



На правах рукописи

Городничев Михаил Геннадьевич

**Методология создания информационной системы
управления транспортными потоками**

Специальность 2.3.8. Информатика и информационные процессы
(технические науки)

Автореферат
диссертации на соискание учёной степени
доктора технических наук

Москва — 2026

Работа выполнена на кафедре «Математическая кибернетика и информационные технологии» ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ).

Научный консультант: доктор технических наук, доцент
Яшина Марина Викторовна

Официальные оппоненты: **Келлер Андрей Владимирович**,
доктор технических наук, профессор
директор Набережночелнинского института (филиала) ФГАОУ ВО КФУ

Шевцова Анастасия Геннадьевна,
доктор технических наук, доцент
директор института дополнительного образования и профессионального обучения «Высшая технологическая школа БГТУ имени В.Г. Шухова» БГТУ им. В.Г. Шухова

Олейникова Светлана Александровна,
доктор технических наук, профессор
профессор кафедры автоматизированных и вычислительных систем Факультета информационных технологий и компьютерной безопасности ФГБОУ ВО «Воронежский государственный технический университет»

Ведущая организация: ФГБОУ ВО «Пензенский государственный университет»

Защита состоится 16 октября 2026 г. в 13 часов на заседании диссертационного совета 55.2.002.02 на базе ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ) по адресу: Москва, ул. Авиамоторная 8А.

С диссертацией можно ознакомиться в библиотеке и на сайте МТУСИ: <https://dis.mtuci.ru/dis/02/>

Автореферат разослан _____.

Ученый секретарь
диссертационного совета 55.2.002.02
(приказ ректора МТУСИ от 01.06.2026
г., №158-о)
доктор физико-математических наук



С.Ю. Казанцев

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Развитие городских агломераций, рост уровня автомобилизации, повышение интенсивности перевозок и усложнение структуры транспортных сетей приводят к существенному увеличению требований к эффективности управления дорожным движением. Транспортная инфраструктура функционирует в условиях высокой динамичности, пространственной неоднородности и воздействия множества внешних факторов, включая изменения интенсивности движения, погодные условия, инциденты и особенности поведения участников дорожного движения. В этих условиях традиционные подходы к управлению транспортными потоками, основанные преимущественно на статических схемах регулирования и локальной автоматизации, оказываются недостаточными для обеспечения требуемого уровня пропускной способности, безопасности и устойчивости транспортной системы. Современный этап развития интеллектуальных транспортных систем характеризуется переходом от отдельных автоматизированных решений к интегрированным информационным системам, использующим технологии больших данных, машинного обучения, распределённых вычислений и искусственного интеллекта. Интеллектуальные транспортные системы рассматриваются как сложные социотехнические комплексы, объединяющие средства мониторинга, коммуникационные сети, платформы обработки данных и механизмы автоматизированного управления.

Методы прогнозирования транспортного потока играют важную роль в современных интеллектуальных транспортных системах. На макроуровне они используются для оценки параметров транспортного потока, включая интенсивность движения, скорость и загруженность дорожной сети, тем самым обеспечивая основу для адаптивного управления, навигационных услуг и систем информирования пользователей. На уровне микромоделирования основная задача состоит в описании поведения отдельных транспортных средств и оценке характера их взаимодействия в изменяющейся дорожной обстановке. Ключевой характеристикой транспортных процессов является их нестационарный характер, который формируется динамикой дорожной ситуации, человеческим фактором и изменчивостью статистических свойств наблюдаемых данных.

Несмотря на значительный объем исследований в области интеллектуальных транспортных систем, существующие методы и архитектурные решения имеют ряд ограничений. Пространственно-временные модели прогнозирования часто демонстрируют снижение эффективности при изменении структуры транспортной сети или статистических характеристик входных данных. Традиционные микроскопические модели требуют трудоемкой калибровки и обладают ограниченной адаптивностью к изменяющимся условиям движения, тогда как модели, основанные исключительно

на данных, могут демонстрировать накопление ошибок и аномальное поведение при многократном применении. В то же время практическая реализация интеллектуальных методов осложняется отсутствием формализации в архитектурах выполнения моделей искусственного интеллекта, а также проблемами, связанными с масштабируемостью, отказоустойчивостью и соблюдением требований к качеству обслуживания в условиях нестационарной нагрузки.

Дополнительную актуальность приобретают задачи построения распределенной инфраструктуры интеллектуальных транспортных систем, способной обеспечивать надежный сбор, хранение и обработку больших объемов разнородной информации, а также воспроизводимую верификацию алгоритмов и устойчивое функционирование вычислительных контуров.

Таким образом, актуальность темы исследования определяется необходимостью разработки комплексной методологии создания информационной системы управления транспортными потоками, объединяющей методы прогнозирования макроскопических и микроскопических характеристик транспортного потока, формализованные подходы к построению архитектуры исполнения моделей искусственного интеллекта и методологию построения инфраструктуры интеллектуальных транспортных систем. Решение данных задач имеет существенное научное и практическое значение для повышения эффективности, устойчивости и масштабируемости современных интеллектуальных транспортных систем.

Степень разработанности темы исследования. Степень разработанности темы определяется широким кругом исследований, соответствующих направлениям, в части разработки компьютерных методов и моделей описания, оценки и оптимизации информационных процессов и ресурсов, технологий извлечения и анализа информации в больших базах данных, а также архитектур программно-аппаратных комплексов и методов обеспечения надёжной обработки информации в распределённых интеллектуальных транспортных системах. Существенный массив работ посвящён пространственно-временному анализу и прогнозированию транспортных характеристик на основе данных сенсорных сетей и моделей машинного обучения, включая графовые и нейросетевые подходы, развиваемые в исследованиях Zheyi Wu, Junbo Li, Bin Yu, Chaoqun Chen, Jie Zhou, Jun Zhang, Thomas N. Kipf, William L. Hamilton, Petar Veličković, Max Welling, Yoshua Bengio, Geoffrey Hinton, Yann LeCun.

Значительная часть публикаций относится к микроскопическому моделированию транспортной динамики и моделям следования за лидером, где сформированы классические параметризованные модели и методы их калибровки, а также рассматривается влияние нестационарности условий движения и вариативности поведения водителя. В данном направлении развиваются исследования Denos C. Gazis, Robert Herman, Richard W.

Rothery, Peter G. Gipps, Martin Treiber, Arne Kesting, Dirk Helbing, Masao Bando, Rui Jiang, Carlos Laval, Vincenzo Punzo.

Отдельное направление составляют методы, ориентированные на устойчивость моделей, в том числе при последовательном применении, и на формализуемый контроль качества результатов прогнозирования и управления, включая подходы обучения с подкреплением и методы регуляризации поведения политик. Данные вопросы отражены в работах Timothy P. Lillicrap, Scott Fujimoto, Tuomas Haarnoja.

В части построения и эксплуатации информационных систем обработки транспортных данных развиваются исследования распределённых информационных систем, алгоритмов согласования, обеспечения отказоустойчивости, масштабируемости и соблюдения требований к качеству сервиса. Существенный вклад в соответствующие области внесён Leslie Lamport, Diego Ongaro, John Ousterhout, Jeffrey Dean, Sanjay Ghemawat, Armando Fox.

В отечественной научной школе вопросы интеллектуальных транспортных систем, транспортной телематики и архитектурной организации ИТС рассматриваются в работах В.В. Комарова, С.А. Гарагана, И.А. Евстигнеева, С.В. Жанказиева, Р.В. Душкина и М.В. Яшиной, посвященных вопросам стандартизации, проектирования, внедрения и развития интеллектуальных транспортных систем, цифровой транспортной инфраструктуры и информационно-управляющих комплексов транспортной отрасли.

Отдельное направление отечественных исследований связано с моделированием транспортных потоков, транспортной динамики и разработкой методов управления движением. Существенный вклад в развитие данного направления внесли А.П. Буслаев, А.Г. Таташев и А.В. Гасников, чьи работы посвящены математическому моделированию транспортных процессов, анализу транспортных сетей, вопросам транспортного равновесия, оптимизации потоков и развитию методов управления транспортными системами в условиях высокой динамичности и неопределенности среды.

Методологические и вычислительные основы моделирования сложных динамических процессов, обработки данных и построения высокопроизводительных распределенных вычислительных систем получили развитие в работах А.С. Бугаева, А.М. Валуева и Б.Н. Четверушкина. Их исследования в области прикладной математики, вычислительной физики, численных методов и высокопроизводительных вычислений сформировали научную основу для разработки и анализа сложных информационно-вычислительных систем, включая системы обработки и моделирования транспортных процессов.

Анализ и обобщение работ по теме показывают, что при высокой степени проработанности отдельных компонентов, включая модели пространственно-временного прогнозирования, классические и модели микродинамики основанные на данных, а также методы построения

распределённых интеллектуальных транспортных систем, сохраняется методологический разрыв между задачами извлечения закономерностей из транспортных данных, устойчивым моделированием поведения в условиях нестационарности и формализуемым обеспечением требований к надёжности и качеству сервиса при внедрении и эксплуатации соответствующих информационных процессов и ресурсов.

Научная проблема исследования заключается в отсутствии комплексных методов и методологий оценки и оптимизации информационных процессов и ресурсов интеллектуальных транспортных систем, обеспечивающих извлечение закономерностей и прогнозирование по пространственно-временным данным при нестационарности. Существенными проявлениями проблемы являются недостаточная адаптивность параметризованных моделей к изменению условий, возникновение аномальных предсказаний у моделей, основанных на данных, при последовательном применении, ограниченная пригодность доступных информационных ресурсов в виде наборов данных для воспроизводимой оценки и обучения, а также недостаточная формализация архитектур и процедур верификации, обеспечивающих устойчивость обработки информации в условиях нестационарной нагрузки.

Целью диссертационной работы является повышение эффективности функционирования интеллектуальных транспортных систем за счёт разработки методов прогнозирования макроскопических и микроскопических характеристик транспортного потока, формирования методологии построения и верификации архитектуры информационной системы сбора, хранения и обработки данных о состоянии среды.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Провести анализ современного состояния интеллектуальных транспортных систем.
2. Разработать метод прогнозирования макроскопических характеристик транспортного потока. Экспериментально оценить предложенный метод на общепринятых наборах данных пространственно-временного прогнозирования транспортных потоков и сравнить его с базовыми методами.
3. Разработать метод прогнозирования микроскопических характеристик транспортного потока, позволяющий учитывать динамически изменяемый контекст сцены. Провести экспериментальное исследование предложенного метода на общепринятых наборах данных, оценив устойчивость при последовательном применении и чувствительность к накоплению ошибок.
4. Разработать и формализовать методологию конфигурирования архитектуры исполнения методов искусственного интеллекта в интеллектуальных транспортных системах, ориентированную на

- соблюдение требований уровня сервиса по задержке, пропускной способности и доступности в условиях нестационарной нагрузки.
5. Разработать архитектуру системы обработки больших данных для интеллектуальных транспортных систем структурированную по функциональным зонам.
 6. Разработать и обосновать методологию построения инфраструктуры ИТС, ориентированную на отказоустойчивость, масштабируемость и адаптируемость сбора, хранения и обработки больших объёмов разнородных данных.

Объектом исследования являются процессы функционирования интеллектуальных транспортных систем, включающие сбор и обработку разнородных данных о дорожной обстановке, а также методы моделирования и прогнозирования характеристик транспортного потока.

Предметом исследования является научно-методический комплекс методов прогнозирования макроскопических и микроскопических характеристик транспортного потока, методология построения, верификации и выбора архитектур исполнения моделей искусственного интеллекта и инфраструктурных решений интеллектуальных транспортных систем при ограничениях по качеству сервиса, масштабируемости и отказоустойчивости.

Научная новизна результатов диссертации. В процессе выполнения диссертационной работы получены следующие оригинальные научные результаты:

1. Разработан метод пространственно-временного прогнозирования макроскопических характеристик транспортного потока PTSTG, *отличающийся от существующих* синтезом патч-ориентированного временного кодирования с адаптивной стохастической по строкам матрицей смежности, обучаемой из данных без использования фиксированной дорожной топологии, и вычислительно простым графовым уточнением узловых представлений.
2. Разработан метод динамической коррекции коэффициентов классической модели следования за лидером, *отличающийся от существующих*, динамической адаптацией параметров во времени на основе текущего состояния движения при сохранении физической интерпретируемости параметров модели.
3. Разработан гибридный метод прогнозирования микроскопических характеристик транспортного потока, основанный на данных, *отличающийся от существующих* реализацией функционального разделения модулей, при котором блок, основанный на данных, обеспечивает точность воспроизведения динамики (краткосрочное и долгосрочное окно), а классический блок задаёт формализуемые ограничения и подавляет аномальные воздействия в замкнутом контуре.

4. Разработана комплексная методология построения архитектуры исполнения моделей искусственного интеллекта в интеллектуальных транспортных системах, *отличающаяся от существующих* формализацией архитектуры в виде ориентированного мультиграфа исполнения с параметризованными вершинами и рёбрами, введением вероятностных ограничений по задержке и устойчивости по интенсивности входа, и обеспечением воспроизводимого синтеза и выбора конфигурации исполнения на основе измеримых метрик.
5. Разработана архитектура системы сбора, хранения и обработки больших разнородных данных для интеллектуальных транспортных систем, *отличающаяся от существующих* структурированием по функциональным зонам, охватывающая весь процесс функционирования системы начиная со сбора данных до аналитики, хранения и визуализации.
6. Разработана методология построения инфраструктуры интеллектуальных транспортных систем, *отличающаяся от существующих* тем, что объединяет архитектуру обработки больших данных с механизмом децентрализованной кластеризации серверов на основе метрики сетевого расстояния (RTT), консенсуса Raft и автоматического выбора лидера сервисом Agent, а также специализированным алгоритмом динамического распределения приоритетных задач при ограничении по времени выполнения и методом генерации статистически обоснованных синтетических данных транспортного потока.

Теоретическая и практическая значимость. Теоретическая значимость диссертационного исследования заключается в развитии методологических основ создания системы управления транспортными потоками на базе интеллектуальной обработки пространственно-временных данных и адаптивного моделирования транспортных процессов. В работе:

- обоснована реализация замкнутого контура управления транспортными потоками с систематизацией сценариев взаимодействия подсистем мониторинга, прогнозирования и формирования управляющих воздействий в составе интеллектуальной транспортной системы;
- установлена связь между выбором алгоритмических и технологических решений (прогнозирование макро- и микроскопических характеристик, адаптивная коррекция параметров моделей, архитектура исполнения вычислений) и показателями качества функционирования системы управления, включая устойчивость моделирования и точность прогнозирования в условиях нестационарности транспортной нагрузки;

- введено новое представление об архитектуре комплекса обработки транспортных данных как распределённой системы информационного обмена, обоснована структурно-функциональная модель, включающая компоненты сбора, хранения и обработки больших разнородных данных, а также средства гарантированного исполнения моделей искусственного интеллекта по требованиям к задержке, пропускной способности и доступности;
- показана концептуальная возможность интеграции пространственно-временного прогнозирования и гибридного моделирования транспортных процессов с распределённым исполнением моделей искусственного интеллекта в едином вычислительном контуре, обеспечивающем масштабируемость на крупные сенсорные сети и устойчивость к изменениям статистических свойств входных данных.

Практическая значимость работы определяется возможностью применения результатов диссертационного исследования при проектировании, внедрении и эксплуатации интеллектуальных транспортных систем и центров управления дорожным движением. Разработанные методы пространственно-временного прогнозирования и гибридного моделирования, а также предложенные решения по архитектуре сбора, хранения и обработки больших разнородных данных и по организации отказоустойчивого распределённого исполнения моделей искусственного интеллекта могут быть использованы:

- в системах мониторинга и прогнозирования состояния улично-дорожной сети для поддержки оперативного управления дорожным движением (адаптивное регулирование, маршрутизация, информирование участников движения);
- в программно-аппаратных комплексах обработки транспортных данных, обеспечивающих протокольную совместимость, интеграцию с внешними информационными системами и соблюдение требований к качеству обслуживания прикладных сервисов;
- при разработке и испытаниях инфраструктуры геораспределённых ИТС, включая воспроизводимую валидацию моделей при ограниченности натуральных данных.

Методы исследования. Методологическую основу работы составляет системный подход, включающий системный анализ предметной области, декомпозицию задач и формализацию требований к качеству функционирования интеллектуальных транспортных систем. В ходе исследования использовались методы математического моделирования и идентификации параметров для описания микроскопической динамики движения и адаптации параметров классических моделей, методы интеллектуального анализа данных и машинного обучения для построения прогнозных моделей по пространственно-временным данным, а также

методы теории графов и анализа временных рядов для учёта пространственных взаимодействий и временных зависимостей. Для обеспечения обоснованности выводов применялись методы теории вероятностей и математической статистики, включая процедуры сравнительного анализа и оценку качества моделей по формализованным метрикам. Исследование включало вычислительный и имитационный эксперимент, в том числе оценку поведения моделей при последовательном применении, а также экспериментальную верификацию на натурных и общепринятых наборах данных. При решении задач проектирования архитектуры исполнения и инфраструктуры применялись методы многокритериальной оптимизации и верификации ограничений по качеству сервиса.

Основные положения, выносимые на защиту:

1. Метод пространственно-временного прогнозирования макроэкономических характеристик транспортного потока PTSTG, *позволяющий* выполнять однократное многогоризонтное прогнозирование на горизонтах 15/30/60 минут при сопоставимой точности и обеспечивать масштабируемость на крупные сенсорные сети за счёт разрежения межузловых взаимодействий.
2. Метод динамической коррекции коэффициентов классической модели следования за лидером, *позволяющий* обеспечивать учёт нестационарности дорожной среды и повышение устойчивости моделирования в условиях вариативности внешних факторов и режимов движения.
3. Гибридный метод прогнозирования микроэкономических характеристик транспортного потока, основанный на данных, *позволяющий* снижать накопление ошибки при последовательном применении модели и повышает устойчивость при сдвиге распределения при сохранении точности в типичных режимах движения.
4. Комплексная методология построения архитектуры исполнения моделей искусственного интеллекта в интеллектуальных транспортных системах, *позволяющая* гарантировать заданные уровни задержки, пропускной способности и доступности в условиях нестационарной нагрузки ИТС.
5. Архитектура системы сбора, хранения и обработки больших разнородных данных для интеллектуальных транспортных систем, *позволяющая* снизить сложность интеграции с внешними системами за счёт учёта протокольной совместимости, адаптированной к типу источника и сценарию взаимодействия.
6. Методология построения инфраструктуры интеллектуальных транспортных систем, *позволяющая* обеспечивать самонастройку и отказоустойчивость геораспределённой ИТС и выполнять воспроизводимое тестирование при отсутствии реальных данных.

Степень достоверности и апробации результатов Основные результаты диссертационного исследования были представлены на 44 международных конференциях, форумах и научных семинарах:

Международная научно-практическая конференция «Quality Management, Transport and Information Security, Information Technologies» (ITQMIS) (Санкт-Петербург, 2018; Сочи, 2019). Международная научная конференция «Wave electronics and its application in information and telecommunication systems» (WECONF) (Санкт-Петербург, 2018-2024). Международная конференция «Технологии и предпринимательство в цифровом обществе» (TEDS) (Москва, 2019). Международная конференция «Systems of signals generating and processing in the field of on board communications» (Москва, 2019-2021, 2023-2025). Международная конференция «Systems of Signal Synchronization, Generating and Processing in Telecommunications» (SYNCHROINFO) (Ярославль, 2019; Светлогорск, 2020-2021; Архангельск, 2022; Псков, 2023; Выборг, 2024; Тюмень, 2025). International Conference on Engineering Management of Communication and Technology (EMCTECH) (Вена, 2020). Международная конференция «Intelligent technologies and electronic devices in vehicle and road transport complex» (TIRVED) (Москва, 2021, 2023, 2025). Международная отраслевая научно-техническая конференция «Технологии информационного общества» (Москва, 2017-2019). Международная конференция «Транспортные потоки на сетях» (Сириус, 2023, 2025). Международная научно-техническая конференция «Новые информационные технологии в телекоммуникациях и почтовой связи» (Минск, 2021-2022). International Conference Computational and Mathematical Methods in Science and Engineering (Кадис, 2025). Международный промышленный форум «Интеллект машин и механизмов» (Севастополь, 2021). Научный семинар «Фундаментальные и прикладные проблемы развития автомобильно-дорожного комплекса России» академика РАН, д.ф.-м.н., профессора Валерия Васильевича Козлова (Москва, 2018-2025).

Результаты также обсуждались на заседании кафедры, а также на научных семинарах в МТУСИ.

Апробация результатов подтверждается свидетельствами о регистрации программ для ЭВМ, представленных в приложении К диссертации.

Личный вклад.

Все основные научные результаты и положения, выносимые на защиту, получены лично автором или при его определяющем участии. Автору принадлежит постановка научной проблемы, формирование цели и задач исследования, разработка методологических, математических, алгоритмических и архитектурных решений, проведение вычислительных и имитационных экспериментов, анализ и интерпретация полученных результатов, а также подготовка публикаций по теме диссертационного исследования. В работах, опубликованных в соавторстве, соискателю

принадлежит определяющая роль в постановке исследовательских задач, формализации предлагаемых методов, разработке архитектурных и алгоритмических решений, выборе экспериментальных сценариев, анализе результатов и обобщении выводов применительно к методологии создания системы управления транспортными потоками.

В работах [1; 2] соискателю принадлежит разработка общего подхода к пространственно-временному прогнозированию макроскопических характеристик транспортного потока, включая обоснование применения патч-ориентированного временного кодирования, адаптивной обучаемой матрицы смежности и графового уточнения узловых представлений.

В работах [3–14], связанных со сбором, определением и моделированием характеристик транспортного потока, соискателю принадлежит постановка и решение задач прогнозирования микроскопических характеристик транспортного потока.

В работах [4; 6; 11; 13; 15–30] соискателем разработаны и исследованы интеллектуальные и гибридные методы обработки транспортных данных, распознавания объектов и характеристик транспортной сцены, анализа устойчивости нейросетевых решений и прогнозирования динамики движения.

В работах [31–48] соискателю принадлежит разработка и обоснование архитектурных решений для сбора, хранения, обработки и защищенного информационного обмена в интеллектуальных транспортных системах.

В работах [31–36; 41–45; 47; 48], связанных с поддержкой функционирования ИТС, обработкой и хранением транспортных данных, соискателем разработана архитектура системы сбора, хранения и обработки больших разнородных данных для интеллектуальных транспортных систем.

В работах [5; 10; 29; 30; 37–40; 46–59], связанных с генерацией синтетических данных, динамическим распределением ресурсов, поддержкой функционирования ИТС, передачей информации между транспортными средствами и адаптивным управлением светофорами, соискателем разработана методология построения инфраструктуры интеллектуальных транспортных систем.

В публикациях [5; 7; 52; 54–60] соискателем дополнительно обоснованы подходы к применению обучения с подкреплением, имитационного моделирования и цифрового представления дорожной среды для задач управления транспортными потоками, выбора траекторий движения и испытания алгоритмов ИТС.

В публикациях [11; 14; 18; 19; 21; 24–29; 61–63] автором исследованы методы сбора и прогнозирования микро и макро-характеристик транспортного потока поступающих из различных источников информации о дорожной ситуации. Эти результаты использованы при формировании подсистем мониторинга и семантического восприятия в составе интеллектуальной транспортной системы.

В публикациях [41–48] соискателем проработаны вопросы разработки информационных систем, обработки данных и применения распределённых технологий в задачах транспортной и цифровой инфраструктуры. Данные результаты использованы при обосновании архитектурных решений, обеспечивающих интероперабельность, надёжность и воспроизводимость информационных процессов ИТС.

Автором лично выполнены систематизация и обобщение результатов исследования, сформулированы выводы по главам и заключение диссертации, определены теоретическая и практическая значимость работы, подготовлены материалы апробации и публикации по теме диссертационного исследования. Совокупность публикаций [1–63] подтверждает полноту апробации основных научных результатов диссертации и их соответствие положениям, выносимым на защиту.

Благодарности. Особую благодарность автор выражает доктору физико-математических наук, профессору Буслаеву Александру Павловичу, научному руководителю кандидатской диссертации, чьи научные идеи, профессиональное наставничество и личный пример оказали значительное влияние на формирование научных взглядов автора и во многом определили направление настоящего исследования.

Реализация и внедрение. Алгоритмы, методы и методология, разработанные в настоящем исследовании, внедрены в следующих организациях:

- «Минимакс-94» (АО) (г. Москва): внедрение указанных результатов позволило повысить точность прогнозирования транспортной ситуации, улучшить качество управления транспортными потоками и повысить оперативность реагирования на изменения дорожной обстановки;
- «Эр Си Технолджис» (ООО) (г. Москва): внедрение указанных результатов позволило повысить надёжность функционирования информационных систем, обеспечить масштабируемость обработки данных и соблюдение требований к качеству сервиса;
- «Транснависофт» (ООО) (г. Москва): внедрение указанных результатов позволило повысить устойчивость моделирования транспортных процессов, обеспечить воспроизводимость оценки качества работы интеллектуальных транспортных систем и повысить эффективность анализа транспортных данных;
- учебный процесс кафедры «Математическая кибернетика и информационные технологии» ордена Трудового Красного Знамени федерального государственного бюджетного образовательного учреждения высшего образования «Московский технический университет связи и информатики» (МТУСИ);

- учебный процесс кафедры «Инженерия и математика прикладных систем искусственного интеллекта» федерального государственного бюджетного образовательного учреждения высшего образования «Московский автомобильно-дорожный государственный технический университет (МАДИ)».

Подтверждается соответствующими актами внедрения результатов диссертационной работы (5 шт.), представленных в приложении Л диссертации.

Соответствие специальности. Тематика и результаты диссертации соответствуют следующим направлениям специальности: 2.3.8. Информатика и информационные процессы.

1. п. 8 - разработка систем принятия решений на основе баз данных и знаний, реализующих имитационные модели прогнозирования изменения материальных процессов и событий, поскольку разработаны подходы к прогнозированию и устойчивому моделированию транспортного движения, включая адаптивную коррекцию параметров микроскопических моделей следования и оценку поведения потока в меняющихся режимах;
2. п. 9 - разработка архитектур программно-аппаратных комплексов поддержки цифровых технологий сбора, хранения и передачи информации в инфокоммуникационных системах, в том числе с использованием «облачных» интернет-технологий, и оценка их эффективности, поскольку обоснована архитектура исполнения моделей искусственного интеллекта в составе ИТС с требованиями к задержке, пропускной способности и доступности, а также предложены принципы организации вычислительной инфраструктуры;
3. п. 13 - разработка и применение методов распознавания образов, кластерного анализа, нейросетевых и нечетких технологий, решающих правил, мягких вычислений при анализе разнородной информации в базах данных, поскольку предложены и экспериментально исследованы интеллектуальные (в том числе гибридные) методы прогнозирования макро- и микроскопических характеристик транспортного потока, ориентированные на устойчивость к нестационарности и дрейфу данных;
4. п. 14 - разработка и исследование принципов организации и функционирования распределенных информационных систем и баз данных, прикладных протоколов информационных сетей, форматов представления данных и языков информационного поиска в распределенных информационных ресурсах, поскольку разработана архитектура системы сбора, хранения и обработки больших разнородных транспортных данных с учетом интероперабельности компонентов и интеграции с внешними информационными системами;

Публикации. По материалам диссертации опубликовано 63 научные работы (в том числе 16 без соавторов), из которых: публикации в рецензируемых периодических изданиях, входящих в перечень ВАК – 10; в научных изданиях, индексируемых в международных базах Scopus/WoS – 38 (из которых: Q1 – 6, Q2 – 2, которые в соответствии с Рекомендацией ВАК от 21.12.2023, приравниваются к публикациям в рецензируемых изданиях категории K1); свидетельства о государственной регистрации программ для ЭВМ – 15 (которые приравниваются к публикациям в рецензируемых изданиях, рекомендованных ВАК).

Содержание работы

Во **введении** обосновывается актуальность темы диссертационного исследования, обусловленная необходимостью повышения эффективности функционирования ИТС за счёт разработки методов прогнозирования макро- и микроскопических характеристик транспортного потока, а также методологии построения, конфигурирования и верификации информационной системы сбора, хранения, обработки и анализа транспортных данных. Показано, что рост интенсивности транспортных потоков, увеличение объёма разнородных данных, развитие сенсорной инфраструктуры, кооперативных транспортных коммуникаций и методов ИИ требуют перехода от отдельных автоматизированных подсистем к комплексной интеллектуальной информационно-управляющей системе, обеспечивающей прогнозирование состояния транспортной сети, поддержку принятия решений и устойчивое функционирование в условиях нестационарной нагрузки.

В **первой главе** выполнен анализ современного состояния ИТС как распределённых социотехнических систем, объединяющих дорожную инфраструктуру, сенсорные подсистемы, средства связи, платформы хранения и обработки данных, методы ИИ и механизмы автоматизированного управления. Показано, что современная ИТС не сводится к автоматизации отдельных технических средств регулирования дорожного движения, а представляет собой многоуровневую информационно-управляющую систему, функционирующую в условиях нестационарности транспортных потоков, ограниченной наблюдаемости, нормативных требований и жёстких эксплуатационных ограничений по задержке, доступности и надёжности.

В главе обосновано, что методологической основой функционирования ИТС является замкнутый управленческий контур, в котором данные транспортного мониторинга преобразуются в прогнозное описание состояния сети и далее используются для формирования управляющих воздействий. Такая постановка позволяет связать задачи сбора и обработки данных, прогнозирования транспортных характеристик, исполнения моделей ИИ и обеспечения качества сервиса распределённой информационной системы. Архитектура ИТС при этом рассматривается как

согласованная совокупность полевого, коммуникационного, аналитического, прикладного и пользовательского уровней, дополненная средствами обеспечения безопасности, контроля качества данных, мониторинга, журналирования и управления жизненным циклом моделей.

Проведённый анализ показал, что развитие ИТС связано с переходом от локальных автоматизированных подсистем к интегрированным платформам, использующим большие транспортные данные, потоковую обработку, кооперативные коммуникации и методы ИИ. Вместе с тем сохраняется методологический разрыв между извлечением транспортных данных, построением прогнозных моделей, безопасным исполнением моделей ИИ и формализованным обеспечением требований к качеству сервиса. Данный разрыв проявляется в ограничениях существующих методов макроэкономического прогнозирования, недостаточной адаптивности микроэкономических моделей, отсутствии проверяемой методологии выбора архитектуры исполнения моделей, неполной формализации инфраструктуры обработки разнородных данных и затруднённой проверке решений при ограниченной доступности реальных транспортных наблюдений.

Таким образом, в первой главе обоснована необходимость разработки совокупности взаимосвязанных методов и методологий: метода прогнозирования макроэкономических характеристик транспортного потока, методов микроэкономического прогнозирования и коррекции моделей следования за лидером, методологии конфигурирования архитектуры исполнения моделей ИИ, а также методологии построения инфраструктуры сбора, хранения, обработки и тестирования данных ИТС. Полученные в главе выводы определяют постановку задач последующих глав диссертационной работы и обеспечивают связь между заявленной научной новизной, положениями, выносимыми на защиту, и практической направленностью исследования.

Во второй главе решается задача прогнозирования макроэкономических характеристик транспортного потока по данным сенсорной сети. Данная задача является одной из ключевых для ИТС, поскольку прогноз скорости, интенсивности и состояния дорожных участков используется при оценке времени в пути, обнаружении перегрузок, поддержке навигационных сервисов и выработке управляющих воздействий. В отличие от локальной обработки одиночного временного ряда, транспортная сеть характеризуется одновременным наличием временной динамики и пространственных зависимостей между дорожными участками. Поэтому модель прогнозирования должна учитывать как долговременные закономерности в ряду наблюдений, так и функциональные связи между узлами сенсорной сети.

Проведённый анализ существующих методов показал, что рекуррентные архитектуры LSTM и GRU позволяют учитывать предыдущее состояние временного ряда, но ограничены последовательной обработкой

и ухудшением передачи дальнего контекста при увеличении горизонта прогноза. Графовые пространственно-временные модели, такие как STGCN, DCRNN и Graph WaveNet, включают структуру дорожной сети, однако часто используют заранее заданную или слабо адаптивную матрицу смежности. Трансформерные модели и патч-ориентированные архитектуры лучше работают с длинным временным контекстом, но в базовой форме не задают явного механизма пространственного взаимодействия между сенсорами. Следовательно, для задач ИТС требуется метод, сочетающий эффективность временного кодирования, адаптивное описание межузловых связей и пригодность к развёртыванию в вычислительной инфраструктуре с ограничениями по задержке и пропускной способности.

Для решения указанной задачи в диссертационной работе разработан метод **PTSTG** (*PatchTST-Graph*), предназначенный для пространственно-временного прогнозирования макроскопических характеристик транспортного потока. Метод объединяет патч-ориентированное временное кодирование, обучаемую стохастическую матрицу смежности и лёгкое графовое уточнение узловых представлений. Временная часть метода основана на идее PatchTST: временной ряд каждого сенсора разбивается на перекрывающиеся патчи, что уменьшает длину последовательности, обрабатываемой механизмом внимания, и позволяет устойчиво учитывать долгосрочный контекст. Пространственная часть метода задаётся не фиксированной дорожной топологией, а матрицей межузловых связей, обучаемой по данным.

Постановка задачи формализуется следующим образом. Пусть заданы исторические наблюдения сенсорной сети:

$$X \in \mathbb{R}^{B \times T_{\text{in}} \times N \times C_{\text{in}}}, \quad (1)$$

и целевые значения:

$$Y \in \mathbb{R}^{B \times H \times N \times C_{\text{out}}}, \quad (2)$$

где B - размер батча, T_{in} - длина входного окна, N - число узлов сенсорной сети, C_{in} и C_{out} - размерности входных и выходных признаков, H - горизонт прогнозирования. Требуется построить параметризованное отображение

$$\hat{Y} = \Phi_{\theta}(X), \quad (3)$$

которое по фиксированному историческому окну формирует прогноз на весь горизонт за один прямой проход. В экспериментах использовался шаг дискретизации 5 минут, поэтому значения $H \in \{3, 6, 12\}$ соответствуют горизонтам 15, 30 и 60 минут.

Временной кодировщик PTSTG выполняет разбиение ряда каждого узла на патчи длины P с шагом s . Для узла i последовательность $x_i \in \mathbb{R}^{T_{\text{in}} \times C_{\text{in}}}$ преобразуется в набор патчей:

$$U_i = \mathcal{P}(x_i; P, s) \in \mathbb{R}^{L \times (PC_{\text{in}})}, \quad (4)$$

где

$$L = \left\lfloor \frac{T_{\text{in}} - P}{s} \right\rfloor + 1 \quad (5)$$

- число патчей. Каждый патч проецируется в пространство признаков размерности d , после чего обрабатывается трансформерным кодировщиком:

$$E_i = U_i W_p + b_p, \quad Z_i = \text{TransformerEncoder}(E_i + \text{PE}). \quad (6)$$

Итоговое представление узла формируется агрегацией по патч-токенам:

$$h_i = \frac{1}{L} \sum_{\ell=1}^L Z_{i\ell}. \quad (7)$$

Такое представление сохраняет локальную динамику внутри патчей и одновременно содержит информацию о более длинном временном контексте.

Пространственный блок PTSTG строится на обучаемой стохастической по строкам матрице смежности. Для её формирования используются две матрицы узловых вложений $E_1, E_2 \in \mathbb{R}^{N \times r}$. Сначала вычисляется неотрицательная матрица аффинностей:

$$S = \text{ReLU}(E_1 E_2^\top), \quad (8)$$

затем выполняется симметризация:

$$\tilde{S} = \frac{1}{2}(S + S^\top), \quad (9)$$

после чего матрица смежности получается строковой нормализацией:

$$A = \text{softmax}_{\text{row}} \left(\frac{\tilde{S}}{\tau} \right), \quad A \mathbf{1} = \mathbf{1}. \quad (10)$$

Каждая строка A задаёт распределение весов по узлам, участвующим в смешивании признаков. В отличие от подходов, где граф задаётся только физической близостью дорожных участков, в PTSTG связи между сенсорами определяются по обучаемым латентным признакам. Это позволяет учитывать функциональные зависимости между участками сети, возникающие вследствие маршрутизации, координации светофорного управления, перераспределения спроса и повторяющихся режимов перегрузки.

После построения матрицы A выполняется графовое уточнение узловых представлений. Пусть

$$H^{(0)} = \text{Stack}(h_1, \dots, h_N) \quad (11)$$

- начальное пространственное представление всех узлов. На каждом графовом слое выполняется межузловое смешивание вида:

$$\hat{H} = H^{(\ell)} + (A H^{(\ell)}) W_\ell, \quad (12)$$

после чего применяется нормализация и позиционный полносвязный блок. Остаточные соединения ограничивают деградацию признаков при увеличении глубины графового уточнения. При необходимости матрица A может быть разрежена по принципу $\text{top-}k$: в каждой строке сохраняются только k наибольших значений с последующей повторной нормализацией. Это снижает вычислительную стоимость межзвучного смешивания и делает метод пригодным для крупных сенсорных сетей.

Ключевое отличие PTSTG от базового PatchTST состоит в том, что PatchTST обрабатывает временные ряды как независимые или слабо связанные каналы, тогда как PTSTG после временного кодировщика вводит явный механизм пространственного взаимодействия. В отличие от графовых моделей с фиксированной матрицей дорожной топологии, предложенный метод обучает матрицу связей по данным. В отличие от тяжёлых пространственно-временных механизмов внимания, PTSTG сохраняет разделение временного и пространственного блоков, что упрощает вычисления и делает модель более удобной для последующего развёртывания в ИТС.

Результаты экспериментов приведены для трёх наборов данных и трёх горизонтов прогноза. Основными метриками являются MAE и RMSE. MAPE используется как дополнительный показатель, поскольку его значение зависит от способа обработки малых скоростей и выбора нижнего ограничения знаменателя.

На наборе METR-LA метод PTSTG показывает качество, близкое к сильным графовым моделям на коротком и среднем горизонтах. На горизонте 15 минут PTSTG достигает MAE 2,85, RMSE 5,33 и MAPE 7,7%. При этом значение RMSE оказывается ниже, чем у DCRNN, для которой RMSE составляет 5,38, однако по MAE и MAPE лучшим остаётся DCRNN со значениями 2,77 и 7,3%. Это показывает, что предложенный метод точнее ограничивает крупные ошибки прогноза, но несколько уступает по средней абсолютной ошибке.

На горизонте 30 минут PTSTG имеет MAE 3,39, RMSE 6,41 и MAPE 9,7%. По RMSE метод остаётся сопоставимым с DCRNN, у которой RMSE составляет 6,45, но уступает ей по MAE и MAPE. На горизонте 60 минут PTSTG показывает MAE 4,19, RMSE 7,69 и MAPE 12,67%, тогда как DCRNN достигает MAE 3,60, RMSE 7,59 и MAPE 10,5%. Следовательно, на METR-LA базовая конфигурация PTSTG наиболее эффективна на коротком и среднем горизонтах, а на дальнем горизонте требует усиления пространственного блока или дополнительной настройки графовой части.

На наборе PEMS-BAY картина отличается. На горизонте 15 минут PTSTG достигает MAE 1,55, RMSE 3,12 и MAPE 3,43%. Эти значения сохраняют сопоставимый порядок ошибки относительно DCRNN, для которой получены MAE 1,38, RMSE 2,95 и MAPE 2,9%. Следовательно, на

краткосрочном прогнозе предложенный метод близок к сильной графовой модели, хотя и не превосходит её.

На горизонте 30 минут PTSTG показывает MAE 2,08, RMSE 4,28 и MAPE 4,84%, тогда как DCRNN имеет MAE 1,74, RMSE 3,97 и MAPE 3,90%. На горизонте 60 минут разрыв увеличивается: PTSTG достигает MAE 2,81, RMSE 5,54 и MAPE 6,81%, а DCRNN показывает MAE 2,07, RMSE 4,74 и MAPE 4,90%. Вероятная причина состоит в том, что в базовой версии PTSTG используется статическая обучаемая матрица смежности. Она отражает усреднённые функциональные связи между узлами, но не перестраивается для каждого входного окна. Поэтому при изменении режима движения, особенно на дальнем горизонте, ошибка возрастает быстрее.

На LargeST метод PTSTG показывает наиболее сильный результат относительно рассматриваемых базовых моделей. На горизонте 3 шага PTSTG достигает MAE 16,42, RMSE 26,42 и MAPE 11,04%. Эти значения уступают лучшей модели D²STGNN, для которой получены MAE 14,92, RMSE 24,95 и MAPE 9,56%, однако остаются на уровне конкурентоспособных графовых моделей.

На горизонте 6 шагов PTSTG показывает MAE 18,86, RMSE 30,37 и MAPE 12,93%. На горизонте 12 шагов метод достигает MAE 22,69, RMSE 36,42 и MAPE 16,59%. При этом по RMSE на дальнем горизонте PTSTG оказывается лучшим среди рассмотренных моделей, что указывает на способность метода ограничивать крупные ошибки при увеличении горизонта прогноза.

В среднем по трём горизонтам PTSTG имеет MAE 17,50, RMSE 28,18 и MAPE 11,07%, что является лучшим средним результатом среди рассмотренных моделей по всем трём метрикам. При этом число параметров PTSTG составляет 0,88 млн. Это больше, чем у DCRNN, GWNET, DGCRN и D²STGNN, но меньше, чем у ASTGCN и DSTAGNN. Для данной работы этот результат является существенным, поскольку LargeST лучше отражает задачу масштабирования на крупную сенсорную сеть. Преимущество PTSTG связано с тем, что временной кодировщик работает по патчам, а пространственное смешивание может быть ограничено разреженной матрицей смежности.

Отдельно оценены вычислительные характеристики метода. В базовой конфигурации PTSTG содержит 0,88 млн параметров. На потребительском GPU NVIDIA GeForce RTX 4060 Ti при батче $B = 32$, длине входного окна 36 шагов и горизонте прогноза 12 шагов средняя задержка прямого прохода составила 56,44 мс, задержка p95 - 61,54 мс, пропускная способность - 564,04 последовательности в секунду. Эти значения показывают, что предложенный метод ориентирован не только на повышение точности прогноза, но и на последующее использование в составе вычислительного контура ИТС.

Таким образом, во второй главе разработан метод прогнозирования макроскопических характеристик транспортного потока, отличающийся сочетанием патч-ориентированного временного кодирования, обучаемой стохастической матрицы смежности и лёгкого графового уточнения. Метод формирует многогоризонтный прогноз за один прямой проход, не требует полного задания дорожной топологии и допускает снижение вычислительной сложности за счёт разрежения межузловых связей. Полученные экспериментальные результаты подтверждают, что PTSTG обеспечивает рациональный компромисс между точностью, числом параметров, вычислительной простотой и применимостью к крупным сенсорным сетям ИТС.

В третьей главе решается задача прогнозирования микроскопических характеристик транспортного потока на основе моделей следования за лидером. В отличие от макроскопического уровня, где анализируются агрегированные характеристики дорожных участков, микроскопический уровень описывает продольное взаимодействие отдельных транспортных средств и реакцию ведомого автомобиля на изменение скорости, дистанции и режима движения лидера. Качество такого прогноза определяет достоверность моделирования заторов, оценки устойчивости потока, анализа опасных сближений и проверки алгоритмов управления в замкнутом контуре.

Показано, что классические модели следования за лидером обладают физической интерпретируемостью, но ограничены использованием фиксированных параметров. Модели, основанные на данных, включая LSTM, CNN-LSTM, TCN и Informer, точнее восстанавливают нелинейную динамику по траекторным данным, но не гарантируют безопасное поведение при последовательном развёртывании. Поэтому в работе используется гибридная постановка, в которой нейросетевой компонент обеспечивает точность прогноза, а классический компонент выполняет физически интерпретируемую коррекцию небезопасных и аномальных выходов.

Первым этапом стала разработка метода сбора микроскопических данных транспортного потока. Необходимость формирования собственного корпуса обусловлена тем, что открытые наборы данных, включая NGSIM, highD и pNEUMA, не полностью отражают целевые условия эксплуатации, связанные с сезонной изменчивостью, зимними сценариями, ухудшенной видимостью, локальной городской инфраструктурой и поведенческой неоднородностью водителей.

Предложенный метод ориентирован на получение синхронизированных наблюдений пары лидера и ведомого. В качестве источников используются мобильное клиент-серверное приложение, координаты, скорость, направление движения, временные метки, характеристики водителя и автомобиля, а также сопоставленные с поездкой метеорологические признаки. На серверной стороне выполняются фильтрация выбросов, исправление единичных артефактов, синхронизация временных рядов,

определение совместного движения транспортных средств, выделение статуса лидера и ведомого, расчёт межмашинной дистанции и формирование обучающих последовательностей.

Сформированный корпус включает данные 37 водителей и 21 автомобиля, более 1800 км поездок за три года наблюдений, секундную дискретизацию, городские и автомагистральные сценарии, сезонное покрытие и погодные классы, включая дождь, снег, туман, пониженную температуру и условия потенциального гололёда. После фильтрации и синхронизации из 32 632 исходных записей получено 23 175 записей, пригодных для обучения и проверки моделей. При использовании окна длины $seq_len = 10$ каждая непрерывная сессия длиной T_j формирует $N_j^{seq} = \max(0, T_j - seq_len)$ обучающих последовательностей, а общий объём выборки определяется как $N^{seq} = \sum_{j=1}^J \max(0, T_j - seq_len)$, где J является числом непрерывных сегментов после предварительной обработки.

Далее в главе разработан метод динамической коррекции коэффициентов классической модели следования за лидером SSSR-Micro. В качестве базового выражения используется модель динамического габарита:

$$s_{req} = C_0 + C_1 v + C_2 v^2,$$

где s_{req} - требуемая безопасная дистанция, v - скорость ведомого транспортного средства, C_0 , C_1 , C_2 - коэффициенты модели. В классической постановке эти коэффициенты задаются статически или калибруются для ограниченного набора условий. В диссертационной работе предложено рассматривать их как динамически изменяемые параметры, зависящие от текущего состояния движения и внешнего контекста.

Пусть в дискретные моменты времени t доступны наблюдения пары лидер-ведомый

$$S_t = (s_t, v_t^f, v_t^\ell, \Delta v_t), \quad \Delta v_t = v_t^f - v_t^\ell,$$

а также контекстный вектор $\mathbf{z}_t$, включающий по крайней мере тип дорожного сценария, сезон и погодный класс, а при наличии также температурные и осадочные показатели и атрибуты водителя и автомобиля. Предполагается, что Δt является шагом дискретизации, согласованным с частотой сбора данных. Предлагаемый метод динамической коррекции коэффициентов модели включает следующие шаги.

Шаг 1. Контекстное разделение данных. Корпус наблюдений разбивается на набор контекстных сегментов $\{\mathcal{D}_k\}_{k=1}^K$, где сегмент задаётся сочетанием факторов (*road_type*, *season*, *weather*) с возможным укрупнением классов при недостатке статистики. Разделение необходимо для устойчивой оценки коэффициентов в условиях неоднородности данных.

Шаг 2. Получение опорной оценки дистанции. Для построения устойчивой опорной зависимости «скорость-дистанция» в каждом сегменте

фиксируется опорная модель $\hat{s}_t$. Выбор опорной модели задаётся результатами проведённого исследования. Линейная и полиномиальная регрессии используются как базовые ориентиры, но из-за ограниченного качества по R^2 и RMSE не рассматриваются как предпочтительные. Для получения более точной опорной оценки применяются либо градиентный бустинг, либо LSTM. Градиентный бустинг рассматривается как предпочтительный метод для опорного прогнозирования дистанции при ограничениях вычислительной сложности и необходимости устойчивой нелинейной аппроксимации. LSTM рассматривается как предпочтительный метод при необходимости учёта последовательной структуры и при наличии выраженной инерционности поведения в данных. Выбор между бустингом и LSTM фиксируется для каждого сегмента по метрикам на валидации. В практическом контуре допускается использование бустинга для дистанции и LSTM для скорости, что согласуется с тем, что LSTM показала наилучшее качество по RMSE в задаче прогнозирования скорости.

Шаг 3. Формирование множества калибровочных точек. В каждом сегменте $\mathcal{D}_k$ формируется множество точек

$$\mathcal{P}_k = \{(v_i, \tilde{s}_i, w_i)\}_{i=1}^{n_k},$$

где v_i берётся как $v_{t_i}^f$, а $\tilde{s}_i$ является целевой дистанцией для калибровки. В качестве $\tilde{s}_i$ допускается использовать либо наблюдаемую дистанцию s_{t_i} , либо опорную оценку $\hat{s}_{t_i}$, либо их сглаженную комбинацию. Использование $\hat{s}_t$ целесообразно в сегментах с повышенным уровнем шумов измерения дистанции. Веса w_i задают доверие к точке и понижают вклад наблюдений с артефактами синхронизации, пропусками и короткими разрывами.

Шаг 4. Калибровка $\hat{C}_{0,k}, \hat{C}_{1,k}, \hat{C}_{2,k}$ по множеству точек. Для каждого сегмента коэффициенты оцениваются решением выпуклой задачи регуляризованной калибровки с ограничениями

$$\hat{C}_k = \arg \min_{C_0, C_1, C_2} \sum_{i=1}^{n_k} w_i \left(\tilde{s}_i - (C_0 + C_1 v_i + C_2 v_i^2) \right)^2 + \lambda_0 C_0^2 + \lambda_1 C_1^2 + \lambda_2 C_2^2, \quad (13)$$

Параметры регуляризации $\lambda_0, \lambda_1, \lambda_2$ выбираются так, чтобы обеспечить устойчивость оценок при переносе между поездками внутри сегмента. Результатом является таблица сегментных коэффициентов $\mathcal{T} = \{(k, \hat{C}_k)\}$.

Шаг 5. Правило обновления коэффициентов в режиме реального времени. В режиме для каждого момента времени t определяется сегмент $k(\mathbf{z}_t)$ и выбираются коэффициенты

$$\hat{C}_t = \hat{C}_{k(\mathbf{z}_t)}.$$

При смене сегмента допускается сглаживание перехода коэффициентов по времени, чтобы избежать скачков ограничения.

Шаг 6. Вычисление динамически корректируемого требования безопасности. На каждом шаге вычисляется требуемая дистанция

$$s_t^{req} = \widehat{C}_{0,t} + \widehat{C}_{1,t}v_t^f + \widehat{C}_{2,t}(v_t^f)^2, \quad (14)$$

после чего применяется нижнее отсечение $s_t^{req} \leftarrow \max(s_t^{req}, s_{\min})$, согласованное с порогом фиксации столкновения ε_c .

Шаг 7. Прогноз модели следования за лидером, основанной на данных. Пусть модель, основанная на данных, формирует прогноз $\widehat{v}_{t+1}^{ML}$ или $\widehat{a}_t^{ML}$ на основе окна длины seq_len . Для совместимости допускается постановка через прогноз скорости. Тогда прогнозируемая дистанция на следующий шаг вычисляется как

$$\widehat{s}_{t+1}^{ML} = \max(0, s_t + (v_t^\ell - \widehat{v}_{t+1}^{ML})\Delta t).$$

Шаг 8. Детекция аномального события относительно классического ограничения. Прогноз маркируется как аномальный, если

$$\widehat{s}_{t+1}^{ML} < s_{t+1}^{req} \quad \text{или} \quad \widehat{s}_{t+1}^{ML} \leq \varepsilon_c, \quad (15)$$

и дополнительно, если требуется, контролируются ограничения на скорость и резкость изменения действия в терминах выбранных эксплуатационных пределов.

Шаг 9. Проекция прогноза на допустимое множество. Если аномалия отсутствует, принимается $\widehat{v}_{t+1}^{ML}$. Если аномалия выявлена, строится безопасный прогноз $\widehat{v}_{t+1}^{safe}$, который минимально отклоняется от $\widehat{v}_{t+1}^{ML}$ и обеспечивает выполнение ограничения. В простейшем варианте используется одномерная проекция по скорости

$$\widehat{v}_{t+1}^{safe} = \arg \min_{v \in [0, v_{\max}]} (v - \widehat{v}_{t+1}^{ML})^2 \quad \text{при условии} \quad \max(0, s_t + (v_t^\ell - v)\Delta t) \geq s_{t+1}^{req}. \quad (16)$$

Итоговый прогноз определяется правилом переключения

$$\widehat{v}_{t+1} = \begin{cases} \widehat{v}_{t+1}^{ML}, & \text{если условие (15) не выполняется,} \\ \widehat{v}_{t+1}^{safe}, & \text{если условие (15) выполняется.} \end{cases}$$

Шаг 10. Логирование срабатываний и итеративное уточнение. Фиксируются контекст $\mathbf{z}_t$, выбранные коэффициенты $\hat{\mathbf{C}}_t$, значения s_t^{req} , исходный прогноз $\hat{v}_{t+1}^{ML}$ и скорректированный прогноз $\hat{v}_{t+1}^{safe}$. Накопленные события используются для уточнения сегментной калибровки и для дообучения модели, основанной на данных, на режимах, в которых она чаще нарушает ограничение.

Результатом является классическая модель с динамически корректируемыми коэффициентами C_0, C_1, C_2 , которая используется как ограничение поверх модели следования за лидером, основанной на данных. Метод обеспечивает динамическое обновление контекста классической модели следования за лидером. Это позволяет подавлять аномальные прогнозы и снижать риск появления аномальных событий в замкнутом контуре, при этом сохраняется точность модели, основанной на данных, в режимах, где её прогнозы не нарушают ограничение.

Следующим этапом разработан гибридный метод прогнозирования микроскопических характеристик транспортного потока. Для выбора его компонентов исследованы Residual IDM, MLP, LSTM, TCN и Informer на наборах highD, NGSIM, DAS1, DAS2, Waymo, Lyft и собственном корпусе данных. Оценка выполнялась по точности прогноза, плавности движения, частоте столкновений и поведению модели в замкнутом контуре.

Полученные результаты показали, что LSTM устойчиво моделирует краткосрочную динамику и инерционность реакции водителя, а Informer эффективнее учитывает длинный временной контекст. При этом чисто нейросетевые архитектуры не устраняют риск аномальных выходов без дополнительного механизма стабилизации. Поэтому предложен метод, в котором предобученная модель формирует признаки для архитектуры Bi-LSTM-LSTM-Informer, а аномальные события подавляются внешним ограничителем на базе SSSR-Micro с динамической коррекцией коэффициентов C_0, C_1, C_2 .

На основании полученных результатов в структуру метода включены следующие блоки.

1. **Предобученный Bi-LSTM энкодер каузальных признаков.** Bi-LSTM формирует информативное представление временного окна за счёт обучения на последовательности в прямом и обратном направлениях. Для онлайн-прогноза используется каузальная форма представления, при которой извлекается forward-ветвь предобученной модели либо применяется замороженный экстрактор признаков с последующей каузализацией окна. Данный блок повышает устойчивость представления состояния без существенного роста вычислительной сложности на инференсе.
2. **Гибридная архитектура LSTM-Informer.** LSTM-компонента моделирует инерционность и задержку реакции, а Informer-компонента обрабатывает длинные зависимости, характерные для

протяжённых переходных процессов. Informer выбран вместо полного Transformer благодаря меньшей вычислительной сложности, обусловленной ProbSparse-вниманием и дистилляцией временного представления.

3. **Механизм интерпретации и диагностики устойчивости.** В метод включён режим анализа Informer-внимания, позволяющий определять временные позиции входного окна, наиболее влияющие на прогноз. Дополнительно используется оценка Липшицевой нормы как индикатор чувствительности модели к возмущениям входа, что важно для диагностики резких управляющих воздействий и роста рывка.
4. **Коррекция аномальных выходов на базе классической модели.** Для подавления редких аномальных выходов нейросети метод дополняется внешним корректирующим контуром на базе SSSR-Micro с динамической коррекцией коэффициентов C_0, C_1, C_2 , разработанной ранее в главе.

Таким образом, предлагаемый метод строится как гибридная композиция. Нейросетевой модуль отвечает за точность и извлечение зависимостей в данных, а классический модуль с динамической коррекцией коэффициентов обеспечивает подавление аномальных режимов и формализуемые ограничения в замкнутом контуре. Формальное описание предлагаемого метода представлено ниже.

Шаг 1. Формирование окна наблюдений. Для каждого шага формируется каузальное окно длины L :

$$X_t = [x_{t-L+1}, \dots, x_t] \in \mathbb{R}^{L \times F}, \quad F = 6.$$

Шаг 2. Извлечение каузальных признаков предобученной Bi-LSTM. Предобученная Bi-LSTM обучается офлайн на последовательностях (в двунаправленном режиме). Для использования в онлайн-контуре выполняется каузализация одним из двух способов:

- **Вариант 1:** извлекается forward-ветвь Bi-LSTM, образуя односторонний энкодер $\Phi_{\rightarrow}$, и вычисляются признаки

$$H_t^{\text{causal}} = \Phi_{\rightarrow}(X_t) \in \mathbb{R}^{L \times D_s}. \quad (17)$$

- **Вариант 2 (для диагностики):** используется замороженный BiLSTMFeatureExtractor с выходом $\mathbb{R}^{L \times 2D_s}$. В этом случае требуется дополнительная каузализация, поскольку двунаправленная обработка использует будущие элементы внутри окна.

Шаг 3. Прогноз нейросетевого действия гибридной моделью LSTM-Informer. Основная модель π_{θ} реализует два параллельных представления окна X_t :

1. **LSTM-ветвь** (или внешняя $\Phi_{\rightarrow}$), формирующая $H_t^{\text{seq}} \in \mathbb{R}^{L \times D_s}$;
 2. **Informer-ветвь**, формирующая $H_t^{\text{inf}} \in \mathbb{R}^{L \times D}$ за счёт каузального разреженного внимания ProbSparse и дистилляции по времени.
- Слияние ветвей выполняется через **fusionconcat** или **fusionadd**:

$$H_t^{\text{fus}} = \begin{cases} [H_t^{\text{seq}}; H_t^{\text{inf}}], & \text{concat,} \\ A(H_t^{\text{seq}}) + H_t^{\text{inf}}, & \text{add,} \end{cases} \quad (18)$$

где $A(\cdot)$ - линейное выравнивание размерностей.

Результатом является прогноз ускорения:

$$\hat{a}_t^{\text{nn}} = \pi_{\theta}(X_t, H_t^{\text{causal}}), \quad (19)$$

где в реализации прогнозируется последовательность $\hat{a}_{t-L+1:t}$, но для замкнутого контура используется текущее значение $\hat{a}_t$.

Шаг 4. Детекция аномального выхода нейросети. Вводится предикат аномальности $\mathcal{A}(\cdot)$, который активируется при нарушении инвариантов безопасного движения или при выходе за допустимые диапазоны. В практической реализации это включает:

$$\mathcal{A}_t = \mathbb{I}[s_t < s_{\min}] \vee \mathbb{I}[\hat{a}_t^{\text{nn}} \notin [a_{\min}, a_{\max}]] \vee \mathbb{I}[\widehat{\text{TTC}}_t < \tau],$$

где $\widehat{\text{TTC}}_t$ вычисляется по прогнозной динамике/текущему состоянию (в зависимости от выбранного протокола). Конкретный набор условий фиксируется в программной реализации эксперимента.

Шаг 5. Коррекция аномального выхода на базе модели SSSR-Micro с динамической коррекцией C_0, C_1, C_2 . Действие нейросети корректируется с использованием классической модели безопасной дистанции SSSR-Micro, в которой вводится динамический габарит:

$$s_{\text{req}}^{\text{BG}}(v_t^f) = C_0 + C_1 v_t^f + C_2 (v_t^f)^2, \quad (20)$$

а коэффициенты C_0, C_1, C_2 *адаптируются во времени* в соответствии с разработанным ранее методом (раздел с динамической коррекцией коэффициентов).

Коррекция реализуется как проекция нейросетевого действия на допустимое множество $\mathcal{U}_t$, заданное условиями безопасности (требованием $s_{t+1} \geq s_{\text{req}}^{\text{BG}}(v_{t+1}^f)$ и ограничениями по ускорению):

$$\hat{a}_t = \begin{cases} \hat{a}_t^{\text{nn}}, & \mathcal{A}_t = 0, \\ \Pi_{\mathcal{U}_t}(\hat{a}_t^{\text{nn}}), & \mathcal{A}_t = 1, \end{cases} \quad (21)$$

где Π_t - оператор проекции, вычисляемый с использованием ограничения и динамически скорректированных коэффициентов C_0, C_1, C_2 . Этот шаг является ключевым отличием предлагаемого метода от подходов, основанных на нейронных сетях. Он гарантирует подавление редких аномальных режимов без необходимости переобучения нейронной сети специально для редких событий.

Шаг 6. Обновление состояния в замкнутом контуре. Скорость и дистанция обновляются по дискретной кинематике:

$$\hat{v}_{t+1}^f = \max \left(0, \hat{v}_t^f + \hat{a}_t \Delta t \right), \quad (22)$$

$$\hat{s}_{t+1} = \max \left(0, \hat{s}_t + (v_t^\ell - \hat{v}_t^f) \Delta t \right). \quad (23)$$

Шаг 7. Диагностика устойчивости и интерпретация Informer. В ходе обучения/валидации дополнительно вычисляется оценка Липшицевой нормы (верхняя оценка чувствительности выхода к возмущениям входа), используемая как диагностический индикатор потенциальной резкости управления и склонности к усилению шумов.

Informer интерпретируется посредством анализа механизма ProbSparse. Выбирается подмножество временных индексов, которые способствуют обновлению контекста. Эти индексы интерпретируются как ключевые фрагменты истории, влияющие на решение модели (фрагменты, отмечающие начало замедления лидера, периоды устойчивой конвергенции и т. д.). Этот анализ повышает прозрачность работы компонента, выполняющегося в течение длительного времени, и предоставляет инструмент для диагностики снижения производительности на конкретных наборах данных.

Тем самым Bi-LSTM обеспечивает устойчивое предварительное представление состояния, LSTM моделирует инерционность реакции, а Informer дополняет прогноз механизмом долговременного контекста с вычислительно эффективным вниманием.

Экспериментальная проверка разработанного метода выполнена на наборах highD, NGSIM, DAS1, DAS2, Waymo, Lyft и собственном корпусе данных. Средние значения метрик составили MSE 19,42, MAE 3,46, RMSE 4,33, при этом во всех рассмотренных наборах зафиксировано нулевое число столкновений. Минимальная ошибка получена на наборе Waymo, где RMSE составила 3,21, а на собственном корпусе данных RMSE составила 3,36 при минимальном значении рывка 1,00. Полученные результаты показывают, что предложенный гибридный метод сочетает точность моделей, основанных на данных, способность учитывать долгосрочный контекст и проводить интерпретируемый анализ ключевых временных сегментов с использованием Informer, а также формализованную коррекцию выбросов

на основе классической модели с динамически адаптируемыми коэффициентами.

В **четвертой главе** решается задача разработки методологии конфигурирования архитектуры исполнения моделей ИИ в ИТС. Необходимость такой методологии обусловлена тем, что после построения прогнозной модели требуется определить проверяемую конфигурацию её исполнения в вычислительной среде. Для ИТС этот этап не может сводиться к выбору сервера, библиотеки машинного обучения или контейнерной платформы, поскольку одновременно должны учитываться задержка ответа, пропускная способность, доступность сервиса, защищённость каналов обмена, устойчивость к изменению интенсивности событий и воспроизводимость проверки.

Показано, что архитектура исполнения зависит от сценария применения модели. Онлайн-прогнозирование, обработка телеметрии, обнаружение инцидентов и оценка времени поездки требуют потокового исполнения с ограничением по времени ответа. Аналитическая обработка и переобучение моделей допускают более длительные вычислительные циклы, но предъявляют повышенные требования к хранилищам, пакетной обработке и повторяемости вычислений. Поэтому одна и та же модель при разных схемах размещения может иметь различные эксплуатационные свойства.

В главе рассмотрены варианты исполнения моделей ИИ с использованием реляционной БД, GPU-контура, Nadoop/Spark, объектного хранилища S3, In-Memory БД, синхронного обращения по HTTPS/gRPC и асинхронного обращения через Kafka. Показано, что сравнение таких вариантов должно выполняться не по названию платформы, а по сквозному маршруту обработки данных. В связи с этим предложена методология, которая переводит описание сценария ИТС в проверяемую конфигурацию исполнения модели с фиксацией маршрута данных, параметров вычислительных узлов, ограничений SLA, инвариантов безопасности и автоматических проверок.

Центральным элементом разработанной методологии является представление архитектуры исполнения в виде ориентированного мультиграфа $\mathcal{G} = (V, E)$, где множество V включает сервисы модели, контроллеры доступа к данным, брокеры сообщений, хранилища, транспортные адаптеры, вычислительные узлы и компоненты наблюдаемости, а множество E задаёт типизированные каналы обмена между ними. Для ребра $e \in E$ фиксируются задержка T_e , скорость обслуживания μ_e , протокол обмена и требования к защищённому соединению. Для вершины $v \in V$ задаются вычислительная способность μ_v , доступная память, политика масштабирования и признаки готовности сервиса. Сценарий исполнения определяется входным процессом $\lambda(t)$, типом и размером полезной нагрузки, требованиями к реакции сервиса и множеством контролируемых путей П.

Сквозной путь $\pi \subseteq E$ соответствует одному маршруту обработки запроса или сообщения. В такой путь входят сетевые переходы, вызовы сервисов, обращения к хранилищам, операции брокера, сериализация, вычисление модели и запись результата. Если стадии выполняются последовательно и перекрытие не используется, верхняя оценка задержки пути задаётся выражением

$$L(\pi) \leq \sum_{e \in \pi} T_e + \sum_{v \in V(\pi)} W_v,$$

где W_v обозначает время ожидания и обработки в вершине v . Для конвейерного исполнения учитывается этап заполнения конвейера. Для серии из K сообщений суммарное время обработки оценивается как

$$T_{\text{tot}}(K) \approx T_{\text{fill}} + \frac{K}{\mu_{\text{bottleneck}}},$$

где $\mu_{\text{bottleneck}}$ - пропускная способность наиболее медленной стадии маршрута.

Пропускная способность полного маршрута определяется не средним значением по компонентам, а самой медленной вершиной или ребром. Поэтому для пути π используется оценка

$$\mu(\pi) = \min \left(\min_{e \in \pi} \mu_e, \min_{v \in V(\pi)} \mu_v \right).$$

Для ИТС это условие имеет принципиальное значение, поскольку интенсивность входных событий изменяется неравномерно, а в периоды пиковой нагрузки рост очередей может привести к нарушению SLA. Условие устойчивости задаётся с запасом по коэффициенту загрузки ρ :

$$\sup_{t \in T_{\text{peak}}} \lambda(t) \leq \rho \mu(\pi), \quad 0 < \rho < 1.$$

После проверки устойчивости вводится вероятностное ограничение по задержке:

$$\mathbb{P}\{L(\pi) \leq \Delta_\pi\} \geq 1 - \varepsilon_\pi, \quad \forall \pi \in \Pi,$$

где Δ_π - допустимая задержка для маршрута π , а ε_π - допустимая вероятность нарушения SLA.

Синтез конфигурации выполняется после выделения контролируемых путей. На этом этапе выбирается не отдельная технология, а набор параметров, определяющих размещение сервисов, параллелизм обработки, режим передачи данных, резервирование и аппаратную платформу. Управляемый вектор конфигурации имеет вид

$$x = (n, c, b, k, P, r, h, p),$$

где n - число реплик сервисов модели, c - суммарная параллельность обработки, b - размер микропакета или окна, k - параллелизм чтения крупных объектов, P - число секций брокера сообщений, r - коэффициенты репликации, h - класс аппаратной платформы, p - режим вычислительной точности. Такая параметризация позволяет сравнивать варианты исполнения по управляемым настройкам, а не по названиям используемых технологий.

Выбор конфигурации формулируется как многокритериальная задача:

$$\min_{x, \Pi} J(x, \Pi) = \alpha_1 \mathbb{E} \left[\max_{\pi \in \Pi} L(\pi) \right] + \alpha_2 C(x) + \alpha_3 E(x),$$

где $C(x)$ - стоимость ресурсов, $E(x)$ - энергопотребление конфигурации, а коэффициенты $\alpha_1, \alpha_2, \alpha_3$ задают относительную значимость задержки, стоимости и энергии. Конфигурация считается допустимой только при выполнении ограничений:

$$\begin{aligned} \mathbb{P}\{L(\pi) \leq \Delta_\pi\} &\geq 1 - \varepsilon_\pi, & \forall \pi \in \Pi, \\ A(x) &\geq A_{\min}, & M(x) \geq M_{\text{need}}, \\ \sup_{t \in \mathcal{T}_{\text{peak}}} \lambda(t) &\leq \rho \mu(\pi), & \forall \pi \in \Pi. \end{aligned}$$

Здесь $A(x)$ обозначает доступность конфигурации, $M(x)$ - доступный объём памяти, M_{need} - требуемую память для выполнения сценария. Таким образом, выбранная конфигурация должна одновременно удовлетворять ограничениям по задержке, доступности, памяти и устойчивости относительно входной нагрузки.

Разработанная методология задаёт порядок перехода от требований сценария ИТС к проверяемой конфигурации исполнения модели.

1. **Формализация сценария и требований.** На первом шаге задаётся входной процесс $\lambda(t)$, множество контролируемых путей Π и пороги $\{\Delta_\pi, \varepsilon_\pi\}_{\pi \in \Pi}$ для вероятностного SLA:

$$P\{L(\pi) \leq \Delta_\pi\} \geq 1 - \varepsilon_\pi, \quad \forall \pi \in \Pi,$$

дополнительно фиксируются требования по доступности $A_{\min}$, стоимости $C_{\max}$, памяти M_{need} и качеству $\text{Ass}_{\min}$, если для выбранной модели вводится ограничение на точность.

2. **Конфигурирование структуры мультиграфа исполнения.** По выбранному сценарию строится ориентированный мультиграф $\mathcal{G} = (V, E)$, включающий только элементы фактического маршрута запроса или потока данных. Вершинам и рёбрам графа сопоставляются сервисы, контроллеры, хранилища, брокеры, вычислительные узлы, транспортные адаптеры и каналы обмена. После этого задаются контролируемые пути Π .

3. **Параметризация конфигурации.** Для выбранной структуры формируется вектор параметров

$$x = (n, c, b, k, P, r, h, p),$$

обозначения приводились выше.

4. **Оценивание пропускной способности и задержки.** Для каждого пути $\pi \in \Pi$ рассчитываются пропускная способность $\mu(\pi)$ и задержка $L(\pi)$. Если в путь входят GPU, Kafka, S3 или БД, дополнительно применяются специализированные оценки. На этом шаге проверяется не отдельный сервис, а весь маршрут прохождения сообщения с учётом передачи данных, ожидания, вычисления и доступа к хранилищу.
5. **Проверка устойчивости по нагрузке.** Для пикового интервала $\mathcal{T}_{\text{peak}}$ проверяется, остаётся ли входной поток ниже допустимой загрузки маршрута с запасом $\rho \in (0,1)$:

$$\sup_{t \in \mathcal{T}_{\text{peak}}} \lambda(t) \leq \rho \mu(\pi), \quad \forall \pi \in \Pi.$$

Это условие исключает работу маршрута у предела пропускной способности и снижает риск накопления очередей.

6. **Выбор параметров как задача оптимизации.** После расчёта задержек, пропускной способности и устойчивости выполняется выбор параметров x и набора путей Π . Этот выбор рассматривается как многокритериальная задача с ограничениями SLA, доступности, памяти и устойчивости. Результатом является пара x^*, Π^* , определяющая допустимую конфигурацию и набор путей выполнения.
7. **Переход к развёртываемой конфигурации.** На основе параметров x^*, Π^* генерируются конфигурации для производственной среды. Они включают количество реплик, ограничения ЦП и памяти, параметры брокера, параметры шардинга и настройки безопасности, включая mTLS, RBAC и секреты.
8. **Верификация инвариантов в CI и нагрузочная проверка SLA.** Для каждого пути $\pi \in \Pi^*$ в CI проверяются инварианты доступности и безопасности Ready, TLS, RBAC, EP, HP, MNT и OK. Затем выполняется статистическая проверка задержки на контрольной нагрузке. При нарушении хотя бы одного инварианта или критерия SLA конфигурация не допускается к выпуску.
9. **Эксплуатационный контур наблюдаемости и адаптации.** После развёртывания контролируются $\lambda(t)$, $L(\pi)$, глубина очередей, доля нарушений SLA и утилизация ресурсов. При выходе нагрузки или задержки за допустимые границы корректируются

параметры $x(t)$, включая $n(t)$, $b(t)$, $c(t)$ или $P(t)$. Такая адаптация рассматривается как управляемое изменение конфигурации, не нарушающее инварианты безопасности и ограничения SLA.

В результате методология задаёт последовательный цикл требования — конфигурирование — оптимизация — развёртывание — верификация — эксплуатация. В этом цикле каждое требование должно быть связано с параметром конфигурации и проверяемой метрикой. Поэтому результат можно проследить от исходного сценария ИТС до конкретных настроек размещения и тестов в CI.

Итоговая конфигурация получается не путём ручного выбора технологии, а через формальное описание маршрута и его параметров. Разные варианты исполнения приводятся к единому набору управляемых величин. Автоматическая проверка фиксирует, выполняются ли требования по задержке, доступности, безопасности и устойчивости. За счёт этого архитектурное решение можно воспроизвести на другом стенде и затем адаптировать под иной профиль транспортной нагрузки.

В **пятой главе** разработана методология построения инфраструктуры ИТС, ориентированная на обеспечение отказоустойчивости, масштабируемости, гибкости обработки больших разнородных данных и воспроизводимого тестирования. В отличие от предыдущей главы, где рассматривалась архитектура исполнения моделей ИИ, здесь объектом исследования является инфраструктура ИТС как распределённая система сбора, хранения, обработки, передачи и проверки транспортных данных.

Необходимость разработки методологии обусловлена высокой неоднородностью источников данных, нестационарностью нагрузки, географическим распределением серверов, требованиями к согласованности данных и ограниченной доступностью реальных транспортных наборов для испытаний. Инфраструктура ИТС должна обеспечивать полный жизненный цикл данных, включая сбор, валидацию, хранение, потоковую и пакетную обработку, аналитическую интерпретацию, визуализацию и интеграцию с прикладными сервисами.

Рассмотрим множество серверов ИТС $S = \{s_1, \dots, s_n\}$, размещённых в разных сетевых зонах. Пусть сетевое расстояние между узлами определяется дву-сторонней задержкой $\text{RTT}(s_i, s_j)$, образующей веса полного графа $G = (S, E)$.

Требуется найти разбиение S на подкластеры $= \{C_1, \dots, C_k\}$, $C_m \subseteq S$, и множество лидеров $L = \{\ell_1, \dots, \ell_k\}$, где $\ell_m \in C_m$, при ограничениях:

$$C_m \cap C_{m'} = \emptyset, \quad \bigcup_{m=1}^k C_m = S, \quad (24)$$

$$|C_m| \geq 3 \quad (\text{минимум для большинства в Raft}), \quad (25)$$

$$\forall s_i \in C_m : \text{assign}(s_i) = \ell_m. \quad (26)$$

Определим функционал качества топологии:

$$J(.,L) = \alpha \sum_{m=1}^k \sum_{s_i \in C_m} \text{RTT}(s_i, \ell_m) + \beta \mathbb{E}[T_{\text{failover}}(.,L)] + \gamma T_{\text{setup}}(.,L),$$

где $\alpha, \beta, \gamma \geq 0$ — весовые коэффициенты; T_{failover} — время переизбрания лидера при отказах; T_{setup} — время включения нового узла.

Целью является минимизация $J(C, L)$ при соблюдении согласованности данных в каждом C_m согласно Raft (безопасность, лог-матчинг, лидерство) и обеспечить динамическую реконфигурацию при изменении и состава без ручного вмешательства администратора.

Референсная архитектура инфраструктуры ИТС задаётся тремя взаимосвязанными представлениями. Функциональное представление фиксирует состав сервисов, группы функций, каналы предоставления данных, серверные службы, аналитические компоненты, зоны хранения, оценки устойчивости и потребления данных. Интеграционное представление описывает типы коннекторов между внешними источниками, сервисами загрузки, транспортным уровнем, посредническим уровнем и прикладными компонентами. Технологическое представление определяет допустимые классы программных средств для реализации сервисов, включая распределённые файловые системы, SQL и NoSQL хранилища, брокеры сообщений, средства потоковой обработки, геоинформационные сервисы и инструменты визуализации.

Такое разделение позволяет рассматривать инфраструктуру ИТС не как набор отдельных программных продуктов, а как воспроизводимую архитектурную схему, в которой функциональные требования, протоколы обмена и технологические решения связаны с задачами сбора, хранения, обработки и предоставления транспортных данных.

Современные одноуровневые схемы кластеризации ограничены при размещении серверов в разных сетевых зонах, поскольку высокий RTT снижает эффективность обработки, увеличивает задержку получения задач и затрудняет поддержание согласованности. Для устранения данного ограничения разработан метод кластеризации серверов, основанный на совместном использовании метрики сетевого расстояния, протокола консенсуса Raft и автоматического выбора лидера сервисом Agent.

Метод предусматривает, что каждый сервер измеряет RTT до других серверов, выбирает ближайшего лидера и включается в соответствующий подкластер. Каждый подкластер содержит не менее трёх серверов, что обеспечивает большинство для Raft. Если ближайший сервер не найден, область поиска расширяется по допустимому сетевому расстоянию до формирования работоспособного подкластера. После группировки серверов выполняется голосование, один узел получает статус лидера, остальные узлы получают статус последователей. Лидер контролирует состояние последователей и исключает узлы при длительной сетевой задержке, некорректной работе сервиса Agent или отказе сервера. Лидеры подкластеров связываются между собой, что формирует децентрализованную инфраструктуру с повторным голосованием при отказе лидера.

Преимущества предложенного метода состоят в динамическом включении новых серверов, автоматическом выборе ближайшего лидера по сетевому расстоянию, повторном присоединении узла при изменении топологии, контроле последователей со стороны лидеров и ограничении межузлового взаимодействия ближайшими связями. Это снижает административные затраты и повышает устойчивость распределённой инфраструктуры ИТС при изменении состава серверов и сетевых задержек.

Для реализации децентрализованной ИТС используется микросервисная архитектура, включающая сервисы client, storage, auth, database и agent. Сервис storage принимает запросы пользователей и агентов, передаёт неструктурированные данные в database и взаимодействует с сервисами auth, database, agent и client. Сервис auth выполняет аутентификацию, авторизацию, регистрацию пользователей и подключение агентов. Сервис database отвечает за хранение данных и использует Raft для обеспечения согласованности. Сервис agent обеспечивает добавление серверов, запуск новых экземпляров database и оповещение storage об изменении состава кластера. Сервис client включает клиентское приложение и библиотеки агентов, предназначенные для получения информации и передачи неструктурированных данных.

Протокол взаимодействия строится вокруг аутентификации пользователя или агента, выдачи токенов, проверки прав доступа через auth, передачи запроса в storage, обращения к database и возврата результата либо сохранения поступивших данных. Такая схема позволяет разделить функции доступа, хранения, масштабирования и клиентского взаимодействия, сохранив возможность размещения сервисов как на одном сервере, так и на разных узлах инфраструктуры.

Эффективность предложенного метода кластеризации оценивалась по времени подготовки сервера ИТС, времени загрузки разнородных данных и времени переключения лидера. Время подготовки сервера сократилось со 121 минуты до 15 минут, что соответствует улучшению

на 87,5%. Время загрузки 10 Гб разнородных данных ИТС сократилось с 46,2 минуты до 24,2 минуты, что соответствует улучшению на 47,6%. Время переключения лидера составило около 730 мс. Полученные результаты объясняются автоматическим присоединением серверов по RGT, локализацией репликации и сокращением ручной конфигурации. Ограничением проверки является лабораторная среда, поэтому для географически распределённых сетей с переменным джиттером и потерями требуется дополнительная настройка параметров Raft и расширение сценариев тестирования.

Далее в главе рассматривается обеспечение отказоустойчивости вычислительной инфраструктуры ИТС при отказах узлов без немедленного ввода дополнительных ресурсов. Используется модель постепенной деградации, при которой отказавшие вычислительные модули исключаются из работы, а задачи перераспределяются между оставшимися работоспособными модулями. Предполагается, что каждый модуль находится в одном из двух состояний, соответствующих работоспособности или отказу.

Далее используется следующее понятие структурного состояния системы:

Структурное состояние (s -состояние) вычислительной системы, которое состоит из n вычислительных модулей, представляется в виде вектора: $s_v = (\sigma_1, \dots, \sigma_n)$, где $\sigma_i = 0$, если i -й элемент работоспособен, $\sigma_i = 1$, если i -й элемент отказал $i = 1, \dots, n$, $v = 1, \dots, 2^n$.

Выделим следующие s -состояния:

- $s^0 = (0, \dots, 0)$ – начальное (все вычислительные модули работоспособны);
- s_v – искажённое – любое s -состояние, не являющееся начальным: $s_v \neq s^0$;
- $s^1 = (1, \dots, 1)$ – конечное (отказ всех ВМ).

В таком случае множество всех состояний ВС будет: $\Omega_s = \{s_v\}$, $v = 1, \dots, 2^n$.

На основе указанного критерия отказа вычислительная система делится на непересекающиеся подмножества. Одно включает работоспособные модули, продолжающие выполнять задачи, а второе – модули, находящиеся в отказе и неспособные к выполнению задач: $\Omega_s = S \cup S^F$, где S – подмножество работоспособных s -состояний, называемое в дальнейшем структурной областью работоспособности системы, S^F – подмножество неработоспособных s -состояний (то есть соответствующих отказу ВС).

Для вычислительной системы без механизмов отказоустойчивости множество S ограничивается стартовым состоянием s^0 – система функционирует без сбоев. В отказоустойчивой системе в S присутствует минимум одно нарушенное состояние s_ω , не тождественное s^1 , которое представляет собой состояние полной неработоспособности. Такие состояния отражают

способность системы функционировать с допустимой деградацией в условиях отказов: $S = \{s^0\} \cup S_r^\omega$, где S_r^ω — некоторое множество искажённых работоспособных s -состояний.

Очевидно, что в общем случае при возможности отказов любых ВМ система в произвольный момент времени t может находиться в любом из s -состояний: $s_\nu = (\sigma_1, \dots, \sigma_n)$, $\nu = 1, \dots, 2^n$.

Отказоустойчивость достигается перераспределением задач отказавших модулей на работоспособные модули. Рассматривается система с фиксированным множеством задач $\Lambda^{(0)}$ без задания порядка выполнения и взаимосвязей. $\Lambda^{(0)}$ — все задачи, которые в начальном состоянии s^0 могут быть выполнены при имеющихся ресурсах.

Считаем, что для состояния s^0 задано начальное распределение задач (РЗ) между всеми ВМ, при котором каждый модуль M_i способен выполнять некоторое подмножество задач: $\Lambda_i^{(0)} = \{U_1^i, U_2^i, \dots, U_{m_i}^i\}$, которые называются собственными задачами модуля M_i .

При отказе каких-либо ВМ система переходит в одно из искажённых состояний s_ω и утрачивает способность выполнять собственные задачи отказавших ВМ. Сохранить способность системы выполнять эти задачи можно с помощью такого перераспределения выполняемых системой задач по оставшимся работоспособным ВМ, при котором собственные задачи отказавших ВМ будут выполняться в работоспособных ВМ.

Есть множество S работоспособных состояний системы, определяющих требуемую степень отказоустойчивости вычислительной системы, которая должна быть достигнута.

При подготовке к работе для каждого состояния s_r^ω , принадлежащего заданному множеству S_r^ω искажённых работоспособных состояний, вычисляется такой план D_r^ω распределения задач по работоспособным ВМ, при котором достигается наилучшее значение некоторого заданного показателя качества функционирования системы при заданных ограничениях на другие выбранные показатели качества и на ресурсы системы.

Указанный план представляет собой план рационального назначения задач в данном состоянии. Все задачи, подлежащие запуску на виртуальной машине M_i согласно любому из планов D_r^ω , в сумме образуют множество $\Omega_i^{(\omega)}$.

Размещение задачи U_j на ВМ M_i определяется её принадлежностью множеству $\Omega_i^{(\omega)}$. При этом неважно, хранится ли модуль задачи в памяти машины или подгружается динамически.

Под основным размещением задач (РмЗ) будем понимать такое распределение, при котором каждая задача U_j из начального множества $\Omega^{(0)}$ закрепляется ровно за одним вычислительным модулем. Задачи множества $U_j \in \Omega^{(0)}$ назовём основными. План основного РмЗ обозначим матрицей D^0 — это двоичная матрица размера $L \times n$: $D^0 = \|d_{ji}^0\|_{L \times n}$, где $d_{ji}^0 = 1$, если задача U_j назначена к выполнению на ВМ M_i и размещена в ней;

$d_{ji}^0 = 0$ — в противном случае. Считаем, что начальному состоянию s^0 соответствует основное РмЗ.

Под дополнительным РмЗ понимаем матрицу $Y = \|y_{ji}\|_{L \times n}$, описывающую размещение резервных копий задач. Она получается путём логического объединения планов $D_r^\omega = \|d_{ji}^\omega\|_{L \times n}$, рассчитанных для всех искажённых, но работоспособных состояний s_r^ω : $y_{ji} = \bigcup_{s_r^\omega \in s_r^\omega} d_{ji}^\omega$.

Матрица ОУ-РмЗ $Z = \|z_{ij}\|_{L \times n}$ строится из двух компонентов — основного $D^0 = \|d_{ji}^0\|_{L \times n}$ и дополнительного $Y = \|y_{ji}\|_{L \times n}$. Элемент Z определяется правилом: $z_{ij} = x_{ij} \oplus y_{ij}$. Здесь операция $\oplus$ используется в смысле объединения информативных признаков с целью формирования расширенного представления состояния системы, обеспечивающего учет нескольких источников данных в единой модели.

Такое решение обеспечивает резервирование: одна и та же задача получает копии на разных ВМ системы, при этом в каждой ВМ допускается не более одной копии данной задачи. В исходном состоянии s^0 запускаются лишь основные задачи согласно начальному плану размещения D^0 (основному РмЗ).

Если из-за отказов части ВМ система переходит в искажённое, но работоспособное состояние s_r^ω , режим работы следует перестроить по плану РмЗ D_r^ω : активировать резервные копии задач на доступных ВМ и выполнять их вместе с основными.

Если в системе с восстановлением ВМ неисправные модули возвращаются к работе, увеличивается состав доступных ВМ, и система переключается в состояние с меньшим числом отказов, активируя план размещения для этого состояния.

Система и её ВМ снабжены ресурсами, достаточными для выполнения задач как по начальному РмЗ, так и по каждому плану D_r^ω . Все полученные планы D_r^ω хранятся в базе данных планов РмЗ и используются процедурами ПЗ/ОУ. При переходе системы в любое из состояний s_r^ω формируется запрос к этой базе данных и из неё извлекается план РмЗ для данного состояния, в соответствии с которым происходит выполнение.

Кластерная вычислительная система рассматривается как клиент-серверная система, в которой главный узел C выполняет функции клиента, а множество вычислительных узлов N образует серверный массив. Взаимодействие между узлами выполняется по HTTP в формате JSON. Для каждой задачи U_j известны вес ω_j , отражающий приоритет выполнения, и время t_j , определяемое по результатам тестирования. В задачах ИТС такой вес позволяет отдавать приоритет обработке событий, связанных с ДТП, опасными нарушениями и другими критическими ситуациями.

Состояние узла n_i характеризуется статусом $\text{status}(n_i)$, суммарным временем уже назначенных задач $T_{\text{зат.}}(n_i)$ и множеством обработчиков $\text{task}_{\text{pod}}(n_i)$, загруженных в память узла. Главный узел хранит глобальное

состояние системы, обновляет его по ответам узлов, обнаруживает работоспособные узлы, сортирует задачи по убыванию веса, сортирует доступные узлы по возрастанию $T_{\text{зат.}}(n_i)$, назначает задачи и сохраняет результаты выполнения в локальной базе данных.

Рассмотрим кластерную вычислительную систему в начальный момент времени сессии решения задач. Система состоит из множества вычислительных узлов: $N = \{n_1, n_2, \dots, n_N\}$, где каждый узел $n_i \in N$ характеризуется своим состоянием. В рамках данной процедуры узлы считаются идентичными по вычислительной мощности.

Пусть A - множество задач, предназначенных для решения в текущей сессии, где $A \subseteq W$. Тогда $T_A(s_v)$ - необходимое время для выполнения заданного набора задач A при распределении, определяемом принятым алгоритмом, когда система находится в состоянии s_v . Получим, что T_A^0 - наименьшее достижимое время выполнения набора задач A в состоянии s^0 : $T_A^0 = \min\{T_A(s^0)\}$.

Также для системы задаётся максимально допустимое время выполнения заданного набора задач A , обозначаемое как $T_A^* > T_A^0$. Эти значения определяются требованиями пользователя и спецификой приложения (ограничения на режим работы в реальном времени).

Областью работоспособности системы S является множество всех её состояний s_i , в которых время выполнения набора задач A не превышает T_A^* : $S = \{s_i \mid T_A(s_i) \leq T_A^*\}$.

Для отказоустойчивой системы S включает начальное состояние s^0 и множество искажённых состояний S^ω : $S = \{s^0\} \cup S^\omega$, где $S^\omega = \{s_v \neq s^0 \mid T_A(s_v) \leq T_A^*\}$.

Событие считается отказом кластерной вычислительной системы, если в результате отказов узлов система переходит в состояние s_v , такое что $T_A(s_v) > T_A^*$, т.е. $s_v \notin S$.

Формальное описание метода динамического распределения задач с приоритетами.

1. **Шаг 1. Инициализация.** Формирование списка задач A для сессии.
2. **Шаг 2. Отсортировать список задач A по убыванию весов ω_j ,** получив упорядоченный список U , где $\omega_{j_1} \geq \omega_{j_2} \geq \dots \geq \omega_{j_{|A|}}$.
3. **Шаг 3. Цикл распределения задач (пока $U \neq \emptyset$):**
 - а) Обновить информацию о состоянии всех узлов N . Состояние каждого узла n_i включает его текущий статус (работоспособен/отказавший) и суммарное затраченное им время.
 - б) Отфильтровать отказавшие узлы.
 - в) Определить множество работоспособных узлов N_{work} .
 - г) Отсортировать множество N_{work} по возрастанию затраченного времени $T_{\text{зат.}}(n_i)$, получив упорядоченный

список $M = (n_{i_1}, n_{i_2}, \dots, n_{i_{|N_{\text{work}}|}})$, где $T_{\text{зат.}}(n_{i_1}) \leq T_{\text{зат.}}(n_{i_2}) \leq \dots \leq T_{\text{зат.}}(n_{i_{|N_{\text{work}}|}})$.

- д) Назначить первую задачу из списка U на выполнение первому узлу n_{i_1} из списка M .

4. Шаг 4. Завершить при $U = \emptyset$. Задачи распределены по узлам в соответствии с приоритетами важности и текущей загруженности.

В начальном состоянии s^0 распределение задач может выполняться с помощью любого алгоритма, направленного на минимизацию времени выполнения заданного набора задач и позволяющего обеспечить «наилучшее» время T_A^0 , в частности с помощью алгоритма, выполняющего распределение задач на основе затрачиваемого процессорного времени и соответствующей сортировки задач и узлов по этому критерию, как это предусмотрено в данном алгоритме. Это даёт возможность распределить задачи таким образом, чтобы затрачиваемое каждым узлом процессорное время было примерно одинаковым. Такой подход, по сравнению со случайным распределением задач по узлам, обеспечивает меньшее общее время выполнения всех задач.

Для проверки метода использовалась КВС из 8 узлов и три набора задач обработки данных с камер наблюдения, включающие 800, 500 и 300 задач. Рассматривались четыре сценария: случайное распределение при работоспособных узлах, распределение по разработанной логике при работоспособных узлах, случайное распределение при постепенном отказе четырёх узлов и распределение по разработанной логике при постепенном отказе четырёх узлов. В качестве порогового времени отказа принималось время выполнения задач при случайном распределении без отказов, после достижения которого оставшиеся задачи отбрасывались.

При 800 задачах разработанный метод без отказов завершил обработку за 42 секунды вместо 96 секунд при случайном распределении. При отказах случайное распределение привело к отбрасыванию 322 задач, а разработанный метод сократил это число до 37. При 500 задачах время без отказов сократилось с 66 до 29 секунд, а число отброшенных задач при отказах уменьшилось со 126 до 12. При 300 задачах время без отказов сократилось с 36 до 20 секунд, а при отказах разработанный метод завершил все задачи за 34 секунды. В среднем метод обеспечил ускорение примерно на 38% и повысил долю выполненных задач с максимальным весом при отказах вычислительных узлов.

Kubernetes и Apache Mesos предназначены для универсальной оркестрации разнородных приложений, масштабирования сервисов и управления сложными микросервисными средами. Разработанный метод не конкурирует с ними в этой области, а ориентирован на специализированный класс задач ИТС, где требуется быстрое выполнение большого числа независимых задач с разными весами при отказах узлов и минимальных накладных расходах. Его целевая область применения включает небольшие кластеры,

пакетную и потоковую обработку данных, а также сценарии, в которых приоритет отдельных задач должен учитываться непосредственно на уровне распределения.

Для проверки инфраструктурного решения реализован программный интерфейс на основе REST API с передачей данных по HTTP/HTTPS. В модели данных используются классы CarInfo и CoordInfo, описывающие идентификатор транспортного средства, номер, скорость, координаты, направление движения и процедуру обновления координат. Интерфейс предоставляется узлами auth и tables, а маршрутизация внешних запросов выполняется через ingress.

Доступ пользователей обеспечивается через токен аутентификации, а добавление записей сервисами источниками данных выполняется по валидному apikey. Пользователь получает доступ к созданию таблиц, созданию источников данных, просмотру и редактированию своих таблиц, а также поиску собранных метрик. Отправка метрик реализована асинхронно через очередь. После приёма метрики сервер возвращает статус доступа, затем обработчик очереди проверяет структуру записи и сохраняет её в соответствующем экземпляре таблицы.

Тестирование системы мониторинга включало два сценария: обработку нагрузки на хранилище данных и обработку нагрузки на обработчиках очереди. В первом сценарии использовался сервис источник, генерирующий синтетические данные о состоянии транспортного потока. На одном узле с четырьмя шардами таблицы нагрузка составляла около 650 операций чтения в секунду. В кластере из трёх узлов со структурой по четыре шарда на узел средняя нагрузка снизилась до 181 операции чтения в секунду на узел. После увеличения коэффициента репликации нагрузка распределялась по всем узлам и в среднем составляла 129 операций чтения в секунду. Это подтверждает, что горизонтальное масштабирование снижает нагрузку на отдельный узел, пока число узлов не достигает числа сегментов таблицы.

Во втором сценарии измерялась скорость полной обработки записи от поступления на сервер до сохранения в базе. Тестовые данные имитировали метрики дорожного движения. Средняя скорость обработки составила 0,2 записи в секунду, что связано с ожиданием записи в очереди. После извлечения из очереди одна запись обрабатывалась примерно за 100 мс. Увеличение числа реплик обработчиков повышало общую производительность потока, но производительность переставала расти, когда число реплик обработчиков равнялось количеству партиций в очереди.

В **заключении** приведены основные результаты работы, которые заключаются в следующем: В диссертационной работе получены следующие основные результаты, соответствующие поставленным задачам исследования.

1. Проведён анализ современного состояния интеллектуальных транспортных систем, рассматриваемых как сложные социотехнические распределённые системы с многослойной архитектурой и сквозными функциями управления. Установлено, что ключевыми ограничениями являются отсутствие единого формализованного аппарата синтеза архитектур эксплуатации моделей ИИ с учётом требований SLA, а также методологический разрыв между задачами прогнозирования, моделирования поведения и обеспечения эксплуатационной надёжности. Выявленные ограничения сформировали научную проблему и обосновали направление дальнейших исследований.
2. Разработан метод прогнозирования макроскопических характеристик транспортного потока PTSTG (PatchTST-Graph), основанный на декомпозиции временного и пространственного моделирования. Метод обеспечивает однопроходное формирование прогноза без авторегрессионного развёртывания, учитывает пространственные зависимости через обучаемую матрицу смежности и демонстрирует конкурентоспособную точность на общепринятых наборах данных при сниженной вычислительной сложности. Экспериментально подтверждена эффективность метода по метрикам MAE и RMSE на различных горизонтах прогнозирования.
3. Разработан метод прогнозирования микроскопических характеристик транспортного потока, основанный на гибридной архитектуре, сочетающей модели, основанные на данных, и классические физически интерпретируемые модели. Предложен гибридный метод на базе Bi-LSTM-LSTM-Informer и внешнего ограничителя на базе модели SSSR-Micro с динамической коррекцией параметров, обеспечивающая устойчивость при замкнутом применении и подавление аномальных предсказаний. Экспериментально показана устойчивость метода к накоплению ошибок и сдвигу распределения, а также его применимость в реальных условиях.
4. Разработана и формализована методология конфигурирования архитектуры исполнения моделей ИИ в интеллектуальных транспортных системах, основанная на представлении системы в виде ориентированного мультиграфа с параметризованными характеристиками рёбер и вершин. Методология обеспечивает синтез архитектурных конфигураций с учётом ограничений по задержке, пропускной способности и доступности, включает формальные модели оценки производительности и механизм верификации SLA в контуре непрерывной интеграции.
5. Предложена архитектура системы обработки больших данных для интеллектуальных транспортных систем, структурированная по функциональным зонам, включающим сбор, передачу, хранение,

обработку и визуализацию данных. Архитектура обеспечивает разделение ответственности компонентов, поддержку интеграции с внешними системами и возможность масштабируемого развёртывания в распределённых вычислительных средах.

6. Разработана и обоснована методология создания инфраструктуры интеллектуальных транспортных систем, ориентированная на отказоустойчивость, масштабируемость и адаптивность. Предложены методы децентрализованной кластеризации с использованием алгоритма консенсуса, динамического распределения задач с учётом приоритетов и генерации синтетических данных для стресс-тестирования. Проведена экспериментальная верификация, подтвердившая работоспособность и эффективность предложенных решений в условиях нестационарной нагрузки.

Полученные результаты в совокупности формируют целостную методологическую и технологическую основу для построения интеллектуальных транспортных систем, обеспечивающих согласованное решение задач прогнозирования, моделирования поведения и эксплуатации моделей машинного обучения в условиях реального времени и нестационарной нагрузки.

Основные публикации по теме диссертации

Работы, опубликованные в изданиях из перечня ВАК РФ

1. *Полянцева, К. А.* Архитектура распределенной системы хранения и обработки больших данных на основе Apache Ozone и Argo Workflows [Текст] / К. А. Полянцева, А. В. Комлев, М. Г. Городничев // Известия Кабардино-Балкарского научного центра РАН. 2026. Т. 28, № 2. С. 34—50. (Перечень рецензируемых научных изданий №1328 от 17.02.2026).
2. *Городничев, М. Г.* Механизм управления алгоритмами построения витрин данных в интеллектуальных транспортных системах [Текст] / М. Г. Городничев, К. А. Полянцева, Н. А. Грязнов // Computational Nanotechnology. 2026. Т. 13, № 1. С. 175—185. (Перечень рецензируемых научных изданий №27 от 17.02.2026).
3. *Иевлев, К. О.* Сравнительный анализ открытых табличных форматов: Apache Iceberg, Apache Hudi, Delta Lake, Apache Paimon [Текст] / К. О. Иевлев, М. Г. Городничев // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ, управление. 2026. № 1. С. 170—188. (Перечень рецензируемых научных изданий №167 от 25.11.2025).
4. *Городничев, М. Г.* О некоторых подходах повышения отказоустойчивости вычислительного кластера интеллектуальной транспортной системы с учетом приоритетов выполнения задач [Текст] / М. Г. Городничев // Модели, системы, сети в экономике, технике, природе и обществе. 2026. 1(XX). С. XX—XX. (Перечень рецензируемых научных изданий №1791 от 25.11.2025).
5. *Городничев, М. Г.* О применении обучения с подкреплением в задаче выбора оптимальной траектории движения [Текст] / М. Г. Городничев // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27, № 2. С. 86—102. (Перечень рецензируемых научных изданий №1319 от 05.02.2025).
6. *Городничев, М. Г.* О задаче применимости синтетических данных при тестировании интеллектуальных транспортных систем [Текст] / М. Г. Городничев // Computational

Nanotechnology. 2025. Т. 12, № 1. С. 105—115. (Перечень рецензируемых научных изданий №27 от 09.12.2024).

7. *Иевлев, К. О.* Сравнительный анализ систем хранения данных HDFS и Apache Ozone [Текст] / К. О. Иевлев, М. Г. Городничев // Computational Nanotechnology. 2025. Т. 12, № 1. С. 26—33. (Перечень рецензируемых научных изданий №27 от 09.12.2024).

8. О задаче построения децентрализованной интеллектуальной транспортной системы на основе протокола Raft и кластеризации по сетевому расстоянию [Текст] / М. Г. Городничев [и др.] // Вестник Рязанского государственного радиотехнического университета. 2025. № 94. С. 59—67. (Перечень рецензируемых научных изданий №735 от 25.11.2025).

9. *Городничев, М. Г.* О задаче построение независимых архитектур обработки данных в интеллектуальных транспортных системах [Текст] / М. Г. Городничев, Д. В. Титов, А. Д. Липатова // Инженерный вестник Дона. 2025. 11(131). С. 62—92. (Перечень рецензируемых научных изданий №1398 от 01.07.2025).

10. *Харрасов, К. Р.* Исследование устойчивости нейронных сетей в задаче распознавания образов [Текст] / К. Р. Харрасов, М. С. Мосева, М. Г. Городничев // Информационные технологии. 2025. Т. 31, № 2. С. 87—92. (Перечень рецензируемых научных изданий №1459 от 09.12.2024).

Статьи, индексируемые в международных базах WoS/Scopus

11. Benchmarking Compact VLMs for Clip-Level Surveillance Anomaly Detection Under Weak Supervision [Текст] / K. Borodin [и др.] // J. Imaging. 2025. Т. 11. С. 400. (Scopus Q1).

12. *Mkrtchian, G.* Patch-Based Transformer-Graph Framework (PTSTG) for Traffic Forecasting in Transportation Systems [Текст] / G. Mkrtchian, M. Gorodnichev // Applied Sciences (Switzerland). 2025. Т. 15, № 19. С. 10468. (Scopus Q1).

13. Development of Software Module for Recognizing Traffic Flows Through Deep Learning [Текст] / M. Gorodnichev [и др.] // Journal of Industrial Integration and Management: Innovation and Entrepreneurship. 2023. Т. 08, № 02. С. 175—199. (Scopus Q1).

14. On the Problem of Restoring and Classifying a 3D Object in Creating a Simulator of a Realistic Urban Environment [Текст] / M. Gorodnichev [и др.] // Sensors. 2022. Т. 22, № 14. С. 5199. (Scopus Q1).

15. End-to-End Train Horn Detection for Railway Transit Safety [Текст] / V.-T. Tran [и др.] // Sensors. 2022. Т. 22. С. 4453. (Scopus Q1).

16. *Gorodnichev, M.* Simulation model of traffic flow based on agent-based modeling [Текст] / M. Gorodnichev // Heritage and Sustainable Development. 2022. Т. 4, № 2. С. 195—190. (Scopus Q1).

17. *Moseva, M.* Robust Neural Network-Based Object Detection Method for Drone Images [Текст] / M. Moseva, K. Kharrasov, M. Gorodnichev // International Journal of Artificial Intelligence. 2024. Т. 22, № 1. С. 1—6. (Scopus Q2).

18. *Gorodnichev, M.* A Road Accident Detection Method Utilizing Deep Learning and Fast Fourier Transform [Текст] / M. Gorodnichev, K. Kharrasov, M. Moseva // Engineering, Technology and Applied Science Research. 2025. Т. 15, № 6. С. 30191—30195. (Scopus Q2).

19. *Gorodnichev, M.* On the Problem of Interpretability of Neural Algorithms for Predicting Microscopic Traffic Flow Data [Текст] / M. Gorodnichev // 2026 Systems of Signals Generating and Processing in the Field of on Board Communications. IEEE. 2026. С. 1—6.

20. *Gorodnichev, M.* On the Problem of Traffic Forecasting Using Graph Neural Networks and Dynamic Adjacency Matrix [Текст] / M. Gorodnichev // 2025 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED). IEEE. 2025. С. 1—4.

21. On the Automated Selection of Neural Network Architectures in the Task of Pattern Recognition for Intelligent Transportation Systems [Текст] / A. Moseva [и др.] // 2025

- Systems of Signals Generating and Processing in the Field of on Board Communications. IEEE. Moscow, Russian Federation, 2025. C. 1—5.
22. *Gorodnichev, M.* On the Applicability of Reinforcement Learning in the Traffic Flow Control Problem [Текст] / М. Gorodnichev // 2025 Systems of Signals Generating and Processing in the Field of on Board Communications. IEEE. Moscow, Russian Federation, 2025. C. 1—6.
 23. *Gorodnichev, M.* On the Task of Designing Efficient Machine Learning Operations Architectures [Текст] / М. Gorodnichev // 2025 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). IEEE. Tyumen, Russian Federation, 2025. C. 1—5.
 24. On the Task of Developing Algorithms for Generating Synthetic Data for Testing Intelligent Transportation Systems [Текст] / А. Lipatova [и др.] // Systems of Signal Synchronization, Generating and Processing in Telecommunications. Т. 7. IEEE. 2024. C. 267—271.
 25. Research on Driver Monitoring Methods [Текст] / М. Gorodnichev [и др.] // Systems of Signals Generating and Processing in the Field of on Board Communications. Т. 7. IEEE. 2024. C. 123—127.
 26. *Gorodnichev, M.* On the problem of developing a fault-tolerant high-loaded cluster of support for an intelligent transportation system [Текст] / М. Gorodnichev, М. Moseva // Wave Electronics and Its Application in Information and Telecommunication Systems. 2023. Т. 6, № 1. C. 122—127.
 27. On the problem of developing a realistic road infrastructure simulator for reinforced learning [Текст] / К. Polyantseva [и др.] // Wave Electronics and Its Application in Information and Telecommunication Systems. 2023. Т. 6, № 1. C. 340—347.
 28. *Gorodnichev, M.* On the applicability of reinforced learning in the route selection task of an unmanned vehicle [Текст] / М. Gorodnichev // Systems of Signal Synchronization, Generating and Processing in Telecommunications. Т. 6. IEEE. 2023. C. 107—115.
 29. *Polyantseva, K.* Ensuring the Reliability of a Highly Loaded Vehicle Monitoring and Traffic Control Platform [Текст] / К. Polyantseva, М. Gorodnichev, М. Moseva // Systems of Signals Generating and Processing in the Field of on Board Communications. Т. 6. IEEE. 2023. C. 365—372.
 30. Providing Fault Tolerance of Cluster Computing Systems Based on Fault-Tolerant Dynamic Computation Planning [Текст] / Е. Tourouta [и др.] // Lecture Notes in Information Systems and Organisation. 2022. C. 143—150.
 31. *Tourouta, E.* Ensuring Fault Tolerance of a Computing System with a Client-Server Architecture Based on Fault-Tolerant Computing Planning [Текст] / Е. Tourouta, М. Gorodnichev, К. Polyantseva // Wave Electronics and Its Application in Information and Telecommunication Systems. 2022. Т. 5, № 1. C. 496—500.
 32. Intelligent System for Recognizing the Main Characteristics of the Traffic Flow [Текст] / М. Gorodnichev [и др.] // 2021 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED). IEEE. 2021.
 33. Development of a system for collecting and processing big data received from terminal devices [Текст] / М. Gorodnichev [и др.] // 2021 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). IEEE. 2021. C. 9488337.
 34. Development of a module for controlling and predicting the movement of unmanned vehicles [Текст] / М. Gromov [и др.] // 2021 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SYNCHROINFO). IEEE. 2021. C. 9488374.
 35. *Degtyareva, V.* Study of Machine Learning Techniques for Transport [Текст] / V. Degtyareva, М. Gorodnichev, М. Moseva // Smart Technologies for Society, State and Economy. 2021. C. 1585—1595.

36. *Gorodnichev, M.* Architecture development for a high-load vehicle tracking application [Текст] / М. Gorodnichev, М. Moseva, D. Belousov // 2021 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE. 2021. С. 9470576.
37. Development of a Platform for Road Infrastructure Digital Certification [Текст] / М. Moseva [и др.] // 2021 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED). IEEE. 2021.
38. On Automated Safety Distance Monitoring Methods by Stereo Cameras [Текст] / М. Gorodnichev [и др.] // 2020 Systems of Signals Generating and Processing in the Field of on Board Communications. IEEE. 2020. С. 9078616.
39. The study of the applicability of clustering methods to identify patterns in traffic flows [Текст] / М. Gorodnichev [и др.] // 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE. 2020. С. 9131435.
40. *Menshih, P.* Efficiency analysis of neural networks ensembles using synthetic data [Текст] / P. Menshih, S. Erokhin, М. Gorodnichev // 2020 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE. 2020. С. 9131495.
41. On methods for recognizing unstructured road maintenance objects using deep learning algorithms [Текст] / М. Gorodnichev [и др.] // 2020 International Conference on Engineering Management of Communication and Technology (EMCTECH). IEEE. 2020. С. 9261568.
42. On the task of recognizing road infrastructure on devices with low computing power [Текст] / М. Moseva [и др.] // 2020 International Conference on Engineering Management of Communication and Technology (EMCTECH). IEEE. 2020. С. 9261559.
43. Neural network methods in real-time recognition of road infrastructure objects [Текст] / М. Gorodnichev [и др.] // IOP Conference Series: Materials Science and Engineering. 2020. Т. 918. С. 012077.
44. *Gorodnichev, M.* Development of Platform for Confirming and Storing Supply Data Using Blockchain Technology [Текст] / М. Gorodnichev, А. Nazarova, М. Moseva // 2019 IEEE International Conference on Quality Management, Transport and Information Security (IT&QM&IS). IEEE. 2019. С. 182—185.
45. *Gorodnichev, M.* Informational Aspects of Automated Monitoring of Dangerous Driving [Текст] / М. Gorodnichev, K. Polyantseva, М. Yashina // 2019 Systems of Signals Generating and Processing in the Field of on Board Communications (SOSG). IEEE. 2019. С. 8706721.
46. On the Applicability of Neural Networks in the Tasks of Detecting Dangerous Movement [Текст] / K. Polyantseva [и др.] // 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF). IEEE. 2019. С. 8840654.
47. The Research of Blockchain Technology for Data Protection in IoT Devices [Текст] / М. Gorodnichev [и др.] // International Conference Technology & Entrepreneurship in Digital Society (TEDS). 2019. С. 94—97.
48. Features of the development of information systems for working with blockchain technology [Текст] / М. Gorodnichev [и др.] // Journal of Physics: Conference Series. 2019. Т. 1399. С. 033039.

Свидетельства о государственной регистрации программ для ЭВМ

49. Свидетельство о гос. регистрации программы для ЭВМ. Интеллектуальная платформа адаптивного управления светофорами [Текст] / Городничев М. Г., Латыпов Э. М., Матюнина Д. Д., Полянцева К. А. ; Ордена Трудового Красного Знамени федерального государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 22.01.2026, № 2026611997.

50. Свидетельство о гос. регистрации программы для ЭВМ. Программная реализация метода моделирования транспортных потоков с учетом физических характеристик одиночных транспортных средств [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 07.05.2025, № 2025661566.
51. Свидетельство о гос. регистрации программы для ЭВМ. Программное обеспечение генерации синтетических данных о состоянии транспортного потока [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 22.05.2025, № 2025662580.
52. Свидетельство о гос. регистрации программы для ЭВМ. Автоматизированная информационная система распознавания и восстановления объектов дорожной инфраструктуры [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 23.05.2025, № 2025662817.
53. Свидетельство о гос. регистрации программы для ЭВМ. Программный модуль динамического распределения ресурсов в подсистемах интеллектуальных транспортных систем мониторинга и управления [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 27.05.2025, № 2025663312.
54. Свидетельство о гос. регистрации программы для ЭВМ. Программная реализация метода поиска нейросетевых алгоритмов управления технологическим процессом на транспорте [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 27.05.2025, № 2025663314.
55. Свидетельство о гос. регистрации программы для ЭВМ. Автоматизированная информационная система поддержки функционирования интеллектуальной транспортной системы с учетом формата обрабатываемой и хранимой информации [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 02.06.2025, № 2025664002.
56. Свидетельство о гос. регистрации программы для ЭВМ. Программная реализация архитектуры, обеспечивающей информационный процесс передачи информации между одиночными транспортными средствами (V2V) [Текст] / Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 09.06.2025, № 2025664922.
57. Свидетельство о гос. регистрации программы для ЭВМ. RouteGenius — интеллектуальная платформа для планирования маршрутов автотуристов [Текст] / Мкртчян Г. М., Павликов А. Е., Тимчук А. В., Мокин А. И., Городничев М. Г. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики». — Оpubл. 24.06.2025, № 2025666160.
58. Свидетельство о гос. регистрации программы для ЭВМ. Программный модуль определения основных характеристик транспортного потока (интенсивность, плотность, качественный состав) на основе видеопотока [Текст] / Городничев М. Г., Мосева М. С., Полянцева К. А. ; Общество с ограниченной ответственностью «Кибер Ай Кью». — Оpubл. 30.07.2021, № 2021662524.

59. Свидетельство о гос. регистрации программы для ЭВМ. Мобильное приложение многопараметрического сбора основных характеристик транспортных потоков под управлением ОС iOS [Текст] / Городничев М. Г., Мосева М. С., Полянцева К. А. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ). — Оpubл. 13.02.2020, № 2020612021.

60. Свидетельство о гос. регистрации программы для ЭВМ. Программная реализация иерархической кластеризации поведенческого фактора водителей методами машинного обучения [Текст] / Городничев М. Г., Мосева М. С. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ). — Оpubл. 25.02.2020, № 2020612467.

61. Свидетельство о гос. регистрации программы для ЭВМ. Программная платформа мониторинга и управления движением беспилотных транспортных средств [Текст] / Городничев М. Г., Мосева М. С., Полянцева К. А. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ). — Оpubл. 25.02.2020, № 2020612482.

62. Свидетельство о гос. регистрации программы для ЭВМ. Мобильное приложение многопараметрического сбора основных характеристик транспортных потоков под управлением ОС Android [Текст] / Городничев М. Г., Мосева М. С., Полянцева К. А. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ). — Оpubл. 25.02.2020, № 2020612483.

63. Свидетельство о гос. регистрации программы для ЭВМ. Программа автоматизированного определения скорости и габаритов транспортных средств на основании изображений, получаемых со стереокамер [Текст] / Городничев М. Г., Полянцева К. А., Гематудинов Р. А., Мосева М. С., Джабраилов Х. А. ; Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ). — Оpubл. 28.11.2019, № 2019665793.