transactions on Intelligent Transportation Systems. – 2025. – Vol. 26, No. 7. – P. 11034–11065. – DOI 10.1109/TITS.2025.3552011.
- [30] Kumar, A. A systematic literature review of defect detection in railways using machine vision-based inspection methods / A. Kumar, S. P. Harsha // International Journal of Transportation Science and Technology. – 2025. – Vol. 18. – P. 207–226. – DOI 10.1016/j.ijtst.2024.06.006.
- [31] Долгий, А. И. Инновационные алгоритмы машинного зрения для диагностики продольного профиля сортировочных путей / А. И. Долгий, А. Е. Хатламаджиян, И. А. Ольгейзер, А. В. Суханов, К. И. Корниенко // Автоматика, связь, информатика. – 2022. – № 8. – С. 7–9. – DOI 10.34649/AT.2022.8.8.002.
- [32] Ольгейзер, И. А. Компьютерное зрение как способ интеллектуализации систем горочной автоматизации / И. А. Ольгейзер, А. В. Суханов, А. М. Лященко, Д. В. Глазунов // Проблемы машиностроения и автоматизации. – 2022. – № 1. – С. 46–53. – DOI 10.52261/02346206_2022_1_46.
- [33] Васильева, С. А. Автоматизация прогноза бокового износа рельсов в кривых как элемент предиктивного анализа состояния пути / С. А. Васильева, А. В. Лохач // Путь и путевое хозяйство. – 2023. – № 11. – С. 5–7.
- [34] Ristić-Durrant, D. A Review of Vision-Based On-Board Obstacle Detection and Distance Estimation in Railways / D. Ristić-Durrant, M. Franke, K. Michels // Sensors. – 2021. – Vol. 21, No. 10. – Art. 3452. – DOI 10.3390/s21103452.
- [35] Pappaterra, M. J. A Systematic Review of Artificial Intelligence Public Datasets for Railway Applications / M. J. Pappaterra, F. Flammini, V. Vittorini, N. Bešinović // Infrastructures. – 2021. – Vol. 6, No. 10. – Art. 136. – DOI 10.3390/infrastructures6100136.
- [36] Gan, J. A Hierarchical Extractor-Based Visual Rail Surface Inspection System / J. Gan, Q. Li, J. Wang, H. Yu // IEEE Sensors Journal. – 2017. – Vol. 17, No. 23. – P. 7935–7944. – DOI 10.1109/JSEN.2017.2761858.
- [37] Ярмолинский, Ф. А. Особенности использования Big Data при исследовании грузопотоков на железнодорожном транспорте / Ф. А. Ярмолинский, О. Д. Покровская, Е. Д. Пасечник, Е. В. Пакулина, А. А. Трапезников // Бюллетень результатов научных исследований. – 2024. – № 1. – С. 107–122. – DOI 10.20295/2223-9987-2024-01-107-122.

- [38] Пат. 2524505 С1 Российская Федерация. Способ управления движением поездов с использованием вариантных графиков / И. Р. Гургенидзе, Б. Н. Курагин, М. Г. Лысиков, С. И. Ляшенко, А. В. Степанов, С. В. Торорошенко; патентообладатель ОАО «НИИАС». – Оpubл. 2014.
- [39] Zhang, Q. A heuristic approach to integrate train timetabling, platforming, and railway network maintenance scheduling decisions / Q. Zhang, R. M. Lusby, P. Shang, X. Zhu // *Transportation Research Part B*. – 2022. – Vol. 158. – P. 210–238. – DOI 10.1016/j.trb.2022.02.002.
- [40] Сидоренко, В. Г. Влияние цифровизации городских рельсовых транспортных систем на процессы планирования и управления движением поездов / В. Г. Сидоренко, М. А. Кулагин, А. Е. Родина // *Автоматика на транспорте*. – 2025. – Т. 11, № 1. – С. 30–43. – DOI 10.20295/2412-9186-2025-11-01-30-43.
- [41] Алексеев, В. М. Построение архитектуры интеллектуальной системы управления городской рельсовой транспортной системой / В. М. Алексеев, Л. А. Баранов, М. А. Кулагин, В. Г. Сидоренко // *Мир транспорта*. – 2021. – Т. 19, № 1. – С. 18–46. – DOI 10.30932/1992-3252-2021-19-1-18-46.
- [42] Сидоренко, В. Г. Разработка информационного обеспечения интеллектуальной системы управления городскими рельсовыми транспортными системами / В. Г. Сидоренко, Л. Н. Логинова, А. И. Сафронов // *Автоматика на транспорте*. – 2023. – Т. 9, № 2. – С. 178–192. – DOI 10.20295/2412-9186-2023-9-02-178-192.
- [43] Сидоренко, В. Г. Опыт и перспективы автоматизации управления перевозочным процессом скоростного транспорта городских агломераций / В. Г. Сидоренко, Е. В. Копылова, А. И. Сафронов, М. А. Туманов // *Автоматика на транспорте*. – 2023. – Т. 9, № 1. – С. 33–48. – DOI 10.20295/2412-9186-2023-9-01-33-48.
- [44] Тарасов, К. А. АСУ «КАМА». Энергооптимальное ведение пакета поездов / К. А. Тарасов, М. Ю. Шеин, Н. А. Бояринова // *Железнодорожный транспорт*. – 2026. – № 5. – С. 18–19.
- [45] Кононов, А. Ф. Проекты в сфере роботизации / А. Ф. Кононов, В. Д. Макогон, А. Е. Хатламаджиян, А. Е. Шапекин // *Автоматика, связь, информатика*. – 2026. – № 3. – С. 15–17. – DOI 10.62994/AT.2026.3.3.003.
- [46] Куприяновский, В. П. Формализованные онтологии и сервисы для высокоскоростных магистралей и цифровой железной дороги / В. П. Куприяновский, О. Н. Покусаев, Ю. И. Волокитин, Д. Е. Намиот, И. П. Петрунина, А. В. Зажигалкин // *International Journal of Open Information Technologies*. – 2018. – Vol. 6, No. 3. – P. 69–86.
- [47] Kasch, J. Design and Operational Assessment of a Railroad Track Robot for Railcar Undercarriage Condition Inspection / J. Kasch, M. Ahmadian // *Designs*. – 2024. – Vol. 8, No. 4. – Art. 70. – DOI 10.3390/designs8040070.
- [48] Liu, H. An autonomous rail-road amphibious robotic system for railway maintenance using sensor fusion and mobile manipulator / H. Liu [et al.] // *Computers & Electrical Engineering*. – 2023. – Art. 108874. – DOI 10.1016/j.compeleceng.2023.108874.
- [49] Daniyan, I. A. Design of a robot for inspection and diagnostic operations of rail track facilities / I. A. Daniyan, K. Mpofu, S. O. Nwankwo // *International Journal of Quality & Reliability Management*. – 2023. – Vol. 40, No. 3. – P. 653–673. – DOI 10.1108/IJQRM-03-2020-0083.
- [50] Brohan, A. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control / A. Brohan [et al.] // *arXiv*. – 2023. – arXiv:2307.15818. – DOI 10.48550/arXiv.2307.15818.
- [51] O'Neill, A. Open X-Embodiment: Robotic Learning Datasets and RT-X Models / A. O'Neill [et al.] // *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. – 2024. – DOI 10.1109/ICRA57147.2024.10611477.
- [52] Kim, M. J. OpenVLA: An Open-Source Vision-Language-Action Model / M. J. Kim [et al.] // *arXiv*. – 2024. – arXiv:2406.09246. – DOI 10.48550/arXiv.2406.09246.
- [53] Попов, П. А. Применение передовых технологий для работы в автоматическом режиме на МЦК / П. А. Попов // *Железнодорожный транспорт*. – 2020. – № 11. – С. 17–21.
- [54] Zhao, Y. VLA-RAIL: A Real-Time Asynchronous Inference Linker for VLA Models and Robots / Y. Zhao, L. Zhao, B. Cheng, G. Yao, X. Wen, H. Gao // *arXiv*. – 2025. – arXiv:2512.24673. – URL: <https://arxiv.org/abs/2512.24673>.

- [55] Кудюкин, В. В. Имитационное моделирование работы робототехнических комплексов, предназначенных для расформирования составов на сортировочных горках / В. В. Кудюкин, А. В. Вуколов, В. С. Кузьмин // Автоматика на транспорте. – 2025. – Т. 11, № 1. – С. 16–29. – DOI 10.20295/2412-9186-2025-11-01-16-29.

CURRENT STATE OF INTELLECTUALIZATION, DIGITALIZATION AND ROBOTIZATION SYSTEMS OF DOMESTIC RAILWAY TRANSPORT

**Bochkov A. V.^{1, a}, Dolgiy A. I.^{1, b}, Zazhigalkin A. V.^{1, 2, c}, Rozenberg E. N.^{1, d},
Sukhanov A. V.^{1, e}, Khatlamadzhiyan A. E.^{1, f}, Chernukha A. V.^{1, g}**

¹ Joint Stock Company «Research and Design Institute for Informatization, Automation and Communication in Railway Transport» (JSC «NIIAS»), Moscow, Russia

² Association «Transport Science», Moscow, Russia

E-mail: ^a a.bochkov@vniias.ru, ^b a.dolgiy@vniias.ru, ^c zashigalkin@mail.ru, ^d S.Lavruhina@vniias.ru,
^e a.suhanov@vniias.ru, ^f a.khatlamadzhiyan@vniias.ru, ^g a.chernukha@vniias.ru

Abstract. The paper systematizes the results of the scientific and practical conference «PRO//Movement. Transport Management Systems» (August 18–20, 2026, Nizhny Novgorod), dedicated to the digitalization of the transportation process, artificial intelligence, and the robotization of railway transport. The conference was attended by the management of the Russian Railways holding, the scientific sector complex, and manufacturing enterprises: managers, scientists, and engineers. The target model for managing freight flows, the «Digital Railway Station» project, and technologies such as virtual coupling, machine vision, predictive diagnostics, and physical artificial intelligence were discussed. Special attention is paid to the role of the scientific activities of JSC «NIIAS» in shaping a unified data architecture, models, and executive systems, as well as to issues of functional safety, economic efficiency, and the replication of developments. The prospects for the development of artificial intelligence in the railway industry are discussed.

Keywords: *railway transport, digital station, traffic control, artificial intelligence, computer vision, robotics, digital twin, functional safety.*

References

- [1] Quaglietta, E., Pellegrini, P., Goverde, R. M. P., Albrecht, T., Jaekel, B., Marlière, G., Rodriguez, J., Dollevoet, T., Ambrogio, B., Carcasole, D., Giaroli, M., & Nicholson, G. (2016). The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture. *Transportation Research Part C*, 63, 23–50. <https://doi.org/10.1016/j.trc.2015.11.014>

For citation:

Bochkov A. V., Dolgiy A. I., Zazhigalkin A. V., Rozenberg E. N., Sukhanov A. V., Khatlamadzhiyan A. E., Chernukha A. V. Current State of Intellectualization, Digitalization and Robotization Systems of Domestic Railway Transport. *Intelligent transport*, 2026, no. 3(39), pp. 115–158. DOI: 10.24412/3033-6007-2026-339-115-158. (in Russian)

- [2] Tang, R., De Donato, L., Bešinović, N., Flammini, F., Goverde, R. M. P., Lin, Z., Liu, R., Tang, T., Vittorini, V., & Wang, Z. (2022). A literature review of artificial intelligence applications in railway systems. *Transportation Research Part C*, 140, Article 103679. <https://doi.org/10.1016/j.trc.2022.103679>
- [3] Bešinović, N., De Donato, L., Flammini, F., Goverde, R. M. P., Lin, Z., Liu, R., Marrone, S., Nardone, R., Tang, T., & Vittorini, V. (2022). Artificial intelligence in railway transport: Taxonomy, regulations, and applications. *IEEE Transactions on Intelligent Transportation Systems*, 23(9), 14011–14024. <https://doi.org/10.1109/TITS.2021.3131637>
- [4] Shubinskiy, I. B., & Rozenberg, E. N. (2023). General provisions for justifying the functional safety of intelligent systems in railway transport. *Dependability*, 23(3), 38–45. <https://doi.org/10.21683/1729-2646-2023-23-3-38-45> (in Russian)
- [5] Shubinskiy, I. B., Rozenberg, E. N., & Bochkov, A. V. (2024). *Reliability, risks, and safety of control systems in railway transport*. Infra-Inzheneriya. (in Russian)
- [6] Bochkov, K. A., Kharlap, S. N., & Litvinov, E. P. (2025). Review of architecture, features and methods of ensuring safety in automated control systems for critical technological processes of railway transport. In *Problemy bezopasnosti na transporte: Materialy XIV Mezhdunarodnoy nauchno-prakticheskoy konferentsii, posvyashchennoy pyatiletke kachestva* (pp. 140–143). Belorusskiy gosudarstvennyy universitet transporta. (in Russian)
- [7] Rozenberg, E. N., Popov, P. A., Talalayev, D. V., Olshanskiy, A. M., & Boyarinova, N. A. (2022). General approaches to safety justification of autonomous systems. *Automation, Communications, Informatics*, (1), 2–9. (in Russian)
- [8] Rozenberg, E. N., Dezhkov, M. A., & Novikov, V. G. (2025). Application of virtual coupling technology at the Eastern polygon. *Automation, Communications, Informatics*, (8), 7–11. <https://doi.org/10.62994/AT.2025.8.8.002> (in Russian)
- [9] Felez, J., & Vaquero-Serrano, M. A. (2023). Virtual coupling in railways: A comprehensive review. *Machines*, 11(5), Article 521. <https://doi.org/10.3390/machines11050521>
- [10] Dolgiy, A. I., & Sukhanov, A. V. (2026). Fuzzy-logic approach to automatic analysis of freight station indicators based on “from the wheel” data. In *Intellektualnyye transportnyye sistemy: Materialy V Mezhdunarodnoy nauchno-prakticheskoy konferentsii (Moskva, 21 maya 2026 g.)* (pp. 139–145). RUT (MIIT). <https://doi.org/10.30932/9785002709564-2026-139-145> (in Russian)
- [11] Dolgiy, A. I., Khatlamadzhian, A. E., Olgezyer, I. A., & Sukhanov, A. V. (2025). Classification of railway station operation modes based on real “from the wheel” data. *Vestnik RGUPS*, (1(97)), 58–68. https://doi.org/10.46973/0201-727X_2025_1_58 (in Russian)
- [12] Khatlamadzhian, A. E., Olgezyer, I. A., Sukhanov, A. V., & Iyerusalimov, V. S. (2024). Formation of objective indicators of marshalling yard operation based on “from the wheel” data. *Automation on Transport*, 10(3), 254–268. <https://doi.org/10.20295/2412-9186-2024-10-03-254-268> (in Russian)
- [13] Goverde, R. M. P. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B*, 41(2), 179–201. <https://doi.org/10.1016/j.trb.2006.02.003>
- [14] D’Ariano, A., Corman, F., Pacciarelli, D., & Pranzo, M. (2008). Reordering and local rerouting strategies to manage train traffic in real time. *Transportation Science*, 42(4), 405–419. <https://doi.org/10.1287/trsc.1080.0247>
- [15] Lusby, R. M., Larsen, J., Ehrgott, M., & Ryan, D. (2011). Railway track allocation: Models and methods. *OR Spectrum*, 33(4), 843–883. <https://doi.org/10.1007/s00291-009-0189-0>
- [16] Corman, F., D’Ariano, A., Pacciarelli, D., & Pranzo, M. (2012). Bi-objective conflict detection and resolution in railway traffic management. *Transportation Research Part C*, 20(1), 79–94. <https://doi.org/10.1016/j.trc.2010.09.009>
- [17] Andreyev, V. E., Dolgiy, A. I., Kudyukin, V. V., Khatlamadzhian, A. E., Grishayev, S. Yu., & Olgezyer, I. A. (2023). Digital railway station – from concept to real implementation. *Automation, Communications, Informatics*, (9), 2–6. <https://doi.org/10.34649/AT.2023.9.9.001> (in Russian)

- [18] Kaewunruen, S., Sresakoolchai, J., & Lin, Y.-H. (2023). Digital twins for managing railway bridge maintenance, resilience, and climate change adaptation. *Sensors*, 23(1), Article 252. <https://doi.org/10.3390/s23010252>
- [19] van Dinter, R., Tekinerdogan, B., & Catal, C. (2022). Predictive maintenance using digital twins: A systematic literature review. *Information and Software Technology*, 151, Article 107008. <https://doi.org/10.1016/j.infsof.2022.107008>
- [20] Padovano, A., Longo, F., Manca, L., & Grugni, R. (2024). Improving safety management in railway stations through a simulation-based digital twin approach. *Computers & Industrial Engineering*, 187, Article 109839. <https://doi.org/10.1016/j.cie.2023.109839>
- [21] Zhang, J., & Zhang, J. (2023). Artificial intelligence applied on traffic planning and management for rail transport: A review and perspective. *Discrete Dynamics in Nature and Society*, Article 1832501. <https://doi.org/10.1155/2023/1832501>
- [22] Rozenberg, I. N. (Ed.). (2021). *Proceedings of JSC "NIIAS"* (Issue 11, Vol. 3, pp. 7–68). T8 Izdatelskiye Tekhnologii. (in Russian)
- [23] Chen, M., Tworek, J., Jun, H., Yuan, Q., Ponde de Oliveira Pinto, H., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., ... Zaremba, W. (2021). *Evaluating large language models trained on code* (arXiv:2107.03374). arXiv. <https://doi.org/10.48550/arXiv.2107.03374>
- [24] Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection. In *Proceedings of the 16th ACM Workshop on Artificial Intelligence and Security (AISec '23)* (pp. 79–90). ACM. <https://doi.org/10.1145/3605764.3623985>
- [25] Moiseyev, V. V. (2022). Ensuring cybersecurity of microprocessor and relay-processor interlocking systems on the network of JSC "Russian Railways": Prospects and practical application of protection tools. *Automation on Transport*, 8(3), 266–275. <https://doi.org/10.20295/2412-9186-2022-8-03-266-275> (in Russian)
- [26] Letskiy, E. K., & Gurov, A. I. (2025). Integration of artificial intelligence into functional testing of critical information infrastructure of railway transport. *Intellectual Technologies on Transport*, (1), 14–19. <https://doi.org/10.20295/2413-2527-2025-141-14-19> (in Russian)
- [27] De Donato, L., Flammini, F., Marrone, S., Mazzariello, C., Nardone, R., Sansone, C., & Vittorini, V. (2022). A survey on audio-video based defect detection through deep learning in railway maintenance. *IEEE Access*, 10, 65376–65400. <https://doi.org/10.1109/ACCESS.2022.3183102>
- [28] Alif, M. A. R., Tucker, G., & Hussain, M. (2026). Advances in computer vision for comprehensive railway engineering: From track inspection to rolling stock and safety monitoring. *Railway Engineering Science*. <https://doi.org/10.1007/s40534-026-00434-7>
- [29] Olivier, B., Guo, F., Qian, Y., & Connolly, D. P. (2025). A review of computer vision for railways. *IEEE Transactions on Intelligent Transportation Systems*, 26(7), 11034–11065. <https://doi.org/10.1109/TITS.2025.3552011>
- [30] Kumar, A., & Harsha, S. P. (2025). A systematic literature review of defect detection in railways using machine vision-based inspection methods. *International Journal of Transportation Science and Technology*, 18, 207–226. <https://doi.org/10.1016/j.ijst.2024.06.006>
- [31] Dolgiy, A. I., Khatlamadzhyan, A. E., Olgeyzer, I. A., Sukhanov, A. V., & Korniyenko, K. I. (2022). Innovative machine vision algorithms for diagnosing the longitudinal profile of marshalling tracks. *Automation, Communications, Informatics*, (8), 7–9. <https://doi.org/10.34649/AT.2022.8.8.002> (in Russian)
- [32] Olgeyzer, I. A., Sukhanov, A. V., Lyashchenko, A. M., & Glazunov, D. V. (2022). Computer vision as a way of intellectualizing hump automation systems. *Engineering Automation Problems*, (1), 46–53. https://doi.org/10.52261/02346206_2022_1_46 (in Russian)
- [33] Vasilyeva, S. A., & Lokhach, A. V. (2023). Automation of lateral rail wear prediction in curves as an element of predictive analysis of track condition. *Put i putevoye khozyaystvo*, (11), 5–7. (in Russian)

- [34] Ristić-Durrant, D., Franke, M., & Michels, K. (2021). A review of vision-based on-board obstacle detection and distance estimation in railways. *Sensors*, 21(10), Article 3452. <https://doi.org/10.3390/s21103452>
- [35] Pappaterra, M. J., Flammini, F., Vittorini, V., & Bešinović, N. (2021). A systematic review of artificial intelligence public datasets for railway applications. *Infrastructures*, 6(10), Article 136. <https://doi.org/10.3390/infrastructures6100136>
- [36] Gan, J., Li, Q., Wang, J., & Yu, H. (2017). A hierarchical extractor-based visual rail surface inspection system. *IEEE Sensors Journal*, 17(23), 7935–7944. <https://doi.org/10.1109/JSSEN.2017.2761858>
- [37] Yarmolinskiy, F. A., Pokrovskaya, O. D., Pasechnik, E. D., Pakulina, E. V., & Trapeznikov, A. A. (2024). Features of using Big Data in the study of freight flows in railway transport. *Bulletin of Research Results*, (1), 107–122. <https://doi.org/10.20295/2223-9987-2024-01-107-122> (in Russian)
- [38] Gurgenidze, I. R., Kuranin, B. N., Lysikov, M. G., Lyashenko, S. I., Stepanov, A. V., & Tororoshenko, S. V. (2014). *Method of train traffic control using variant schedules* (Russian Federation Patent No. 2524505 C1). OAO “NIIAS”. (in Russian)
- [39] Zhang, Q., Lusby, R. M., Shang, P., & Zhu, X. (2022). A heuristic approach to integrate train timetabling, platforming, and railway network maintenance scheduling decisions. *Transportation Research Part B*, 158, 210–238. <https://doi.org/10.1016/j.trb.2022.02.002>
- [40] Sidorenko, V. G., Kulagin, M. A., & Rodina, A. E. (2025). Impact of digitalization of urban rail transport systems on train planning and traffic management processes. *Automation on Transport*, 11(1), 30–43. <https://doi.org/10.20295/2412-9186-2025-11-01-30-43> (in Russian)
- [41] Alekseyev, V. M., Baranov, L. A., Kulagin, M. A., & Sidorenko, V. G. (2021). Building the architecture of an intelligent control system for an urban rail transport system. *World of Transport and Transportation*, 19(1), 18–46. <https://doi.org/10.30932/1992-3252-2021-19-1-18-46> (in Russian)
- [42] Sidorenko, V. G., Loginova, L. N., & Safronov, A. I. (2023). Development of information support for an intelligent control system of urban rail transport systems. *Automation on Transport*, 9(2), 178–192. <https://doi.org/10.20295/2412-9186-2023-9-02-178-192> (in Russian)
- [43] Sidorenko, V. G., Kopylova, E. V., Safronov, A. I., & Tumanov, M. A. (2023). Experience and prospects of automation of transportation process management for high-speed transport in urban agglomerations. *Automation on Transport*, 9(1), 33–48. <https://doi.org/10.20295/2412-9186-2023-9-01-33-48> (in Russian)
- [44] Tarasov, K. A., Shein, M. Yu., & Boyarinova, N. A. (2026). ACS “KAMA”: Energy-optimal operation of a train package. *Zheleznodorozhnyy transport*, (5), 18–19. (in Russian)
- [45] Kononov, A. F., Makogon, V. D., Khatlamadzhiyan, A. E., & Shapekin, A. E. (2026). Projects in the field of robotization. *Automation, Communications, Informatics*, (3), 15–17. <https://doi.org/10.62994/AT.2026.3.3.003> (in Russian)
- [46] Kupriyanovskiy, V. P., Pokusayev, O. N., Volokitin, Yu. I., Namiot, D. E., Petrunina, I. P., & Zazhigalkin, A. V. (2018). Formalized ontologies and services for high-speed railways and the digital railway. *International Journal of Open Information Technologies*, 6(3), 69–86. (in Russian)
- [47] Kasch, J., & Ahmadian, M. (2024). Design and operational assessment of a railroad track robot for railcar undercarriage condition inspection. *Designs*, 8(4), Article 70. <https://doi.org/10.3390/designs8040070>
- [48] Liu, H., Rahman, M., Rahimi, M., Starr, A., Durazo-Cardenas, I., Ruiz-Carcel, C., Ompusunggu, A., Hall, A., & Anderson, R. (2023). An autonomous rail-road amphibious robotic system for railway maintenance using sensor fusion and mobile manipulator. *Computers & Electrical Engineering*, Article 108874. <https://doi.org/10.1016/j.compeleceng.2023.108874>
- [49] Daniyan, I. A., Mpofu, K., & Nwankwo, S. O. (2023). Design of a robot for inspection and diagnostic operations of rail track facilities. *International Journal of Quality & Reliability Management*, 40(3), 653–673. <https://doi.org/10.1108/IJQRM-03-2020-0083>

- [50] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., Florence, P., Fu, C., Gonzalez Arenas, M., Gopalakrishnan, K., Han, K., Hausman, K., Herzog, A., Hsu, J., Ichter, B., ... Zitkovich, B. (2023). *RT-2: Vision-language-action models transfer web knowledge to robotic control* (arXiv:2307.15818). arXiv. <https://doi.org/10.48550/arXiv.2307.15818>
- [51] O'Neill, A., Rehman, A., Maddukuri, A., Gupta, A., Padalkar, A., Lee, A., Pooley, A., Gupta, A., Mandlekar, A., Jain, A., Tung, A., Bewley, A., Herzog, A., Irpan, A., Khazatsky, A., Rai, A., Gupta, A., Wang, A., Singh, A., ... Lin, Z. (2024). Open X-Embodiment: Robotic learning datasets and RT-X models. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. <https://doi.org/10.1109/ICRA57147.2024.10611477>
- [52] Kim, M. J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E., Lam, G., Sanketi, P., Vuong, Q., Kollar, T., Burchfiel, B., Tedrake, R., Sadigh, D., Levine, S., Liang, P., & Finn, C. (2024). *OpenVLA: An open-source vision-language-action model* (arXiv:2406.09246). arXiv. <https://doi.org/10.48550/arXiv.2406.09246>
- [53] Popov, P. A. (2020). Application of advanced technologies for automatic operation at the Moscow Central Circle. *Zheleznodorozhnyy transport*, (11), 17–21. (in Russian)
- [54] Zhao, Y., Zhao, L., Cheng, B., Yao, G., Wen, X., & Gao, H. (2025). *VLA-RAIL: A real-time asynchronous inference linker for VLA models and robots* (arXiv:2512.24673). arXiv. <https://arxiv.org/abs/2512.24673>
- [55] Kudyukin, V. V., Vukolov, A. V., & Kuzmin, V. S. (2025). Simulation modeling of robotic complexes designed for train breaking-up at marshalling humps. *Automation on Transport*, 11(1), 16–29. <https://doi.org/10.20295/2412-9186-2025-11-01-16-29> (in Russian)

