

Παράδειγμα

Парадигма

Paradigma

№5, 2026

Часть 7

Електронно
научно
списание

Парадигма
Електронно
научно списание

БРОЙ 5/2026
Часть 7

Издател:
"ЦЕНТЪР ЗА НАУЧНИ
ИЗСЛЕДВАНИЯ И
ИНФОРМАЦИЯ
"ПАРАДИГМА"" ЕООД
БЪЛГАРИЯ, гр. Варна
9002,
р-н Одесос, ул.
Опълченска No 27
E-mail:
niiparadigma@gmail.com

www.paradigma.science

ISSN 2367-8658

Договори на размещение:
eLIBRARY.RU
CYBERLENINKA

Публикационна политика:



Редакционен съвет

Абакаров Д. К., к.с.н. (г. Брянск, Россия); **Анжиганова Л. В.**, д.ф.н., профессор (г. Абакан, Россия); **Антамошкин А. Н.**, д.т.н., профессор (г. Красноярск, Россия); **Арпентьева М. Р.**, д. психол. наук, доцент, член-корреспондент РАЕ академик МАЕ (г. Калуга, Россия); **Багоцкий С. В.**, к.б.н., доцент МИОО, ученый секретарь Московского общества испытателей природы (г. Москва, Россия); **Белобрыкина О. А.**, к.психол.н., доцент, академик Академии полярной медицины и экстремальной экологии человека (г. Новосибирск, Россия); **Бобкова Е. Ю.**, к.пед.н., доцент (г. Самара, Россия); **Валигова И. Е.**, к.пс.н., (г. Брест, Республика Беларусь); **Галкина А. И.**, с.н.с., начальник отдела ФГБНУ "ИУО РАО", руководитель Объединенного фонда электронных ресурсов «Наука и образование», почетный работник науки и техники РФ (г. Москва, Россия); **Галчева К. Б.**, доцент, доктор по педагогика Пловдивски университет „П.Хилендарски“ (г. Пловдив, Республика България); **Заславский А. А.**, к. пед. наук, доцент, (г. Москва, Россия); **Заславская О. Ю.**, д.пед.н., профессор; **Землянухина Н. С.**, д.э.н., профессор (г. Саратов, Россия); **Землянухина С. Г.**, д.э.н., профессор (г. Саратов, Россия); **Ищанова Г. Т.**, к.э.н., (г. Алматы, Казахстан); **Капрусова М. Н.**, к.ф.н., доцент (г. Борисоглебск, Россия); **Костригин А. А.** (г. Нижний Новгород, Россия); **Кошенова М. И.**, к.пс.н., доцент, зав.каф. социальной психологии и виктимологии (г. Новосибирск, Россия); **Кравец О. Я.**, д.т.н., профессор (г. Воронеж, Россия); **Магсумов Т. А.**, к.и.н., доцент (г. Набережные Челны, Россия); **Няголова М. Д.**, канд. психол. наук, доцент истории психологии Великотърновского университета имени Святых Кирилла и Мефодия (г. Велико Търново, Республика България); **Останков А. В.**, д.т.н., профессор (г. Воронеж, Россия); **Перова М. Б.**, д.э.н., профессор (г. Вологда, Россия); **Поляков Ю. А.**, к.т.н., доцент, (г. Москва, Россия); **Садчиков А. П.**, д.б.н., профессор Международного биотехнологического центра МГУ имени М.В.Ломоносова, вице-президент Московского общества испытателей природы (г. Москва, Россия); **Саенко Л. В.**, к.ю.н., доцент (г. Саратов, Россия); **Седов В. А.**, к.ф.-м.н., доцент, заведующий кафедрой теоретических основ электротехники (г. Владивосток, Россия); **Седова Н. А.**, к.т.н., доцент (г. Владивосток, Россия); **Семенютина А. В.**, д.с.-х.н., зав. отделом биологии древесных растений ВНИАЛМИ (г. Волгоград, Россия); **Сидоровнин Г. П.**, директор Европейского Стольпинского инфоцентра (г. Майнц, Германия); **Соловьева А. Г.**, к.б.н., профессор РАЕ, с.н.с. (г. Нижний Новгород, Россия); **Суркова И. Ю.**, д. социол. н., доцент (г. Саратов, Россия); **Трендафилова А. Т.**, ассистент Факультета общественного здоровья Медицинский университет-София (г. София, республика България); **Фурсов А. Л.**, к.э.н., директор научно-исследовательского института «Парадигма» (г. Саратов, Россия); **Хусяинов Т. М.** (г. Нижний Новгород, Россия)

Главен редактор

Андрей Фурсов

ТЕХНОЛОГИИ

УДК 004.912:004.8:61

Шамрин М. Д.

МИРЭА – Российский технологический университет, Москва, Россия

РАСПОЗНАВАНИЕ ПЕРСОНАЛЬНЫХ ДАННЫХ ДЛЯ ОБЕЗЛИЧИВАНИЯ РУССКОЯЗЫЧНЫХ МЕДИЦИНСКИХ ДОКУМЕНТОВ С ПРИМЕНЕНИЕМ МОДЕЛИ АРХИТЕКТУРЫ BERT

Аннотация: в данной статье рассматривается проблема автоматизированного обезличивания неструктурированных медицинских документов на русском языке. Предложен гибридный подход, комбинирующий нейросетевую модель архитектуры трансформер для задачи распознавания именованных сущностей (NER) и алгоритмы на основе регулярных выражений для извлечения строго формализованных атрибутов. Для обучения моделей сформирован набор данных объемом 2500 клинических записей, включающий 62 627 размеченных сущностей по 24 категориям. Экспериментальная оценка показала, что разработанная гибридная модель достигает показателя F1-меры 95.21%. Доказано преимущество гибридного подхода перед изолированными методами извлечения информации в медицинских документах.

Ключевые слова: NER, обезличивание, персональные данные, медицинские документы, глубокое обучение, NLP, трансформеры, BERT, регулярные выражения.

Введение. В условиях цифровизации здравоохранения и повсеместного перехода на электронные медицинские карты (ЭМК), объемы текстовой информации в виде клинических заметок и выписных эпикризов растут в геометрической прогрессии. Данные массивы представляют высокую ценность для медицинских исследований и обучения систем искусственного интеллекта. Однако законодательство Российской Федерации (ФЗ-152 «О персональных данных» и ФЗ-323 «Об основах охраны здоровья граждан») накладывает строгие ограничения на использование реальных данных пациентов, защищая врачебную тайну [1, 2]. Специфика электронных медицинских документов заключается в наличии не только прямых персональных данных (ПДн), но и квазиидентификаторов, комбинация которых может привести к повторной идентификации субъекта [3]. В связи с этим разработка локальных систем автоматизированного обезличивания неструктурированных медицинских текстов на русском языке является актуальной задачей.

Анализ существующих подходов. Процесс обезличивания текстовых данных в современных исследованиях рассматривается как процедура, состоящая из двух этапов:

1. Распознавание именованных сущностей (Named Entity Recognition, NER) — поиск и извлечение именованных сущностей в виде персональных данных в неструктурированном тексте.

2. Трансформация — применение алгоритмов изменения найденных сущностей для защиты конфиденциальности с сохранением полезности текста.

Методы извлечения персональных данных в научной литературе разделяются на следующие группы [4]:

1. Методы на основе правил.
2. Методы глубокого обучения.

Традиционные подходы к задаче обезличивания базируются на использовании регулярных выражений и словарей. Данные методы демонстрируют высокую точность (Precision) при поиске строго формализованных атрибутов (номера телефонов, СНИЛС, адреса электронной почты) [5]. Тем не менее, они обладают низкой гибкостью и не способны учитывать семантический контекст, что ведет к критическому снижению полноты (Recall) при наличии опечаток, сокращений или синонимов [6].

Современным стандартом в задачах обработки естественного языка (NLP) выступают модели на базе архитектуры трансформер, в частности, семейство BERT. За счет механизма внутреннего внимания (Self-Attention) они способны анализировать двунаправленный контекст [7,8]. Генеративные LLM-модели позволяют выполнять обезличивание через промпты без предварительного дообучения. Однако их применение в медицинских учреждениях ограничено следующими недостатками: облачные API нарушают требования законодательства о защите ПДн, локальное развертывание требует значительных вычислительных ресурсов, а склонность к галлюцинациям снижает надежность результатов [9]. Таким образом, для высокозащищенных медицинских контуров наиболее оптимальным выбором являются локально разворачиваемые гибридные архитектуры.

Общая архитектура предлагаемой системы. Для компенсации недостатков отдельных методов спроектирована гибридная архитектура конвейера обезличивания (рис. 1). Целевая архитектура разделяет процесс на этапы извлечения и структурной трансформации данных.

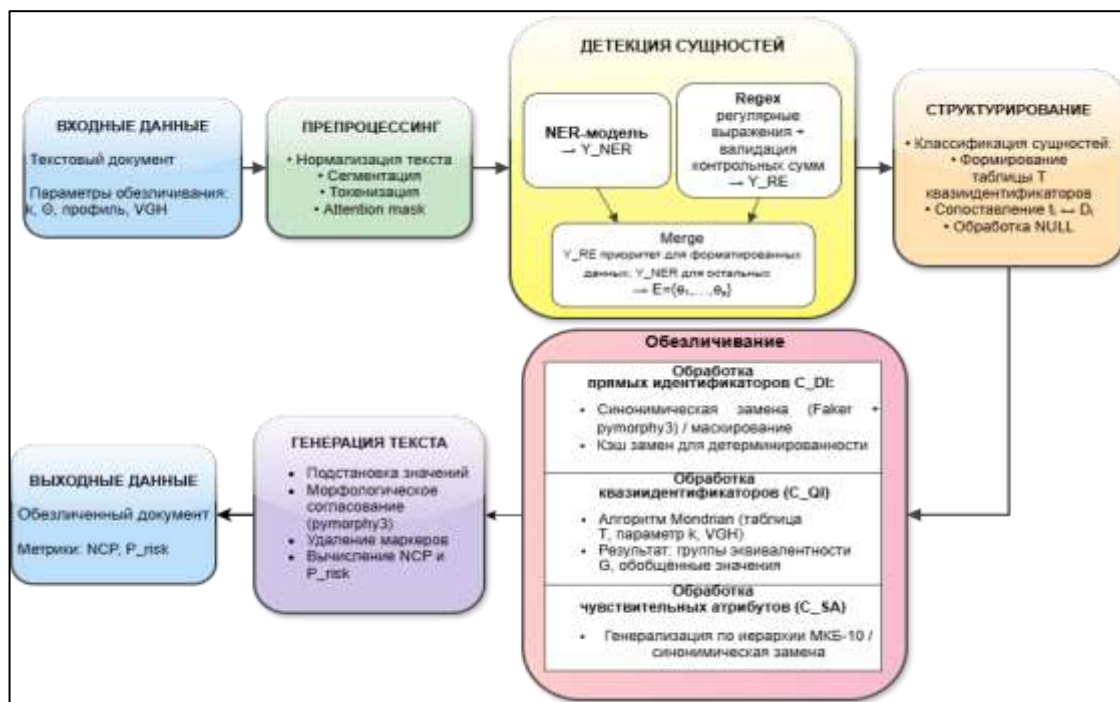


Рисунок 1. Целевая архитектура системы обработки и обезличивания документа

На вход системы поступает медицинский документ в неструктурированном текстовом формате. Процесс извлечения персональных данных распараллеливается на два модуля:

1. Модуль на основе правил выполняет поиск формализованных прямых идентификаторов (ИНН, паспорта, полисы ОМС) с применением оптимизированных регулярных выражений.

2. Нейросетевой модуль NER осуществляет контекстный поиск неструктурированных сущностей (ФИО, симптомы, диагнозы) с использованием модели глубокого обучения.

После извлечения сущностей результаты обеих подсистем объединяются. При наличии пересечений координат найденных токенов применяется алгоритм разрешения конфликтов: для строго формализованных сущностей приоритет отдается регулярным выражениям, в то время как для контекстно-зависимых сущностей более надежным считается решение нейросетевой модели.

Подготовка и разметка набора данных. Для обучения нейросетевой модели был сформирован набор данных, состоящий из 2500 медицинских документов. В качестве основы использовались набор данных амбулаторных записей RuMedPrime [10] и синтетические тексты, сгенерированные фреймворком MedSyn [11, 12]. Схема формирования набора данных представлена на рисунке 2.

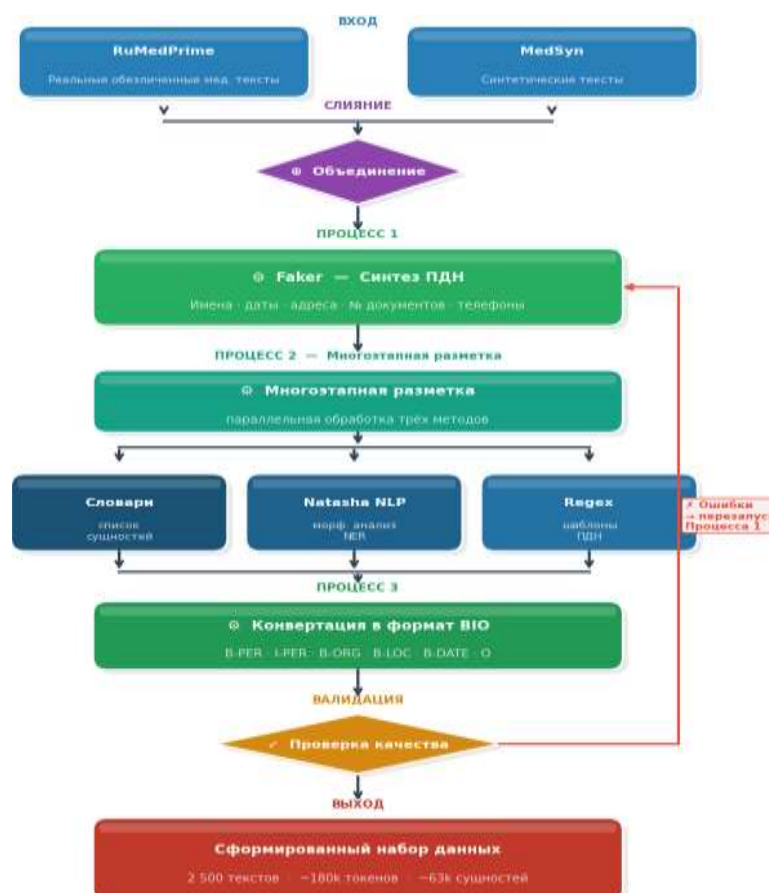


Рисунок 2. Процесс формирования и разметки набора данных

Разметка текстов осуществлялась в формате BIO, в которой первый токен каждой сущности маркируется специальной меткой B (Begin), а все последующие токены той же сущности — меткой I (Inside). Токены не принадлежащие сущности помечаются как O (Outside). В ходе классификации было выделено 24 категории обезличиваемых сущностей, разделенных на три группы: прямые идентификаторы, квазиидентификаторы и чувствительные медицинские атрибуты (симптомы, диагнозы). Общее количество размеченных сущностей в сформированном датасете составило 62627. Выборка характеризовалась естественным дисбалансом классов с преобладанием упоминаний медицинских симптомов, что отражает реальную структуру клинических записей. Распределение сущностей по классам представлено на рисунке 3.

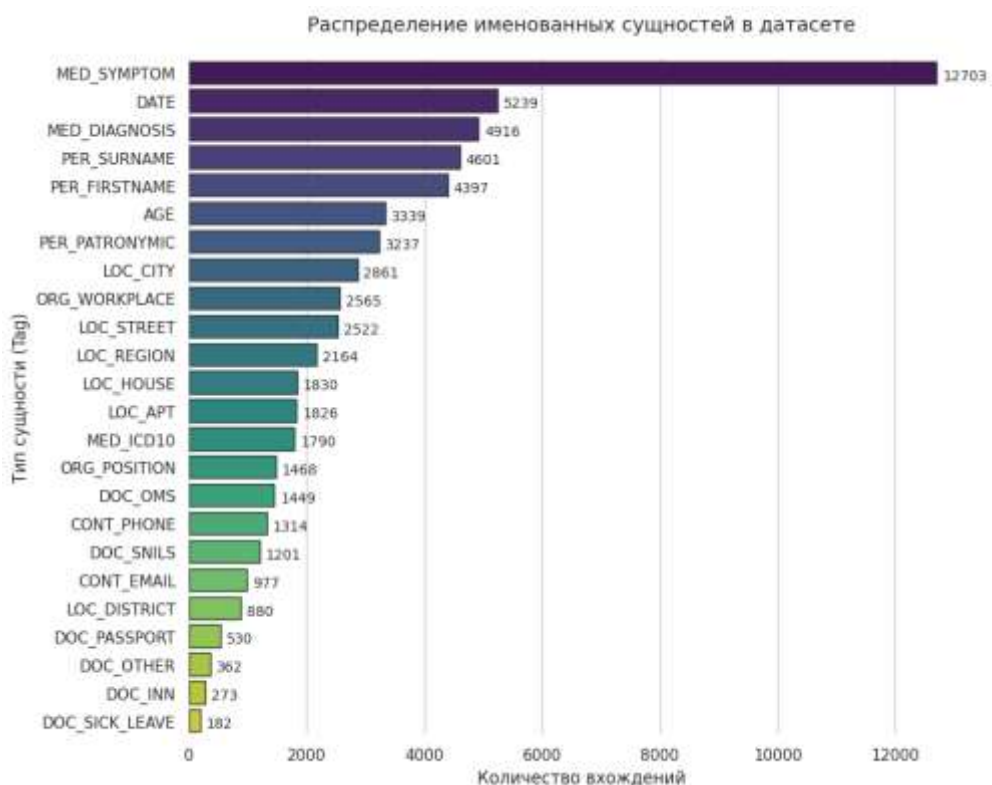


Рисунок 3. Распределение сущностей по классам в сформированном наборе данных

Обучение модели. В качестве нейросетевого ядра системы была выбрана предобученная русскоязычная модель ai-forever/ruBert-base [13]. Выбор обусловлен полнотой её словаря для русского языка, а также оптимальным балансом между точностью распознавания и требованиями к вычислительным ресурсам при локальном развертывании.

Дообучение модели проводилось на собранном наборе данных. Подготовленный набор данных был разделен в соотношении 80% — обучающая выборка, 10% — валидационная и 10% — тестовая выборка. Для выбора оптимальной конфигурации классификатора было проведено обучение двух вариантов модели: базовой (Baseline) и итоговой (Final). Целью экспериментов являлся подбор гиперпараметров, предотвращающих переобучение на медицинских текстах. Результаты сравнительного анализа представлены в таблице 1.

Таблица 1. Оценка влияния гиперпараметров на качество обучения

Параметр	Модель 1 (Baseline)	Модель 2 (Final)
Архитектура	ai-forever/ruBert-base	
Learning Rate	2e-5 (0,00002)	1e-5 (0,00001)
Количество эпох	5	15
Weight Decay	0.01	0.01

Параметр	Модель 1 (Baseline)	Модель 2 (Final)
Batch Size	16	16
Результаты	F1 = 90.60%	F1 = 92.38%

Как видно из таблицы 1, увеличение количества эпох до 15 в сочетании со снижением скорости обучения (Learning Rate) до $1e-5$ позволило улучшить качество классификации. Снижение шага оптимизатора обеспечило более плавную настройку скрытых весов классификатора. Анализ графика функции потерь (рис. 4) показал устойчивое снижение ошибки на обучающей и валидационной выборках без признаков переобучения, что подтверждает высокую способность модели к обобщению.

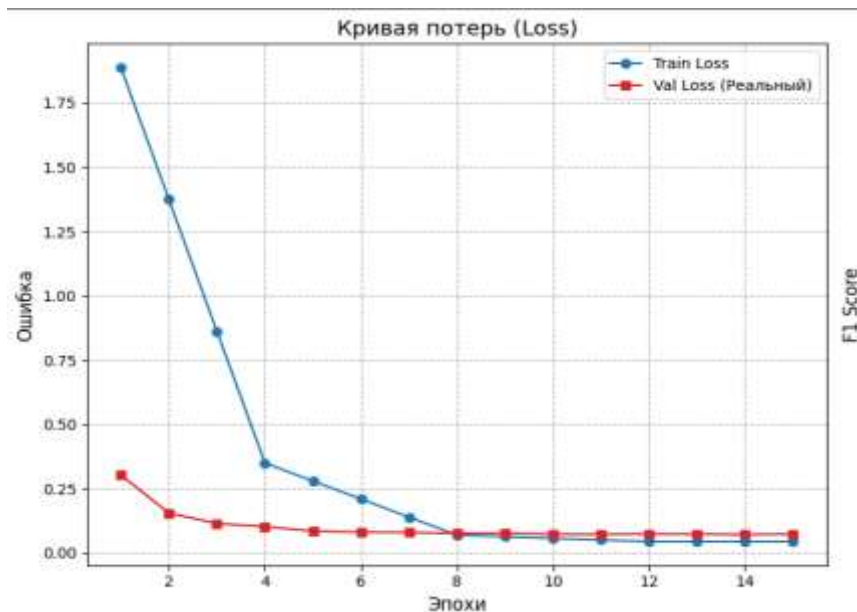


Рисунок 4. График функции потерь

Метрики оценки качества. Для оценки эффективности этапа распознавания именованных сущностей в работе использованы стандартные метрики классификации: точность (Precision), полнота (Recall) и F1-мера (F1-score). Расчет метрик производился на основе анализа следующих параметров:

- True Positive (TP) — количество ПДн, которые были верно найдены и классифицированы системой;
- False Positive (FP) — количество ошибочных срабатываний на словах, не являющихся чувствительными данными;
- False Negative (FN) — количество пропущенных системой персональных данных.

Формулы для расчета метрик имеют следующий вид:

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F_1 = \frac{2 \cdot P \cdot R}{P+R} \quad (1)$$

В контексте извлечения персональных данных приоритетной метрикой является Recall. Данный показатель отражает способность системы найти все вхождения чувствительных данных в тексте. Низкое значение Recall свидетельствует о наличии пропусков, что ведет к прямой утечке конфиденциальной информации и нарушению законодательства о защите ПДн.

Экспериментальная оценка и сравнение подходов. Для подтверждения эффективности комбинирования методов проведен сравнительный анализ трех подходов: изолированное использование регулярных выражений (Regex), изолированное применение нейросети (BERT) и предложенный гибридный подход, объединяющий Regex и BERT. Сводные метрики представлены в таблице 2.

Таблица 2. Сводные метрики качества разных подходов

Метрика \ Метод	Регулярные выражения	Модель BERT	Гибридный подход
Precision (точность)	0.799	0.934	0.934
Recall (полнота)	0.658	0.915	0.973
F1-мера	0.711	0.924	0.952

Сравнительная оценка выявила характерные особенности каждого метода. Использование исключительно регулярных выражений показало крайне низкую полноту (Recall = 0.658). Данный подход демонстрирует безошибочное извлечение структурированных данных (Precision = 1.0) для классов DOC_INN, DOC_PASSPORT, CONT_EMAIL (табл. 3). Однако данный метод не применим для поиска таких сущностей, как MED_SYMPTOM или PER_SURNAME, из-за высокой лингвистической вариативности.

Модель BERT существенно повысила базовое качество, однако внедрение гибридного подхода позволило увеличить полноту нахождения конфиденциальной информации до 0.973, что является важным показателем для задачи извлечения персональных данных.

Для сравнения работы методов по разным типам сущностей результаты представлены в таблице 3.

Таблица 3. Метрики качества по сущностям

Категория	Сущность	Pres	Rec.	F1	Pres	Rec.	F1	Pres	Rec.	F1
		Regex			Модель BERT			BERT и Regex		
Персональн ые данные	PER_SURNA ME	0.39 3	0.22 1	0.28 3	0.98 3	0.88 2	0.93 0	0.96 7	0.91 1	0.93 8
	PER_FIRSTN AME	0.41 3	0.22 3	0.29 0	0.98 8	0.95 1	0.96 9	0.96 6	0.97 1	0.96 9
	PER_PATRO NYMIC	0.41 2	0.27 5	0.33 0	1.00 0	0.98 2	0.99 1	0.99 1	1.00 0	0.99 6
	AGE	0.97 2	0.76 8	0.85 8	0.99 3	0.89 3	0.94 1	0.99 4	0.99 4	0.99 4
Контактны е данные	CONT_EMAI L	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
	CONT_PHON E	0.88 2	0.88 2	0.88 2	1.00 0	1.00 0	1.00 0	1.00 0	0.95 3	0.97 6
	DATE	0.98 8	0.99 0	0.98 9	0.97 2	0.99 6	0.98 4	0.94 8	1.00 0	0.97 3
Документы	DOC_INN	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
	DOC_OMS	1.00 0	1.00 0	1.00 0	0.99 2	0.80 1	0.88 6	0.99 3	1.00 0	0.99 7
	DOC_OTHER	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
	DOC_PASSP ORT	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
	DOC_SICK_L EAVE	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
	DOC_SNILS	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
Адреса	LOC_APT	1.00 0	0.54 4	0.70 5	1.00 0	0.96 7	0.98 3	1.00 0	1.00 0	1.00 0
	LOC_CITY	0.80 9	0.45 0	0.57 9	0.94 4	0.89 4	0.91 8	0.90 4	0.90 4	0.90 4
	LOC_DISTRI CT	0.00 0	0.00 0	0.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0	1.00 0
	LOC_HOUSE	1.00 0	0.54 6	0.70 7	1.00 0	0.93 4	0.96 6	1.00 0	0.99 5	0.99 7
	LOC_REGIO N	0.98 7	0.71 7	0.83 1	1.00 0	0.99 5	0.99 8	1.00 0	0.99 5	0.99 8
	LOC_STREET	0.08 1	0.06 7	0.07 3	1.00 0	0.93 7	0.96 7	0.98 4	0.99 2	0.98 8
Медицинск ие данные	MED_DIAGN OSIS	0.81 7	0.48 0	0.60 5	0.87 0	0.91 2	0.89 0	0.78 5	0.84 0	0.81 2
	MED_ICD10	0.99 5	1.00 0	0.99 7	1.00 0	0.98 9	0.99 5	0.98 9	1.00 0	0.99 5

Категория	Сущность	Pres	Rec.	F1	Pres	Rec.	F1	Pres	Rec.	F1
		Regex			Модель BERT			BERT и Regex		
	MED_SYMPTOM	0.977	0.992	0.985	0.771	0.791	0.781	0.839	1.000	0.912
Организации	ORG_POSITION	1.000	0.688	0.816	0.993	1.000	0.996	0.993	1.000	0.996
	ORG_WORKPLACE	1.000	0.629	0.772	0.992	0.996	0.994	0.988	0.996	0.992
ИТОГО	Weighted avg	0.799	0.658	0.711	0.934	0.915	0.924	0.934	0.973	0.952

Использование исключительно модели BERT позволяет достичь высоких показателей на неструктурированных данных, однако сохраняется вероятность пропуска (ошибка II рода) жестко заданных идентификаторов в нестандартных форматах записи. Модель практически безошибочно (F1-score > 0.99) идентифицирует сущности со строгой позиционной структурой: контакты (CONT_PHONE), должности (ORG_POSITION) и коды заболеваний (MED_ICD10). Наиболее низкий результат был зафиксирован для категории MED_SYMPTOM (F1-score = 0.7811), однако гибридный подход повышает показатель до 0.912. Это обусловлено высокой лингвистической сложностью описания жалоб пациентов в свободном тексте. В категории диагнозов (MED_DIAGNOSIS, F1-score = 0.8900) ошибки классификации преимущественно связаны с семантическим наложением, когда клинический термин совпадает с общеупотребительным словом. Также зафиксировано незначительное снижение полноты (Recall = 0.8820) при поиске фамилий (PER_SURNAME), что объясняется их высокой вариативностью и присутствием редких форм в тестовой выборке. Гибридный подход повышает данный показатель до 0.911. Сравнение эффективности подходов по макрокатегориям изображено на рисунке 5.

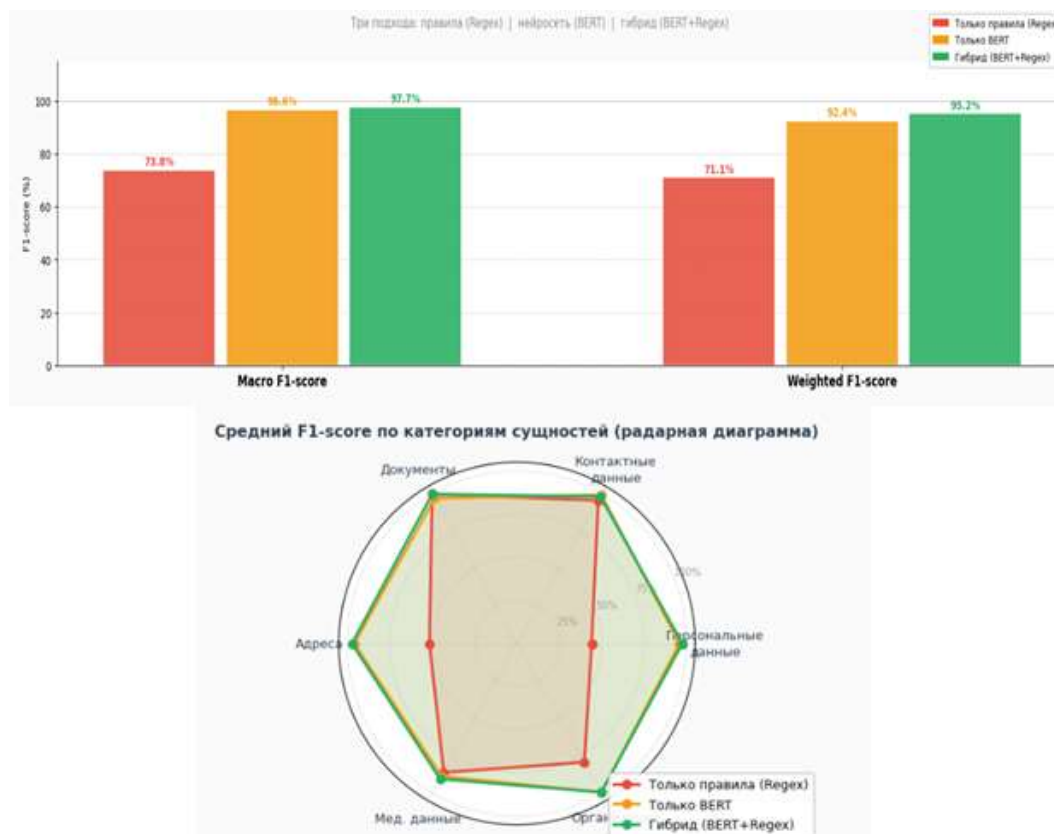


Рисунок 5. Сравнительный анализ извлечения персональных данных (F1-score, %)

Как видно из приведенной диаграммы (рис. 5), гибридный подход минимизирует данные недостатки. Комбинирование подходов обеспечивает максимальную полноту нахождения персональных данных ($Recall = 0,973$) и гарантирует стопроцентное маскирование шаблонных данных правилами Regex, оставляя ресурсы нейросети для выявления сложных семантических конструкций неструктурированных данных.

Для визуализации работы предложенного гибридного подхода на рисунке 6 представлен пример обработки фрагмента синтетического документа. Система успешно идентифицирует персональные данные, заменяя их соответствующими тегами, что позволяет сохранить структуру документа для последующего анализа.

НАПРАВЛЕНИЕ НА ЛАБОРАТОРНОЕ ИССЛЕДОВАНИЕ				НАПРАВЛЕНИЕ НА ЛАБОРАТОРНОЕ ИССЛЕДОВАНИЕ			
№ заказа	17145678	Фамилия пациента	Иванов	№ заказа	[DOC_OTHER]	Фамилия пациента	[PER_SURNAME]
Дата регистрации заказа	02.05.2025	Имя пациента	Иван	Дата регистрации заказа	[DATE]	Имя пациента	[PER_FIRSTNAME]
Пол пациента	мужской	Дата рождения пациента	01.02.2002	Пол пациента	мужской	Дата рождения пациента	[DATE]
Адрес пациента	Цветной бульв. д 33 в1 кв 110	№ страхового полиса	7834 5672 1234 5678	Адрес пациента	[LOC_STREET] [LOC_STREET] [LOC_HOUSE] [LOC_HOUSE] кв [LOC_APT]	№ страхового полиса	[DOC_OMS]
Номер истории болезни пациента		Направляющий врач	Котов А.С.	Номер истории болезни пациента		Направляющий врач	[PER_SURNAME] [PER_FIRSTNAME] [PER_PATRONYMIC]
Адрес	123456, г.Москва, мкр. Северный, ул. Мира, д. 5	Отделение		Адрес	[DOC_OTHER] г [LOC_CITY], [LOC_STREET], [LOC_STREET], [LOC_HOUSE]	Отделение	
	ООО «Клиника Здоровье»	Палата			[ORG_WORKPLACE]	Палата	
Назначенные исследования:				Назначенные исследования:			
<ul style="list-style-type: none"> 1. Общий анализ мочи 2. Общий анализ крови (с лейкоцитарной формулой) 3. Биохимический анализ крови (глюкоза, общий белок, АЛТ, АСТ) 4. Анализ крови на ВИЧ, гепатит В и С 				<ul style="list-style-type: none"> 1. Общий анализ мочи 2. Общий анализ крови (с лейкоцитарной формулой) 3. Биохимический анализ крови (глюкоза, общий белок, АЛТ, АСТ) 4. Анализ крови на ВИЧ, гепатит В и С 			
Краткие клинические данные:				Краткие клинические данные:			
<p>Пациент Иванов Иван Иванович, 24 год, обратился с жалобами на частое мочеиспускание, боли в поясничной области. Температура тела 37,2°C. Состояние удовлетворительное. Контактный телефон: +7 (916) 123-45-67. Направлен из ООО «Клиника Здоровье», расположенной по адресу: г. Москва, мкр. Северный, ул. Мира, д. 5. Лечащий врач — Петрова Ирина Владимировна.</p>				<p>Пациент [PER_SURNAME] [PER_FIRSTNAME] [PER_PATRONYMIC], [AGE] год, обратился с жалобами на [MED_SYMPTOM], [MED_SYMPTOM]. Температура тела 37,2°C. Состояние удовлетворительное. Контактный телефон: [CONT_PHONE]. Направлен из [ORG_WORKPLACE], расположенной по адресу: [LOC_REGION], [LOC_DISTRICT], [LOC_STREET], д. [LOC_HOUSE]. Лечащий врач — [PER_SURNAME] [PER_FIRSTNAME] [PER_PATRONYMIC].</p>			
Врач: Петрова Ирина Владимировна				Врач: [PER_SURNAME] [PER_FIRSTNAME] [PER_PATRONYMIC]			
Дата: 02.05.2025				Дата: [DATE]			
М.П. Подпись: _____				М.П. Подпись: _____			

Рисунок 6. Пример обработки фрагмента медицинского документа

Заключение. В ходе исследования разработана и реализована локальная гибридная система автоматизированного извлечения чувствительных данных из медицинских текстов на русском языке. Применение нейросетевой модели BERT в связке с алгоритмами на основе правил позволило обеспечить высокую эффективность распознавания ПДн (F1-score = 95.21%).

Перспективным направлением развития системы является реализация различных методов трансформации извлеченных персональных данных, в частности, замена синтетическими данными и внедрение математической модели k-анонимности, например, алгоритма Mondrian [14, 15]. А использование древовидных структур для иерархического обобщения извлеченных квазиидентификаторов [9] (адреса, даты, диагнозы) позволит математически минимизировать риск реидентификации пациентов при сохранении аналитической ценности данных.

Список литературы:

1. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных» [Электронный ресурс]. URL: <http://pravo.gov.ru/proxy/ips/?docbody&nd=102108261> (дата обращения: 06.02.2026)
2. Федеральный закон от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации». [Электронный ресурс]. URL: <https://minzdrav.gov.ru/documents/7025> (дата обращения: 06.02.2024).

3. Sweeney, L. Simple Demographics Often Identify People Uniquely / L. Sweeney // *Health (San Francisco)*. – 2000. – Vol. 671. – P. 1-34.
4. Negash, B. De-identification of free text data containing personal health information: a scoping review of reviews / B. Negash [et al.] // *International Journal of Population Data Science*. — 2023. — Vol. 8, no. 1. — Art. 29.
5. Бабак, Н. Г. Автоматическое обезличивание конфиденциальной информации / Н. Г. Бабак [и др.] // *Russian Technological Journal*. — 2023. — Т. 11, № 5. — С. 7–18.
6. Liu, Z. De-identification of clinical notes via recurrent neural network and conditional random field / Z. Liu, B. Tang, X. Wang, Q. Chen // *Journal of Biomedical Informatics*. — 2017. — Vol. 75. — P. S34–S42.
7. Devlin, J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // *Proceedings of NAACL-HLT*. — 2019. — P. 4171–4186.
8. Vaswani A. Attention Is All You Need / A. Vaswani, L. Jones, N. Shazeer [et al.] // *31st Conference on Neural Information Processing Systems (NIPS 2017)*. – Long Beach, CA, USA, 2017
9. Jia, J. A flexible two-stage anonymization framework for narrative medical records adapting to various language models / J. Jia, H. Nishi // *Computers in Biology and Medicine*. — 2025. — Vol. 195. — Art. 110624.
10. Starovoytova, E., Kulikov, E., Fedosenko, S., Shmyrina, A., Kirillova, N., Vinokurova, D., Balaganskaya, M.: *RuMedPrimeData* — 2021. [Электронный ресурс]. URL: <https://zenodo.org/records/5765873> (дата обращения: 31.03.2026).
11. Kumichev, G. *MedSyn: LLM-based Synthetic Medical Text Generation Framework, Synthetic dataset*. [Электронный ресурс]. URL: <https://huggingface.co/datasets/Glebkaa/MedSyn-synthetic> (дата обращения: 31.03.2026).
12. Kumichev, G. *MedSyn: LLM-based Synthetic Medical Text Generation Framework* / G. Kumichev [et al.] // *MIL Team, Sber AI Lab*. — 2024. — 18 p.
13. Zmitrovich D., Abramov A., Kalmykov A. [et al.]. A Family of Pretrained Transformer Language Models for Russian // *arXiv preprint arXiv:2309.10931*. 2023. [Электронный ресурс]. URL: <https://arxiv.org/abs/2309.10931> (дата обращения: 30.04.2026).
14. Slijepčević, D. k-Anonymity in practice: How generalisation and suppression affect machine learning classifiers / D. Slijepčević [et al.] // *Computers & Security*. — 2021. — Vol. 111. — Art. 102488.
15. LeFevre, K. Mondrian multidimensional k-anonymity / K. LeFevre, D. J. DeWitt, R. Ramakrishnan // *Proceedings of the 22nd International Conference on Data Engineering (ICDE '06)*. — 2006

УДК 624.131.3:624.131.5

Рогов И. А.

Санкт-Петербургский государственный архитектурно-строительный университет,
Санкт-Петербург, Россия

ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ АНИЗОТРОПИИ ПРОЧНОСТНЫХ ХАРАКТЕРИСТИК ЛЕНТОЧНОГО СУГЛИНКА НА ГРАНИЦЕ С РАЗНЫМИ МАТЕРИАЛАМИ

Аннотация. Представлены результаты экспериментального исследования прочностной анизотропии ленточных суглинков отобранных на территории Санкт-Петербурга. Определены характеристики сопротивления сдвигу как по грунту вдоль и поперёк слоистости, так и на контакте с разными материалами. Установлено, что прочность на срез поперёк слоёв значительно выше, чем вдоль, а переход к системе «грунт–материал» сопровождается снижением прочности. Выявлены количественные коэффициенты прочностной анизотропии. Показано, что угол внутреннего трения на контакте зависит от типа материала. Характер изменения сцепления на границе с материалами противоположен изменению угла трения и демонстрирует резкое падение при срезе поперёк напластования. Полученные данные важны для расчёта несущей способности свай и других подземных конструкций в слоистых глинистых грунтах.

Ключевые слова: прочность на срез, прочностная анизотропия, ленточные суглинки, система «грунт-грунт», система «грунт-материал».

Введение

Анизотропия – это зависимость физических и механических свойств материала от направления приложения нагрузки. В механике грунтов различают несколько видов анизотропии: деформационную, фильтрационную и прочностную [1, 2]. Прочностная анизотропия проявляется в том, что предельное сопротивление сдвигу, описываемое углом внутреннего трения φ и удельным сцеплением c , существенно зависит от ориентировки поверхности разрушения относительно структурных элементов грунтового массива [1]. Именно прочностная анизотропия оказывает прямое влияние на несущую способность оснований, устойчивость откосов и боковое сопротивление свай.

Главной причиной прочностной анизотропии глинистых грунтов является их текстура – пространственная ориентировка глинистых частиц и микроагрегатов, слоистость, наличие ориентированных систем трещин и включений [2]. Наиболее ярко текстурная анизотропия выражена у отложений, сформировавшихся в спокойных водных бассейнах с сезонной слоистостью. Классическим примером таких грунтов служат озёрно-ледниковые ленточные суглинки, широко распространённые на территории Северо-Запада России, в том числе в Санкт-Петербурге [3, 4].

Озёрно-ледниковые отложения представлены ритмичным чередованием тонких пылевато-глинистых слоёв и более плотных глинистых прослоев, что

придает породе характерную ленточную текстуру [3]. Каждый годовой цикл образует одну ленту толщиной от долей миллиметра до нескольких сантиметров. Именно эта упорядоченная микроструктура обуславливает резко выраженную прочностную анизотропию: сопротивление сдвигу вдоль слоистости, как правило, ниже, чем поперёк неё, а деформационные и фильтрационные свойства радикально различаются по направлениям [1, 2]. Для ленточных суглинков характерны высокие значения угла внутреннего трения при нагружении поперёк напластования и заметно пониженные – при срезе по ослабленным контактам между слоями.

Несмотря на то, что анизотропия механических свойств ленточных глин изучается достаточно давно, основное внимание исследователей уделялось поведению самого грунтового массива [1, 2, 3]. Вместе с тем, при работе подземных конструкций, таких как висячие сваи, определяющим является её сопротивление трению по боковой поверхности, т.е. на границе системы «грунт–материал» [5]. Однако количественные данные об изменении прочностных характеристик на контакте с конструкционными материалами при различной ориентации слоёв остаются ограниченными. Это затрудняет обоснованное проектирование свайных фундаментов в условиях распространённости анизотропных ленточных суглинков [4]. Целью настоящей работы является экспериментальное определение параметров анизотропии прочности на сдвиг ленточных суглинков как по грунту, так и на границе со сталью и стеклопластиком.

Проведение испытания и результаты

Для исследований свойства анизотропии прочностных характеристик на границе с разными материалами на территории Санкт-Петербурга были отобраны монолиты суглинка с ярко выраженным чередованием слоёв разного гранулометрического состава. Фотография монолита приведена на рисунке 1.



Рисунок 1. Текстура ленточного суглинка

В качестве материалов были использованы круглые пластины диаметром 71 мм, изготовленные из стали и стеклопластика.



Рисунок 2. Испытываемая стальная пластина

С помощью режущего кольца из монолитов были вырезаны образцы вдоль и поперёк направления слоистости и проведены стандартные испытания

на срез на приборах АСИС компании ООО «Геотек» в соответствии с рекомендациями актуальной нормативной документации [6].

После чего образцы грунта изымались из прибора, разделялись на две половины по линии среза и заново подготавливались к проведению испытания на сдвиг на границе с материалами. Испытания проводились в соответствии с методикой, описанной в [5,7].

Даная методика базируется на предположении, что процесс установки конструкций фундамента, например, свай, обеспечивает более плотное притирание грунта к поверхности материала, что приводит к увеличению прочности системы «грунт-материал». Таким образом реальные значения прочностных характеристик системы можно определить путём проведения серии последовательных повторных сдвигов без разбора прибора.

Все течение всего испытания образцы находились в мягкопластичной консистенции, показатель текучести грунта I_L варьировался в диапазоне от 0,73 до 0,56.

За конечное значение максимальной касательного напряжения принималось среднее значение результатов трёх последовательных испытаний на срез, в которых значения τ_{max} отличались не более, чем на 3%.

Результаты определения прочностных характеристик ленточного суглинка приведены на рисунке 3.

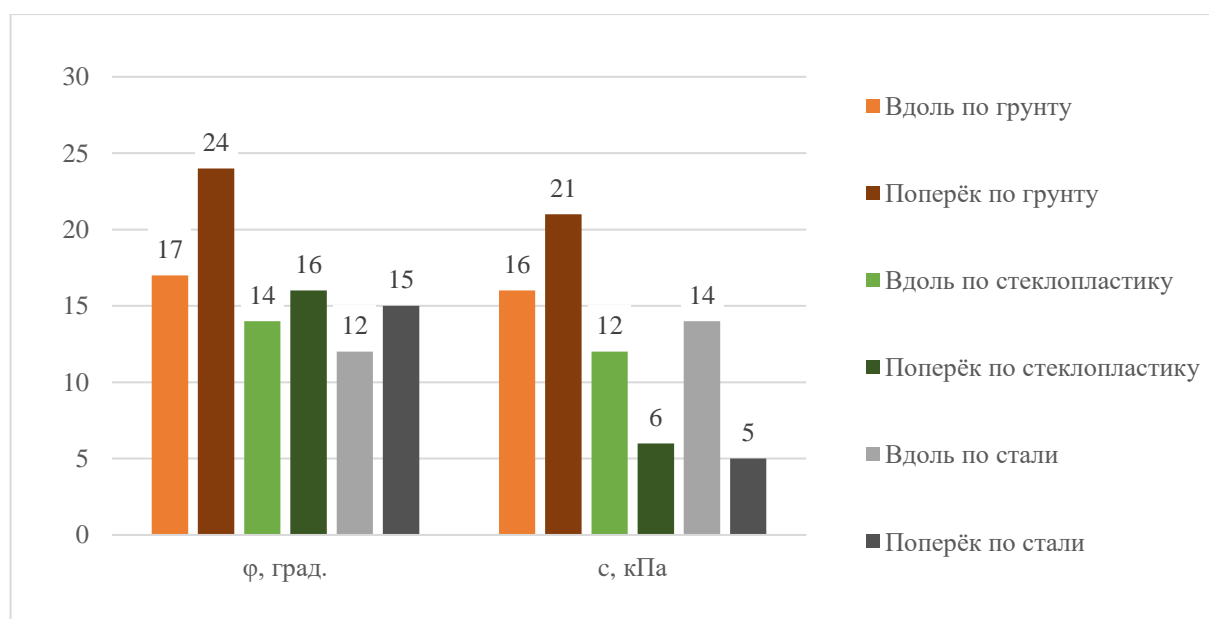


Рисунок 3. Прочностные характеристики ленточного суглинка

Результаты исследования показали, что при срезе поперёк слоёв значения прочностных характеристик выше, чем при срезе вдоль.

Значения угла внутреннего трения ϕ при переходе от среза на границе системы «грунт-грунт» к границе системы «грунт-материал» в случае среза вдоль слоёв показали снижение на 20 и 30% для стеклопластика и стали

соответственно, для случая среза поперёк слоёв снижения составили 30 и 40%. Показатели угла внутреннего трения на границе системы «грунт-материал» при ориентации слоёв поперёк так же оказались выше, чем при ориентации слоёв вдоль линии среза.

Характер изменений удельного сцепления c отличается от характера изменений угла внутреннего трения ϕ : во-первых, значение сцепления вдоль слоёв на границе со сталью оказалось выше, чем на границе со стеклопластиком, во-вторых, при срезе по материалам поперёк линии напластования наблюдается резкое падение сцепления.

Коэффициенты анизотропии для угла внутреннего трения ϕ и удельного сцепления c определялись по формулам (1) и (2), и приведены на рисунке 4.

$$K_{a\phi} = \frac{\phi_{\text{поперёк}}}{\phi_{\text{вдоль}}}(1)$$

$$K_{ac} = \frac{c_{\text{поперёк}}}{c_{\text{вдоль}}}(2)$$

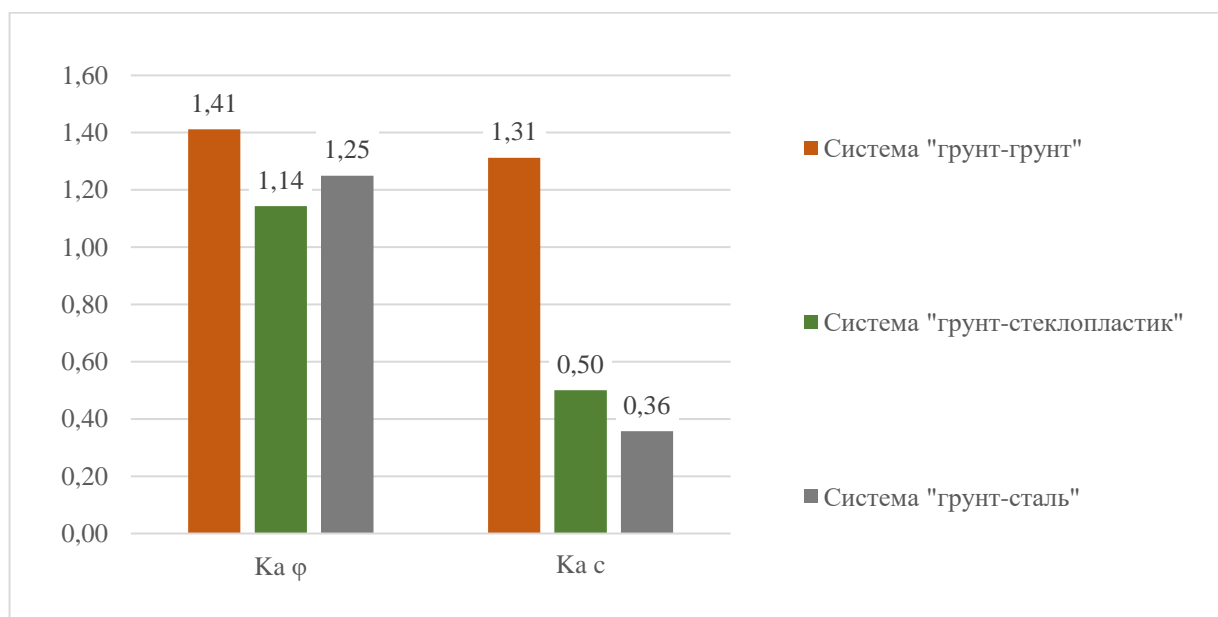


Рисунок 4. Коэффициенты прочностной анизотропии

Выводы и обсуждение

Проведённые испытания подтвердили, что исследуемый ленточный суглинок обладает ярко выраженной прочностной анизотропией. Коэффициент анизотропии угла внутреннего трения $K_{a\phi}$ при срезе по грунту составил 1,41, т.е. сопротивление сдвигу поперёк слоистости почти в полтора раза выше, чем вдоль. Коэффициент анизотропии сцепления K_{ac} оказался

равным 1,25, что также свидетельствует о более высоком структурном сцеплении в направлении поперёк слоистости.

Система «грунт–грунт» во всех случаях показала более высокую прочность, чем система «грунт–материал». Объясняется это тем, что на контакте системы «грунт–материал» сопротивление определяется преимущественно адгезионными силами и внешним трением по относительно гладкой поверхности, что ведёт к снижению как угла внутреннего трения, так и удельного сцепления.

Угол внутреннего трения на границе «грунт–материал» существенно зависит от шероховатости и физико-химической природы материала. Стеклопластик, обладающий большей шероховатостью, обеспечил более высокие значения φ , чем сталь. При срезе вдоль слоёв значение φ уменьшилось по сравнению с грунтом на 20 % для стеклопластика и 30 % для стали; при срезе поперёк слоёв – 30 и 40 % соответственно. Абсолютные значения φ при ориентации поперёк слоёв оставались выше, чем при срезе вдоль, сохраняя общую тенденцию анизотропии.

Динамика изменения удельного сцепления c на контакте с материалами принципиально иная. При срезе поперёк напластования зафиксировано резкое падение сцепления: для системы «грунт–сталь» $K_{ac} = 0,36$, для «грунт–стеклопластик» $K_{ac} = 0,50$. Это означает, что структурное сцепление, обусловленное цементационными связями и ориентацией глинистых частиц поперёк слоёв, практически не передаётся на границу с инородным материалом, особенно с гладкой стальной поверхностью. Вдоль слоистости снижение сцепления меньше, однако его величина на контакте со сталью оказалась выше, чем на границе со стеклопластиком, что может быть связано с особенностями адсорбции глинистых частиц на окисленной стальной поверхности.

Полученные результаты демонстрируют, что при проектировании свайных фундаментов в ленточных суглинках необходимо учитывать не только исходную анизотропию грунта, но и пониженные прочностные характеристики на контакте с материалом ствола сваи. Особое внимание следует уделять случаям, когда направление сдвига ориентировано поперёк слоёв, так как потеря сцепления на границе может быть критической.

Список литературы:

1. Цытович Н.А. Механика грунтов. – М.: Высшая школа, 1983. – 288 с.
2. Гольдштейн М.Н. Механические свойства грунтов. – М.: Стройиздат, 1979. – 367 с.
3. Дашко Р.Э., Каган А.А. Механика грунтов в инженерно-геологической практике. – М.: Недра, 1990. – 267 с.
4. Улицкий В.М., Шашкин А.Г. Основания и фундаменты. Проектирование и расчёт в условиях плотной городской застройки. – СПб.: Стройиздат СПб, 2004. – 392 с.

5. Бояринцев А.В. Учёт влияния материала конструкции свайного фундамента на совместную работу с грунтовым основанием при его промерзании: дис. на соиск. уч. степ. канд. тех. наук / Бояринцев А.В.: СПбГАСУ. – СПб, 2022. – 199 с.

6. ГОСТ 12248.1-2020. Грунты. Определение характеристик прочности методом одноплоскостного среза. – М.: Стандартиформ, 2020. – 26 с.

7. Патент № 2849850 С1 Российская Федерация, МПК E02D 1/02. способ определения параметров трения немерзлого грунта о поверхности фундамента и конструкции подземных сооружений: заявл. 14.02.2025: опубл. 30.10.2025 / А. В. Бояринцев; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования "Санкт-Петербургский государственный архитектурно-строительный университет". – EDN VTBOFV.

УДК 004.738.5:004.056.55

Зацепин Д. А.

Московский автомобильно-дорожный государственный технический университет
(МАДИ), Москва, Россия

ИССЛЕДОВАНИЕ И АНАЛИЗ ПРОБЛЕМ И ТЕНДЕНЦИЙ РАЗВИТИЯ БЛОКЧЕЙН-ТЕХНОЛОГИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Аннотация. В статье рассматриваются современное состояние, ключевые тенденции и практические проблемы внедрения блокчейн-технологий в информационные системы. Анализируется эволюция технологии от стадии хайпа к зрелому промышленному применению, исследуются глобальные тренды 2025–2026 годов, включая модульные архитектуры, конвергенцию с искусственным интеллектом, развитие DePIN и институциональную токенизацию. Особое внимание уделяется российскому контексту: нормативно-правовому регулированию криптовалют, развитию рынка цифровых финансовых активов (ЦФА), проектам в государственном управлении, банковском секторе и строительной отрасли. Анализируются ключевые проблемы масштабируемости, интероперабельности, безопасности и регулирования, а также предлагаются перспективные направления дальнейших исследований.

Ключевые слова: Блокчейн-технологии, Информационные системы, Тенденции развития блокчейна, Проблемы внедрения, Смарт-контракты, Управление строительными контрактами, Цифровые финансовые активы (ЦФА), Регулирование криптовалют в России, Масштабируемость блокчейна, Интероперабельность Web3-систем, Токенизация реальных активов (RWA), DePIN (децентрализованные физические сети), Zero-Knowledge Proofs (доказательства с нулевым разглашением), Конвергенция блокчейна и ИИ, Отечественные блокчейн-платформы (Мастерчейн, Waves Enterprise).

1. Введение

Технология распределенных реестров (блокчейн) за прошедшее десятилетие прошла путь от радикальной криптоанархической идеи до признанного инструмента цифровой трансформации. Если в 2017–2019 годах доминировал ажиотаж и множество малоприменимых «блокчейн-стартапов», то к 2026 году технология обрела зрелые формы и устойчивую практическую ценность [4]. Как отмечает Осама Бари, технический директор D24 Fintech, «2025 год стал годом, когда блокчейн наконец превратился в корпоративную инфраструктуру» [6]. Согласно оценкам McKinsey, рынок блокчейн-технологий к 2030 году может достичь \$2 трлн.

В современной дискуссии сместился фокус с вопроса «нужен ли блокчейн» на вопрос о том, решает ли технология конкретную задачу лучше, чем традиционная база данных [1]. Блокчейн обладает рядом фундаментальных преимуществ: неизменяемость записей, прозрачность для участников, децентрализация и возможность автоматизации через смарт-контракты. Однако

эти преимущества достигаются ценой определенных компромиссов, что создает множество проблем при интеграции в реальные информационные системы.

Данная статья представляет собой исследование и анализ актуальных тенденций и проблем развития блокчейн-технологий в контексте их применения в информационных системах с учетом глобальных трендов и российской специфики.

2. Современное состояние и ключевые мировые тенденции

2.1. Переход к зрелости: от экспериментов к промышленной инфраструктуре

К 2026 году технология блокчейн пережила глубокую структурную перестройку. Схлынула первоначальная эйфория, и на смену ей пришло осознанное внедрение там, где технология действительно дает добавленную стоимость. Ключевыми драйверами зрелости стали:

Стабилизация L1 и L2-решений. Сети первого уровня (Layer 1) и решения масштабирования второго уровня (Layer 2) достигли уровней стабильности, скорости и дешевизны, позволяющих строить продукты для массового использования. Роллапы (rollups) снизили комиссии за газ до менее \$1 и увеличили пропускную способность до 2000 TPS, решая проблему перегрузки сети Ethereum.

Потребность в верифицируемости («Flight to Verifiability») [6]. После глобальных кризисов целостности данных 2025 года спрос на непротиворечивые и верифицируемые записи достиг беспрецедентного уровня. Крупные предприятия начали активно рассматривать блокчейн как способ создания «единого источника правды» в сложных экосистемах.

2.2. Эра модульных блокчейнов

Одним из важнейших трендов 2026 года является переход от монолитных к модульным блокчейн-архитектурам [6]. Платформы, такие как EigenLayer и Celestia, позволили разделить функции выполнения (execution), достижения консенсуса (consensus) и доступности данных (data availability), что радикально повышает масштабируемость, интероперабельность и экономическую эффективность. Для разработчиков корпоративных систем это означает, что в ближайшем будущем интеграция блокчейн-SDK станет такой же рутинной операцией, как подключение облачных API десятилетие назад.

2.3. Токенизация реальных активов (RWA)

Институциональная токенизация вышла на передний план. Банки и инвестиционные фонды начали выпускать токенизированные казначейские облигации и бонды, торгующиеся 24/7 на блокчейне [6]. По данным D24 Fintech, это привело к перемещению триллионов долларов реальных активов на блокчейн-рельсы, устраняя простои традиционных рынков, закрытых по выходным. Крупнейшим барьером остается разрозненность регулирования в разных юрисдикциях, требующая создания единого глобального правового поля.

2.4. Конвергенция блокчейна и искусственного интеллекта

2026 год стал точкой бифуркации, когда блокчейн, ИИ и платежные системы начали объединяться в единый технологический стек. ИИ становится «решающим двигателем» (decision engine) блокчейн-стека. Ключевые направления конвергенции включают:

Аудит и верификация смарт-контрактов: ИИ-модели автоматически проверяют безопасность и корректность кода смарт-контрактов.

Автоматизированный трейдинг и арбитраж: ИИ-боты в режиме реального времени отслеживают возможности арбитража на децентрализованных биржах и исполняют сделки через смарт-контракты без участия человека.

AML/KYC-комплаенс: Модели машинного обучения сканируют ончейн-транзакции для выявления подозрительной деятельности в реальном времени.

Децентрализованный ИИ-инжиниринг: Блокчейн обеспечивает проверяемость данных, используемых для обучения ИИ, и создает рынки для краудсорсинга вычислительных ресурсов (в рамках модели DePIN).

2.5. DePIN (Decentralized Physical Infrastructure Networks)

DePIN представляет собой значимый тренд 2026 года, соединяющий цифровое пространство блокчейна с физической инфраструктурой: энергосетями, сетями беспроводной связи, системами хранения данных и вычислительными ресурсами. Согласно прогнозам Messari, годовой доход DePIN может удвоиться в 2026 году, превысив \$100 млн. Примечательно, что рынок прошел естественный отбор: проекты без реального бизнес-дохода, опиравшиеся исключительно на спекулятивный ажиотаж, были отсеяны еще до 2026 года.

2.6. Zero-Knowledge-технологии и конфиденциальность

Блокчейн эволюционирует в сторону «избирательной прозрачности» (selective transparency) за счет технологий ZK-proofs и конфиденциальных вычислений. Это решает одну из ключевых проблем корпоративного внедрения: противоречие между публичной верифицируемостью и необходимостью защиты коммерческой тайны. ZK-доказательства позволяют подтвердить корректность транзакции без раскрытия ее конкретных данных, что открывает возможности для использования блокчейна в банкинге и промышленности, где конфиденциальность критична [6].

3. Блокчейн в информационных системах: практические проблемы внедрения

Несмотря на значительный прогресс, внедрение блокчейн-технологий в реальные информационные системы сталкивается с рядом фундаментальных проблем, которые активно исследуются как в научной литературе, так и в инженерной практике.

3.1. Проблемы масштабируемости и производительности

Тема масштабирования является классической проблемой блокчейн-систем [1], изучению которой посвящены десятки научных работ. Основная дилемма заключается в «трилемме блокчейна», сформулированной Виталиком Бутериным: децентрализация, безопасность и масштабируемость находятся в

противоречии, и одновременное достижение всех трех свойств в высокой степени представляет собой нерешенную задачу. Публичные блокчейны (Bitcoin, Ethereum) обеспечивают высокий уровень децентрализации и безопасности, но ограничены в пропускной способности и стоимостных затратах [9].

В то же время новые подходы дают обнадеживающие результаты. Метод периодической агрегации с двойным хеш-якорением (PADHA) позволяет сокращать объем хранимых данных в PoW-блокчейнах без нарушения аудируемости и децентрализации за счет синергии обрезки блоков и хеш-функций-хамелеонов. Приватные и корпоративные блокчейны жертвуют частью децентрализации в пользу производительности, что делает их более пригодными для промышленного использования.

3.2. Интероперабельность Web3-систем

Разнообразие блокчейн-платформ привело к фрагментации экосистемы. В России используются различные платформы: Ethereum (Мастерчейн, Токен, НРД), Hyperledger Fabric (Атомайз, СПб Биржа, Спутник ЦФА), Waves (Конфидент, МЧД) и другие [2]. Каждая из этих платформ решает задачи масштабируемости и устойчивости по-своему, используя разные модели консенсуса: CFT, BFT, PoA.

Технологически эта проблема решаема: отрасль уже умеет строить мосты между разными блокчейнами. Существуют три основных типа архитектур мостов:

Атомарные мосты на основе Hash-Time-Lock-контрактов обеспечивают полную децентрализацию за счет криптографической безопасности.

Lock-and-Mint-мосты блокируют актив в одной сети и выпускают его «обертку» в другой под управлением единого оркестратора.

Мосты на основе guardians и оракулов (LayerZero, Wormhole) передают сообщения через легкие клиенты и ретрансляторы.

Однако куда более жесткими ограничениями становятся регуляторные требования ФСТЭК России, ФСБ России и Банка России к средствам криптографической защиты информации (СКЗИ). Именно требования к СКЗИ чаще всего тормозят проекты, поскольку плохо сочетаются с архитектурой распределенных реестров, особенно при попытке применения ГОСТ-криптографии в открытых блокчейнах.

3.3. Стоимость транзакций и энергопотребление

Высокая стоимость транзакций и значительное энергопотребление (особенно в PoW-сетях) остаются существенными проблемами при широкомасштабном внедрении [1]. Хотя переход Ethereum на PoS (The Merge, 2022) существенно сократил энергопотребление этой сети, PoW-системы вроде Bitcoin продолжают потреблять значительные энергоресурсы. Для корпоративных решений эта проблема часто решается выбором частных сетей с более энергоэффективными алгоритмами консенсуса, хотя за это приходится платить снижением уровня децентрализации.

3.4. Безопасность и приватность

Хотя блокчейн сам по себе обеспечивает высокий уровень защищенности от несанкционированного изменения данных, уязвимости часто возникают на пограничных уровнях информационной системы. Проблемы включают ошибки в реализации смарт-контрактов, уязвимости мостов между блокчейнами, атаки на механизмы консенсуса (включая «атаку 51%»), а также соотношение между необходимой прозрачностью для верификации и защитой конфиденциальных данных.

В здравоохранении, например, публичные блокчейны не рассматриваются для хранения медицинских записей напрямую, но используются для хранения хеш-сумм и ссылок на зашифрованные данные, решая вопрос конфиденциальности через гибридные архитектуры.

4. Российский контекст: регулирование и практические проекты

4.1. Нормативно-правовое регулирование криптовалют

Россия переживает фундаментальный поворот в отношении к криптовалютам и блокчейн-технологиям [3]. В конце 2024 года майнинг получил первый легальный статус в стране — сектор стал признанной регулируемой деятельностью. В марте 2025 года стартовал экспериментальный правовой режим (ЭПР) для ограниченного трансграничного использования криптовалют, а в мае были разрешены производные инструменты (фьючерсы на Bitcoin и Ethereum).

Наиболее значимые изменения запланированы на 2026–2027 годы. До 1 июля 2026 года должна быть принята полная концепция комплексного регулирования, предложенная Банком России. С июля 2026 года резиденты будут обязаны уведомлять ФНС об открытии и закрытии блокчейн-кошельков в иностранных системах. Использование криптовалют для расчетов внутри страны запрещено, транзакции разрешены только через лицензированных посредников (биржи, брокеры, криптообменники) с соблюдением KYC/AML-требований, хранением данных на территории РФ и прохождением процедур блокчейн-аналитики. С июля 2027 года вводится ответственность за незаконную деятельность на крипторынке, аналогичная ответственности за нелегальные банковские операции.

При этом, согласно оценкам Chainalysis, за период июль 2024 — июль 2025 годов оборот российского крипторынка достиг \$379,3 млрд — лидирующая позиция в Европе. Российские сервисы обмена криптовалют, по оценкам компании «Шард», провели за 2025 год транзакций на 44 трлн рублей. Россия также занимает 16% мирового хешрейта Bitcoin, став одной из крупнейших стран по добыче.

4.2. Цифровые финансовые активы (ЦФА)

Рынок ЦФА в России демонстрирует устойчивый экспоненциальный рост. По итогам 2025 года объем первичных размещений ЦФА превысил 1,65 трлн рублей, что почти втрое выше показателя 2024 года [7]. За три года совокупный объем рынка достиг 1,5 трлн рублей, причем в основном выпускаются

краткосрочные (до месяца) инструменты. Уже 18 операторов информационных систем работают на рынке цифровых активов в России.

Сбербанк выступает ключевым драйвером этого направления: клиентам уже доступны инвестиционные продукты с привязкой к криптоактивам (структурные облигации и ЦФА, привязанные к Bitcoin и Ethereum), продолжается работа над собственной блокчейн-платформой. Интересно, что опрос Сбербанка показал: 83% россиян знакомы с термином «криптовалюта», а 43% хотят иметь удобный российский криптокошелек в инфраструктуре знакомых банков и брокеров.

4.3. Государственные проекты на блокчейне

Российское государство активно экспериментирует с блокчейн-технологиями в различных сферах:

«Прозрачный блокчейн» Росфинмониторинга. Сервис завершил пилотное тестирование в пяти крупнейших банках и в 2026 году продолжит интеграцию в национальную антиотмывочную систему. Сервис помогает банкам выявлять связи между операциями с криптовалютой и обычными деньгами.

Цифровой рубль. Госдума приняла законодательство о поэтапном внедрении CBDC с полным запуском, намеченным на сентябрь 2026 года. Крупные финансовые институты обязаны принять инфраструктуру цифрового рубля к этому сроку.

Цифровая платформа оценки налогоплательщиков. Федеральная налоговая служба планирует создать платформу оценки налогоплательщиков на базе блокчейна, которая позволит бизнесу быстрее получать кредиты на выгодных условиях за счет подтвержденной налоговой истории.

Полностью отечественный блокчейн. Университет Иннополис запатентовал первый отечественный блокчейн с открытым кодом, поддерживающий российскую криптографию и смарт-контракты.

4.4. Блокчейн в строительстве и управлении контрактами

Показателен пример пилотного проекта ФНС России, Альфа-Банка и Gaskar Group по созданию цифровой системы для взаиморасчетов между заказчиком и подрядчиками по строительно-монтажным работам [3]. Информационная система через блокчейн передает в модуль банкинга основание для оплаты в рамках строительного контракта — данные из платформы управления строительством.

Этот кейс демонстрирует ключевые преимущества блокчейна в управлении контрактами: бóльшую прозрачность, снижение операционных издержек и минимизацию споров об оплате. В контексте управления строительными контрактами технология позволяет создавать неизменяемые записи всех транзакций и этапов работ, автоматизировать оплату по актам КС-2 и КС-3 при наступлении заданных условий, а также отслеживать поставки материалов с подтверждением соответствия сертификатам.

5. Перспективные направления дальнейших исследований

На основе проведенного анализа можно выделить несколько перспективных направлений для дальнейших исследований в области блокчейн-технологий в информационных системах:

Интеграция блокчейна с BIM-моделями. Создание архитектуры «цифрового двойника» строительного объекта на блокчейне, где смарт-контракты автоматически инициируют платежи при поступлении верифицированных данных от датчиков IoT и BIM-модели, является малоизученным направлением с высоким научным и практическим потенциалом. Разработка программного оркестратора, интегрирующего BIM-моделирование и блокчейн-расчеты, может стать прорывным решением.

Интероперабельность Web3-систем с ГОСТ-криптографией [2]. В условиях российского регулирования задача построения защищенных криптографических мостов между локальными блокчейн-платформами (с поддержкой СКЗИ) и публичными сетями представляет собой нерешенную научно-техническую проблему.

Децентрализованные модели управления доступом на основе блокчейна. Разработка и экспериментальное сравнение производительности и безопасности различных DApp-архитектур для управления доступом к корпоративным ИС (RBAC, ABAC на смарт-контрактах) является перспективным исследовательским направлением.

Применение Zero-Knowledge Proofs в корпоративных блокчейн-системах. Исследование возможностей интеграции ZK-технологий для обеспечения конфиденциальности транзакций в частных блокчейн-сетях без потери верифицируемости.

DePIN в российской инфраструктуре. Анализ применимости децентрализованных физических сетей в сферах совместного использования вычислительных мощностей, распределенной энергогенерации и IoT-инфраструктуры в специфических российских условиях.

6. Заключение

Проведенное исследование позволяет сделать следующие выводы:

Блокчейн-технологии к 2026 году перешли от стадии эксперимента и хайпа к зрелому промышленному применению. Ключевыми трендами стали модульные архитектуры, конвергенция блокчейна с искусственным интеллектом, развитие DePIN и институциональная токенизация.

Основными практическими проблемами внедрения остаются масштабируемость (трилемма блокчейна), интероперабельность разнородных систем, высокая стоимость транзакций, вопросы безопасности и приватности данных, а также разрозненность глобального правового регулирования.

В России наблюдается фундаментальный поворот в регулировании: от полного отсутствия правил к структурированной системе с лицензированием участников, KYC/AML-требованиями и формированием национальной криптоинфраструктуры. Рынок ЦФА демонстрирует устойчивый рост, а государственные проекты («Прозрачный блокчейн», цифровой рубль,

платформа ФНС) активно внедряют блокчейн в систему государственного управления.

Практические кейсы в строительстве (пилот ФНС, Альфа-Банка и Gaskar Group), логистике (гибридные блокчейны для управления цепочками поставок [17†L13-L16]) и образовании (цифровые дипломы на блокчейне [20†L4-L9]) демонстрируют реальные преимущества технологии для информационных систем: повышение прозрачности, снижение издержек и минимизацию спорных ситуаций.

Перспективными направлениями дальнейших исследований являются интеграция блокчейна с BIM-моделями, ГОСТ-совместимая интероперабельность Web3-систем, децентрализованные модели управления доступом, применение ZK-proofs в корпоративных блокчейнах и анализ применимости DePIN в российской инфраструктуре.

Список литературы:

1. Грепан В.Н. Практические проблемы использования блокчейн-технологий // Столыпинский вестник. 2024. №9.
2. Якушкин М. Интероперабельность Web3-систем по ГОСТу [Электронный ресурс] // IT Security. 2026. URL: <https://www.itsec.ru/articles/interoperabelnost-web3-po-gostu> (дата обращения: 10.05.2026).
3. Китанин С.С., Чемерис О.С., Прокопов М.А. Внедрение блокчейн-технологий в финансовом секторе экономики // Московский экономический журнал. 2025. № 3. URL: <https://1economic.ru/lib/122927> (дата обращения: 10.05.2026).
4. Тянько В.Д., Молодецкая М.В., Залётова М.В., Симонова Н.А., Усманов К.П. Блокчейн и смарт-контракты в строительстве // Финансовая экономика. 2022. № 2. С. 214–216.
5. Солохов И.Р., Хайруллин Р.З., Афанасьев Г.И. Гибридный блокчейн с динамическими смарт-контрактами для автоматизации логистических и таможенных процессов // Comp. nanotechnol. 2025. Т. 12, вып. 3. С. 105–114.
6. Гусева А.А., Власова М.С. Регулирование блокчейн-технологий в финансовом секторе // Вестник Воронежского государственного университета. Серия: Экономика и управление. 2025. № 4(67). URL: <https://sciup.org/regulirovanie-blokchejn-tehnologij-v-finansovom-sektore-523-regionalnaja-i-148332714> (дата обращения: 10.05.2026).
7. Давыдова Н.Л., Бухтик А.А. Цифровые финансовые активы (цифровые знаки): компаративный анализ регулирования и тенденции развития рынка в РФ и Беларуси // Детерминанты развития экономики и общества. Том 2. М.: Университетская книга, 2025. С. 99–106.
8. Мамыкина А.А., Уразбахтина Л.Р. Тенденции и перспективы развития рынка цифровых финансовых активов // Вопросы устойчивого развития общества. 2024. № 8.
9. Razuvaev D.D., Ustinov S.M. Mitigating data growth in PoW blockchains: Storage reduction methods for scalability without compromising decentralization // Computing, Telecommunication and Control, 2025.
10. Israfilov A. PERSPECTIVES ON THE USE OF BLOCKCHAIN FOR DECENTRALIZED AND SECURE DATA EXCHANGE // Бюллетень науки и практики. 2024. №12. URL: <https://cyberleninka.ru/article/n/perspectives-on-the-use-of-blockchain-for-decentralized-and-secure-data-exchange> (дата обращения: 10.05.2026).

УДК 004.891:004.85:37.091.26

Асташкина А. А.

МИРЭА – Российский технологический университет, Москва, Россия

ПРИМЕНЕНИЕ RAG-АРХИТЕКТУРЫ И БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ АВТОМАТИЧЕСКОЙ ПРОВЕРКИ РАЗВЕРНУТЫХ ОТВЕТОВ ОБУЧАЮЩИХСЯ

Аннотация. Статья посвящена разработке интеллектуальной системы автоматической проверки развернутых ответов обучающихся с использованием RAG-архитектуры и больших языковых моделей. Рассматриваются особенности применения Retrieval-Augmented Generation для повышения качества автоматического оценивания учебных работ. Предложенный подход сочетает использование учебных материалов и экспертных примеров проверки, что позволяет учитывать как предметные знания, так и особенности экспертного оценивания ответов обучающихся. Проведено экспериментальное исследование качества извлечения релевантной информации и точности автоматической проверки ответов. Полученные результаты показывают, что использование внешнего контекста позволяет повысить согласованность оценок системы и эксперта.

Ключевые слова: RAG-архитектура, большие языковые модели, автоматическая проверка ответов, интеллектуальные образовательные системы, развернутые ответы обучающихся.

Введение

Контроль знаний является важной частью образовательного процесса, позволяющей оценить уровень усвоения материала и выявить пробелы в знаниях обучающихся. При этом традиционная ручная проверка требует значительных временных затрат и подвержена влиянию субъективного фактора, что делает задачу автоматизации оценивания особенно актуальной.

Наибольшую сложность представляет автоматическая оценка открытых заданий с развернутым ответом, поскольку она требует анализа содержания, логики рассуждений и корректности формулировок. Данная задача относится к направлению Automated Short Answer Grading (ASAG).

Существующие подходы к решению задачи ASAG, такие как методы, основанные на правилах, алгоритмы машинного обучения и модели глубокого обучения, обладают рядом ограничений, включая зависимость от заданных шаблонов, необходимость наличия размеченных данных и ограниченную способность выявлять сложные логические связи в ответах обучающихся.

В последние годы большое внимание уделяется применению больших языковых моделей (LLM). Они демонстрируют высокий потенциал в задачах

анализа текстов и оценки ответов, однако остаются проблемы, связанные с достоверностью генерируемой информации и объяснимостью результатов.

Одним из перспективных подходов к повышению точности и надежности проверки ответов является Retrieval-Augmented Generation (RAG), который сочетает возможности поиска релевантной информации и генерации ответа на основе проверенных внешних источников знаний.

Целью данной работы является разработка интеллектуальной системы автоматической проверки развернутых ответов обучающихся на основе RAG-архитектуры и больших языковых моделей.

1. Обзор существующих подходов к автоматической оценке ответов обучающихся

Автоматическая оценка кратких ответов (ASAG) представляет собой задачу оценки ответов на объективные вопросы, написанных на естественном языке, с использованием вычислительных методов. В отличие от заданий с выбором ответа, ASAG требует анализа содержания текста, выявления ключевых смысловых элементов и сопоставления их с эталонным ответом.

1.1. Классические подходы к ASAG

На ранних этапах развития систем ASAG использовались методы, основанные на правилах. К ним относится, например, метод концептуальных карт, при котором ответ обучающегося представляется в виде набора ключевых тезисов, а оценка определяется на основе наличия или отсутствия этих тезисов в ответе [1].

Другим подходом является извлечение информации, при котором из ответа выделяются конкретные факты с помощью шаблонов и правил (например, регулярных выражений) [2]. Полученные структурированные элементы ответа затем сопоставляются с эталоном.

Позднее получили развитие методы, основанные на корпусах текстов, предполагающие сравнение ответа обучающегося с набором эталонных или близких по смыслу ответов [3]. Такие методы используют статистические характеристики текстов и позволяют учитывать вариативность формулировок.

С развитием методов машинного обучения стали применяться модели классификации и регрессии, использующие признаки, извлекаемые из текста (например, n-граммы или метрики сходства) [4]. В рамках этих подходов ответ обучающегося может классифицироваться как «верный», «частично верный» или «неверный», либо оцениваться числовым баллом [5].

Однако, несмотря на достигнутые результаты, перечисленные методы имеют ряд ограничений: необходимость ручного задания правил или признаков, чувствительность к формулировкам и ограниченная способность учитывать сложные смысловые связи между элементами ответа.

1.2. Подходы на основе глубокого обучения и языковых моделей

Дальнейшее развитие в области ASAG связано с применением методов глубокого обучения, включая сверточные и рекуррентные нейронные сети, которые способны автоматически извлекать признаки из текстовых данных.

Особое распространение получили предварительно обученные языковые модели (PLM), такие как модели семейства BERT. Эти модели позволяют учитывать контекст и семантические связи между словами, что значительно повышает точность определения сходства между ответом обучающегося и эталонным ответом [6].

Современным этапом развития являются большие языковые модели (LLM), обладающие способностью понимать сложные структуры текста, интерпретировать смысл и генерировать развернутые ответы. В задачах ASAG такие модели могут выполнять сразу несколько функций: анализировать ответ обучающегося, определять степень его соответствия заданию и формировать объяснение выставленной оценки [7].

Однако, несмотря на высокий потенциал, применение LLM в задачах автоматической оценки сопровождается рядом проблем. Такие модели могут генерировать правдоподобные, но фактически неверные результаты, особенно в задачах, требующих специализированных знаний. Кроме того, они не всегда корректно учитывают критерии оценивания и могут игнорировать эталонные ответы, что приводит к несоответствию между оценкой системы и преподавателя.

1.3. Подход Retrieval-Augmented Generation (RAG)

Одним из перспективных направлений повышения точности и надежности систем ASAG является использование подхода Retrieval-Augmented Generation (RAG).

RAG представляет собой архитектуру, объединяющую два компонента: модуль поиска, извлекающий релевантную информацию из внешних источников, и генеративную LLM-модель, формирующую итоговый ответ на основе найденного контекста [8].

Принцип работы RAG включает три основных этапа:

1. индексирование данных (разбиение документов на фрагменты и преобразование их в векторное представление);
2. поиск релевантной информации на основе семантического сходства;
3. генерация ответа с использованием найденного контекста.

Ключевым преимуществом данного подхода является возможность использования внешних знаний при формировании ответа. Это позволяет снизить вероятность генерации недостоверной информации и повысить обоснованность принимаемых моделью решений.

2. Архитектура разработанной интеллектуальной системы

2.1. Общая архитектура системы

В рамках исследования разработана интеллектуальная система автоматической проверки развернутых ответов обучающихся с использованием RAG-архитектуры и больших языковых моделей. Предложенный подход направлен на повышение точности оценки ответов за счет интеграции внешних источников знаний и генеративных возможностей больших языковых моделей.

Функционирование системы реализуется в виде последовательного конвейера обработки данных, включающего этапы индексирования исходных данных, извлечения релевантной информации, ранжирования и генерации итоговой оценки. Общая архитектура предложенной системы представлена на рисунке 1.

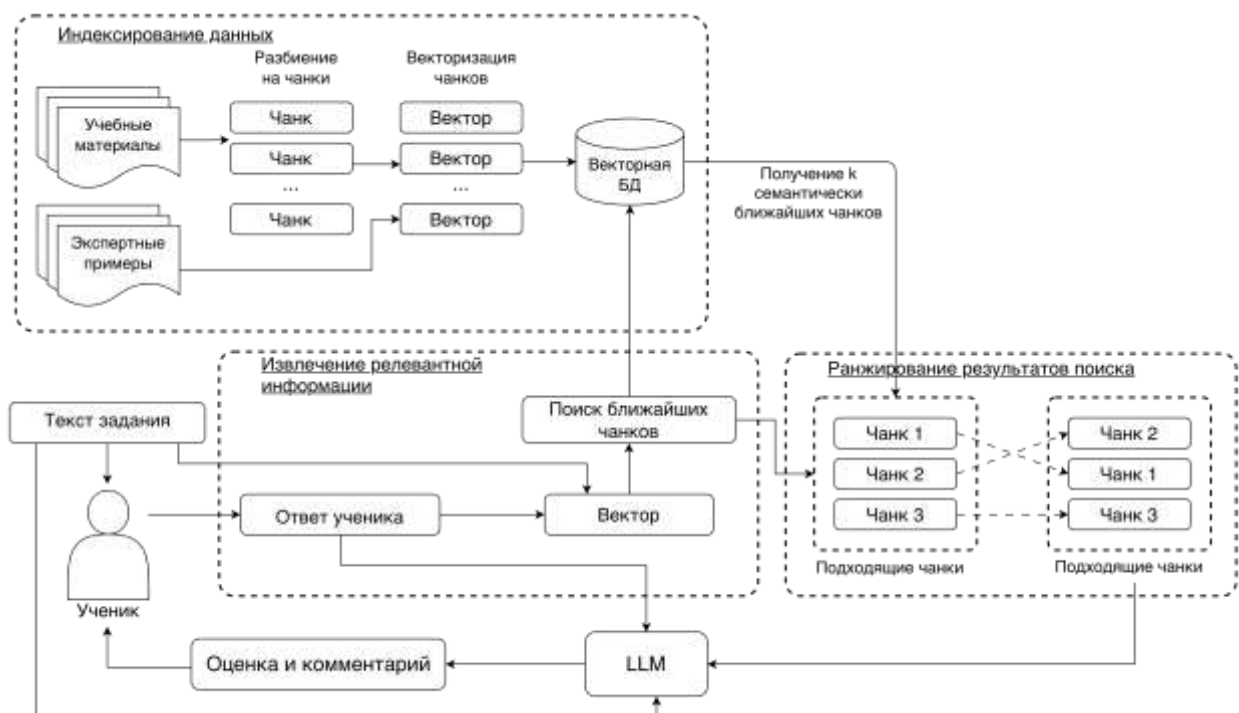


Рисунок 1. Архитектура системы автоматической проверки развернутых ответов обучающихся

Таким образом, система реализует расширенную RAG-архитектуру, адаптированную для задачи автоматической оценки ответов обучающихся.

2.2. Индексирование данных

Целью индексирования является преобразование исходных текстовых данных в структурированный и подходящий для поиска формат, чтобы обеспечить наиболее точное извлечение релевантной информации. В системе используются два типа данных: учебные материалы и экспертные примеры проверки работ.

Обработка учебных материалов включает:

- извлечение текста из исходных документов;

- разбиение текста на фрагменты (чанки);
- преобразование фрагментов в векторные представления;
- сохранение в векторной базе данных.

Разбиение текста на фрагменты осуществляется для обеспечения более эффективного поиска, поскольку языковые модели имеют ограниченную длину контекста [9]. Каждый фрагмент с помощью модели эмбедингов преобразуется в числовое векторное представление, после чего получает уникальный идентификатор. В конце обработанный фрагмент вместе со своим векторным представлением сохраняется в векторную базу данных.

Экспертные примеры индексируются в векторной базе данных как единый текстовый документ, включающий следующие элементы: текст задания, ответ обучающегося, оценку преподавателя, пояснение к оценке, критерии оценивания и эталонный ответ.

Такая структура позволяет использовать экспертные данные как дополнительный источник знаний при оценке новых ответов.

2.3. Извлечение релевантной информации

После этапа индексирования выполняется поиск фрагментов, наиболее релевантных анализируемому ответу обучающегося. Поиск осуществляется в векторной базе данных на основе семантического сходства между запросом и индексированными документами. Запрос формируется на основе текста задания и ответа обучающегося.

Особенностью предложенного подхода является использование двух независимых источников контекста: учебных материалов и экспертных примеров. Первый источник обеспечивает теоретическую основу для проверки ответа, а второй – доступ к ранее оцененным аналогичным работам.

Результатом работы модуля является набор кандидатов – текстовых фрагментов, потенциально полезных для последующей оценки ответа. Поиск реализуется с помощью векторного поиска и фильтрации по метаданным.

2.4. Ранжирование результатов поиска

Поскольку результаты первичного поиска могут содержать нерелевантные фрагменты, в системе предусмотрен дополнительный этап ранжирования. На этом этапе используется модель кросс-энкодера, которая оценивает степень соответствия между запросом и найденным фрагментом.

В отличие от классического поиска, основанного на сравнении векторных представлений, кросс-энкодер анализирует пару «запрос-документ» совместно, что позволяет более точно учитывать контекст и смысловые связи. В результате каждому кандидату присваивается оценка релевантности, на основе которой формируется упорядоченный список наиболее значимых фрагментов.

Применение этапа ранжирования позволяет существенно повысить качество отбора контекста, уменьшить количество нерелевантной информации и улучшить качество последующей генерации ответа.

2.5. Генерация оценки с использованием языковой модели

Финальный этап обработки выполняется с использованием большой языковой модели, которая выступает в роли генеративного компонента архитектуры RAG.

На вход модели подаются текст задания, ответ обучающегося и извлеченный контекст (учебные материалы и экспертные примеры), сформированный на предыдущих этапах. На основе этой информации модель формирует итоговую оценку и сопровождающее ее объяснение.

В системе используется внешняя языковая модель, доступ к которой осуществляется через API. Для обеспечения стабильности результатов используются фиксированные параметры генерации, включая температуру и ограничение длины ответа.

3. Экспериментальное исследование

3.1. Постановка эксперимента

Целью экспериментального исследования является комплексная оценка эффективности предложенной системы автоматической проверки ответов обучающихся.

В рамках исследования анализируются два взаимосвязанных аспекта:

- качество извлечения релевантной информации;
- точность итоговой оценки ответов.

Такой подход позволяет проследить влияние качества поиска контекста на результат генерации и определить вклад отдельных компонентов архитектуры.

3.2. Данные и подготовка выборки

В рамках эксперимента используются два типа данных: учебные материалы и экспертные примеры проверки заданий с развернутым ответом.

Учебные материалы включают учебные пособия по истории. Данные материалы используются в качестве источника теоретического контекста при работе с модулем поиска.

Экспертные примеры представлены в виде набора ответов обучающихся на задания с развернутым ответом в формате ОГЭ и ЕГЭ. Эти данные используются для формирования набора экспертных примеров, отражающих реальные критерии оценивания и практику проверки экзаменационных работ. Структура используемых экспертных данных представлена в таблице 1.

Таблица 1. Структура экспериментальных данных

Название поля	Описание
Предметная область	Учебная дисциплина
Экзамен	Формат экзамена
Номер задания	Порядковый номер задания в экзаменационном варианте
Задание	Вопрос, на который отвечает обучающийся
Максимальный балл	Наивысший балл, который можно получить за задание

Название поля	Описание
Критерии оценивания	Критерии оценивания для полного и частично верного ответа с указанием соответствующего балла
Ответ обучающегося	Развернутый ответ, данный обучающимся на задание
Оценка преподавателя	Балл, выставленный преподавателем за ответ обучающегося по критериям оценивания
Комментарий преподавателя	Пояснение преподавателя к выставленной оценке
Эталонный ответ	Пример возможного верного ответа

Всего используется 152 экспертных примера, из которых 101 пример относится к обучающей выборке, а 51 – к тестовой. Обучающая выборка используется для формирования базы экспертных примеров, индексируемых в векторной базе данных, а тестовая – для оценки качества работы системы.

3.3. Оценка качества извлечения релевантной информации

Для оценки влияния параметров поиска был проведен факторный эксперимент, в котором варьировались следующие параметры:

- модель эмбедингов;
- размер текстового фрагмента (чанка);
- стратегия поиска.

Рассматривались три модели эмбедингов (multilingual-e5-large, bge-m3, paraphrase-multilingual-MiniLM-L12-v2), три значения размера чанка (500, 1000 и 1500 символов) и три стратегии поиска (sparse, dense, hybrid). Таким образом, всего было рассмотрено 27 комбинаций параметров.

Также в качестве модели повторного ранжирования использовалась VAAI/bge-reranker-v2-m3, применяемая к множеству кандидатов, полученных на этапе поиска. Такой подход позволяет уточнить порядок релевантных документов и повысить точность поиска информации.

Для оценки качества извлечения релевантной информации использовались метрики, учитывающие как полноту поиска, так и качество ранжирования результатов:

1. Recall@k – доля релевантных документов, найденных системой среди первых k результатов, относительно общего числа релевантных документов.

2. MRR – среднее значение обратного ранга первого релевантного документа в результатах поиска.

3. nDCG@k – метрика качества ранжирования, которая учитывает не только наличие релевантных документов, но и их позиции в топ-k.

Основные результаты оценки качества извлечения релевантной информации представлены в таблице 2, при этом приведены не все исследованные комбинации параметров, а только конфигурации, отражающие основные выявленные закономерности.

Таблица 2. Результаты эксперимента по влиянию параметров поиска

Модель эмбедингов	Размер чанка	Стратегия поиска	Реранкер	MRR	Recall@5	nDCG@5
MiniLM	1500	hybrid	-	0.3067	0.40	0.3305
e5-large	1500	dense	-	0.5747	0.80	0.6309
bge-m3	500	dense	-	0.5660	0.80	0.6239
e5-large	1000	hybrid	-	0.4073	0.80	0.5043
bge-m3	1000	hybrid	-	0.4680	0.80	0.5506
e5-large	1500	dense	bge-reranker	0.7467	0.96	0.8006
bge-m3	1000	dense	bge-reranker	0.7300	0.92	0.7791

Как видно из таблицы 2, модель paraphrase-multilingual-MiniLM-L12-v2 существенно уступает другим по всем метрикам ($MRR \approx 0.31$, $Recall@5 \approx 0.4$). В то же время модели multilingual-e5-large и bge-m3 демонстрируют сопоставимо высокие результаты ($MRR \approx 0.56-0.57$, $Recall@5 \approx 0.8$).

На рисунке 2 представлено сравнение значений MRR для различных конфигураций.



Рисунок 2. Сравнение значений MRR для различных конфигураций поиска

Анализ показывает, что наибольшее влияние на качество извлечения оказывает применение модели повторного ранжирования. Использование VAAI/bge-reranker-v2-m3 приводит к увеличению MRR с 0.57 до 0.75, а Recall@5 – с 0.8 до 0.96. Это свидетельствует о том, что релевантные документы уже присутствуют среди кандидатов, однако их положение в ранжированном списке требует дополнительного уточнения.

Также установлено, что увеличение размера чанка до 1000-1500 символов способствует повышению качества поиска за счет расширения контекста. Наилучшие значения метрик демонстрирует плотный поиск

(dense), однако гибридный подход (hybrid) также показывает высокое качество извлечения релевантной информации.

Таким образом, анализ показывает, что ключевыми факторами, определяющими качество извлечения, являются выбор модели эмбедингов и применение повторного ранжирования. Оптимальной конфигурацией является использование модели multilingual-e5-large, размера чанка 1000-1500 символов и плотного поиска с последующим повторным ранжированием.

3.4. Оценка точности автоматической проверки ответов

Для количественной оценки точности автоматической проверки использовались метрики, отражающие как строгое совпадение оценок, так и степень отклонения от экспертного результата:

1. Accuracy – отношение числа ответов, для которых оценка модели полностью совпадает с экспертной, к общему числу проверенных ответов.
2. MAE – среднее значение абсолютного отклонения предсказанной оценки от экспертной, характеризующее величину ошибки модели.
3. Within-1 Accuracy – доля ответов, для которых разница между оценкой модели и экспертной не превышает одного балла; метрика отражает практическую приемлемость результатов в образовательном контексте.

Для оценки вклада отдельных компонентов в системе реализовано несколько режимов работы. Их сравнительная характеристика представлена в таблице 3.

Таблица 3. Режимы работы системы

Режим	Используемый контекст
No RAG	отсутствует
Textbook RAG	учебные материалы
Full RAG	учебные материалы + экспертные примеры

В режимах Textbook RAG и Full RAG используется единый процесс извлечения релевантной информации: поиск кандидатов в векторной базе и последующее повторное ранжирование.

Эксперимент по автоматической проверке ответов проводился с использованием большой языковой модели Gemini 2.5 Flash. Результаты сравнительного анализа качества автоматической проверки представлены в таблице 4.

Таблица 4. Сравнение режимов работы системы по метрикам качества

Режим	Accuracy	MAE	Within-1
No RAG	0.67	0.37	0.93
Textbook RAG	0.74	0.26	0.96
Full RAG	0.80	0.21	0.98

Как видно из таблицы 4, использование RAG приводит к улучшению качества автоматической оценки. Базовая модель без внешнего контекста (No

RAG) демонстрирует наименьшую точность (accuracy = 0.67) и наибольшую ошибку (MAE = 0.37), что указывает на ограниченную способность модели воспроизводить экспертную логику оценивания при отсутствии дополнительной информации.

Добавление учебных материалов (Textbook RAG) повышает точность до 0.74 и снижает MAE до 0.26. Таким образом, использование учебных материалов уменьшает количество ошибок и повышает согласованность оценок с экспертными.

Наилучшие показатели достигаются в режиме Full RAG, в котором дополнительно используются экспертные примеры: accuracy = 0.80, MAE = 0.21. Это показывает, что экспертные примеры позволяют модели более точно воспроизводить логику оценивания преподавателя.

Значения Within-1 во всех режимах превышают 0.93, что свидетельствует о высокой практической согласованности оценок: в большинстве случаев отклонение от экспертной оценки не превышает одного балла.

Распределение абсолютной ошибки автоматической проверки по режимам работы системы представлено на рисунке 3. Как видно из графика, использование RAG увеличивает количество случаев полного совпадения оценок модели и преподавателя ($|y - \hat{y}| = 0$) и уменьшает количество ошибок с отклонением на 1-2 балла. Наилучшие результаты демонстрирует режим Full RAG, характеризующийся наибольшей концентрацией предсказаний в области нулевой ошибки.

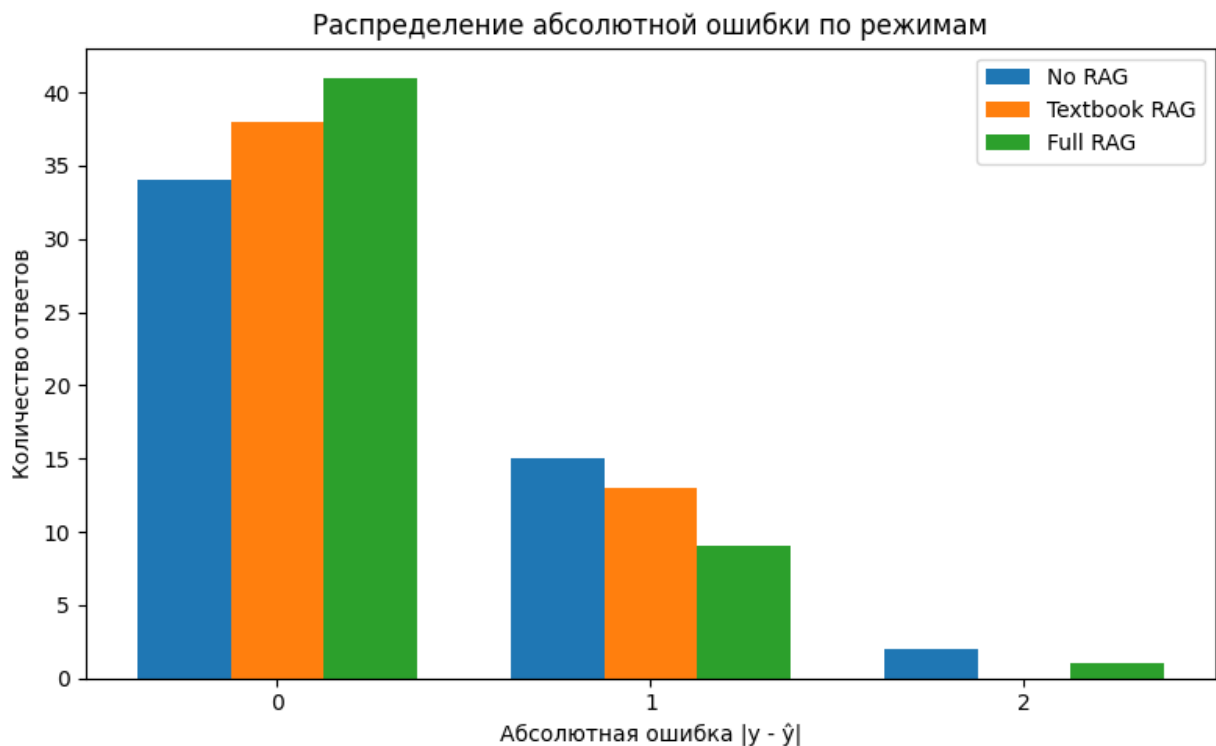


Рисунок 3. Распределение абсолютной ошибки автоматической проверки по режимам работы системы

Матрицы соответствия экспертных и автоматических оценок, представленные на рисунке 4, подтверждают рост качества при переходе к более сложным режимам работы системы. Для режима No RAG наблюдается большее количество расхождений между оценками модели и преподавателя. Использование учебных материалов уменьшает количество ошибок, а режим Full RAG демонстрирует наибольшую концентрацию значений на главной диагонали матрицы, что соответствует точному совпадению автоматических и экспертных оценок.

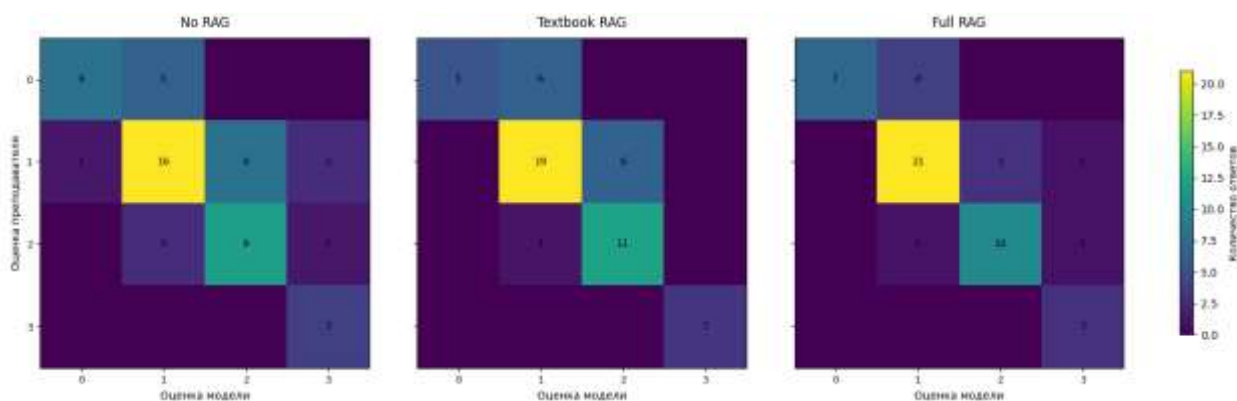


Рисунок 4. Матрица соответствия экспертных и автоматических оценок по режимам работы системы

Заклучение

В работе представлена интеллектуальная система автоматической проверки ответов обучающихся на основе RAG-архитектуры и больших языковых моделей. Предложенный подход сочетает использование учебных материалов и экспертных примеров проверки, что позволяет учитывать как предметные знания, так и логику оценивания ответов преподавателем.

Проведенное исследование показывает, что качество автоматической проверки существенно зависит от релевантности контекста, передаваемого языковой модели. При отсутствии внешнего контекста модель чаще допускает ошибки, связанные с неполным учетом критериев оценивания и неоднозначной интерпретацией ответа обучающегося.

Наибольший вклад в повышение качества извлечения релевантной информации вносит этап повторного ранжирования. Результаты показывают, что основная проблема заключается не столько в отсутствии релевантных документов среди найденных кандидатов, сколько в их исходном порядке. Использование кросс-энкодера позволяет повысить качество ранжирования и улучшить итоговое качество автоматической проверки.

Дополнительное использование экспертных примеров оказывает положительное влияние на согласованность оценок модели с экспертными, поскольку позволяет модели учитывать не только предметные знания, но и реальную логику оценивания ответов преподавателем.

Стоит отметить, что предложенный подход имеет ряд ограничений. Качество работы системы зависит от полноты и качества базы экспертных примеров и используемых учебных материалов. Кроме того, исследование проводилось только на заданиях по истории, что требует дополнительной проверки переносимости результатов на другие предметные области.

Список литературы:

1. Mohler, M. Text-to-Text Semantic Similarity for Automatic Short Answer Grading / M. Mohler, R. Mihalcea. – Text : unmediated // Proceedings of the 12th Conference of the European Chapter of the ACL. – Athens, Greece : Association for Computational Linguistics, 2009. – P. 567-575.
2. A Reliable Approach to Automatic Assessment of Short Answer Free Responses / L. F. Bachman, N. Carr, G. Kamei [et al.]. – Text : unmediated // The 17th International Conference on Computational Linguistics: Project Notes. – Los Angeles, California, USA : Association for Computational Linguistics, 2002.
3. Bukai, O. An Automated Short-Free-Text Scoring System: Development and Assessment / O. Bukai, R. Pokorny, J. Haynes. – Text : unmediated // In Proceedings of the Twentieth Interservice/Industry Training, Simulation, and Education Conference. – Orlando, Florida, USA : National Training and Simulation Association, 2016. – P. 1-11.
4. Chaudhari, R. Deep Learning in Automated Short Answer Grading: A Comprehensive Review / R. Chaudhari, M. Patel. – Text : unmediated // International Conference on Multidisciplinary Approach in Engineering, Technology and Management for Sustainable Development. – 2024. – № 65. – P. 1-12.
5. Zhu, X. Automatic Short-Answer Grading via BERT-Based Deep Neural Networks / X. Zhu, H. Wu, L. Zhang. – Text : unmediated // IEEE Transactions on Learning Technologies. – 2022. – № 15 (3). – P. 364-375.
6. Surya, K. Deep learning for Short Answer Scoring / K. Surya, E. Gayakwad, M. K. Nallakaruppan. – Text : unmediated // International Journal of Recent Technology and Engineering. – 2019. – № 7 (6). – P. 1712-1715.
7. Nye, B. Generative Large Language Models for Dialog-Based Tutoring: An Early Consideration of Opportunities and Concerns / B. Nye, D. Mee, M. G. Core. – Text : unmediated // AIED 2023 Workshop on Empowering Education with LLMs. – Los Angeles, California, USA : CEUR Workshop Proceedings, 2023. – P. 1-14.
8. Retrieval-augmented generation for knowledge-intensive NLP tasks / P. Lewis, E. Perez, A. Piktus [et al.]. – Text : unmediated // Proceedings of the 34th International Conference on Neural Information Processing Systems. – Red Hook, NY, USA : Curran Associates Inc., 2020. – P. 9459-9474.
9. Das, B. C. Security and Privacy Challenges of Large Language Models: A Survey / B. C. Das, M. H. Amini, Y. Wu. – Text : unmediated // ACM Computing Surveys. – 2025. – № 57 (6). – P. 1-39.

УДК 519.237.5:004.85

Филиппенко Д. Р.

МИРЭА – Российский технологический университет, Москва, Россия

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ РЕГУЛЯРИЗАЦИИ В ЗАДАЧАХ ЛИНЕЙНОЙ РЕГРЕССИИ ПРИ МУЛЬТИКОЛЛИНЕАРНОСТИ ПРИЗНАКОВ

Аннотация. В работе рассмотрены три классических метода регуляризации линейной регрессии — гребневая регрессия (Ridge), LASSO и Elastic Net. Приведено теоретическое описание методов, выписаны соответствующие задачи оптимизации и их геометрическая интерпретация. Проведён вычислительный эксперимент на синтетических данных с заданной степенью мультиколлинеарности признаков. На основе среднеквадратичной ошибки на тестовой выборке и числа отобранных признаков получены практические рекомендации по выбору метода регуляризации в зависимости от свойств исходных данных.

Ключевые слова: линейная регрессия, регуляризация, мультиколлинеарность, гребневая регрессия, LASSO, Elastic Net, машинное обучение, отбор признаков

Введение

Линейная регрессия остаётся одним из наиболее востребованных инструментов прикладной статистики и машинного обучения благодаря простоте интерпретации коэффициентов и низкой вычислительной стоимости [1]. Тем не менее, классическая оценка по методу наименьших квадратов (МНК) теряет устойчивость в условиях, когда признаки сильно коррелированы между собой. Это явление, известное как мультиколлинеарность, приводит к плохой обусловленности матрицы Грама и, как следствие, к большой дисперсии оценок коэффициентов модели [2].

Стандартным подходом к преодолению указанной проблемы является введение в функционал потерь штрафного слагаемого — регуляризатора. В зависимости от выбора регуляризатора получаются различные методы: гребневая регрессия (Ridge), LASSO и их линейная комбинация Elastic Net. Каждый из методов обладает собственными статистическими и вычислительными свойствами, что делает актуальной задачу их сравнительного анализа в практических условиях.

Целью настоящей работы является сравнение указанных методов регуляризации по точности предсказания и способности к отбору признаков на модельных данных с контролируемым уровнем мультиколлинеарности. Для достижения поставленной цели в работе решаются следующие задачи: дано формальное описание методов; реализован вычислительный эксперимент; полученные результаты обобщены в виде практических рекомендаций.

1. Постановка задачи и математическая модель

Пусть имеется выборка из n наблюдений, каждое из которых характеризуется вектором признаков $x \in \mathbb{R}^p$ и значением целевой переменной $y \in \mathbb{R}$. Объединим признаки в матрицу $X \in \mathbb{R}^{n \times p}$, а значения целевой переменной — в вектор $y \in \mathbb{R}^n$. Будем считать, что столбцы X центрированы и нормированы, а вектор y центрирован, что позволяет исключить из рассмотрения свободный член.

Классическая оценка МНК получается решением задачи минимизации суммы квадратов невязок:

$$\beta_{\text{МНК}} = \arg \min \|y - X\beta\|_2^2, \quad (1)$$

и в явном виде задаётся выражением $\beta_{\text{МНК}} = (X^T X)^{-1} X^T y$. Если матрица $X^T X$ плохо обусловлена, малые возмущения данных приводят к большим изменениям оценок коэффициентов, что снижает практическую ценность модели.

Гребневая регрессия (Ridge) была предложена А. Хоэрлом и Р. Кеннардом в 1970 г. [3] и состоит в добавлении к функционалу (1) квадратичного штрафа на коэффициенты:

$$\beta_{\text{Ridge}} = \arg \min \{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \}, \quad (2)$$

где $\lambda \geq 0$ — параметр регуляризации. Соответствующее решение имеет явный вид $\beta_{\text{Ridge}} = (X^T X + \lambda I)^{-1} X^T y$. Добавление λI к матрице $X^T X$ гарантирует положительную определённость и улучшает обусловленность задачи.

Метод LASSO (Least Absolute Shrinkage and Selection Operator), предложенный Р. Тибширани в 1996 г. [4], использует штраф по норме ℓ_1 :

$$\beta_{\text{LASSO}} = \arg \min \{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \}. \quad (3)$$

В отличие от гребневой регрессии, LASSO не имеет замкнутого решения, однако обладает важным свойством: при достаточно большом значении λ часть коэффициентов в точности обращается в ноль. Тем самым LASSO одновременно решает задачи оценки коэффициентов и отбора признаков. Геометрически это объясняется тем, что шар нормы ℓ_1 имеет угловые точки на осях координат, в которых вероятна точка касания эллипсов уровня функционала потерь.

Метод Elastic Net, введённый Х. Цоу и Т. Хасты в 2005 г. [5], сочетает оба типа регуляризации:

$$\beta_{\text{EN}} = \arg \min \{ \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \}. \quad (4)$$

Параметры λ_1 и λ_2 управляют соотношением между разреживающим и сглаживающим эффектами. Elastic Net преодолевает известное ограничение LASSO: при наличии групп сильно коррелированных признаков LASSO имеет тенденцию выбирать только один из них, тогда как Elastic Net включает группу целиком.

2. Вычислительный эксперимент

Для сравнения методов сгенерирована синтетическая выборка по модели $y = X\beta^* + \varepsilon$, где истинный вектор коэффициентов $\beta^* \in \mathbb{R}^{20}$ имеет только пять

ненулевых компонент со значением 1,5, а остальные равны нулю. Шум $\varepsilon \sim N(0; \sigma^2)$ задаётся с $\sigma = 1$. Объём обучающей выборки $n = 200$, тестовой — 1000.

Признаки моделировались как многомерное нормальное распределение с ковариационной матрицей $\Sigma_{ij} = \rho^{|i-j|}$, где $\rho \in \{0,1; 0,5; 0,9\}$ — параметр, задающий уровень мультиколлинеарности. Параметры регуляризации λ для всех методов выбирались по схеме 10-кратной перекрёстной проверки на обучающей выборке. Вычислительный эксперимент реализован средствами библиотеки scikit-learn языка Python [6].

Качество моделей оценивалось по среднеквадратичной ошибке (MSE) на тестовой выборке и по числу отобранных признаков, то есть числу коэффициентов, превышающих по модулю порог 10^{-4} . Результаты приведены в таблице 1.

Таблица 1. Сравнение методов регуляризации при различных уровнях мультиколлинеарности

Метод	ρ	MSE	Число признаков	Время, с
МНК	0,1	1,12	20	0,01
МНК	0,5	1,47	20	0,01
МНК	0,9	3,85	20	0,01
Ridge	0,1	1,10	20	0,05
Ridge	0,5	1,28	20	0,05
Ridge	0,9	1,71	20	0,05
LASSO	0,1	1,08	6	0,08
LASSO	0,5	1,21	7	0,08
LASSO	0,9	1,93	4	0,09
Elastic Net	0,1	1,09	7	0,12
Elastic Net	0,5	1,19	9	0,12
Elastic Net	0,9	1,54	8	0,13

Анализ данных таблицы 1 позволяет сделать ряд содержательных выводов. При слабой коррелированности признаков ($\rho = 0,1$) все методы демонстрируют сопоставимое качество предсказания, и метод МНК остаётся вполне приемлемым. По мере роста мультиколлинеарности ($\rho = 0,5$ и особенно $\rho = 0,9$) ошибка МНК резко возрастает, тогда как регуляризованные методы сохраняют устойчивость.

Метод LASSO успешно решает задачу отбора признаков, отбирая 4–7 коэффициентов при истинных пяти ненулевых, однако при $\rho = 0,9$ уступает Ridge и Elastic Net по точности из-за упомянутой проблемы выбора одного

признака из коррелированной группы. Метод Elastic Net демонстрирует наилучший баланс: при сильной мультиколлинеарности он показывает MSE, близкое к Ridge, но при этом обеспечивает разреженность модели, что важно для интерпретации.

Заклучение

В работе проведён сравнительный анализ методов Ridge, LASSO и Elastic Net в задаче линейной регрессии при наличии мультиколлинеарности признаков. Дано формальное описание методов, выписаны соответствующие задачи выпуклой оптимизации и приведена их геометрическая интерпретация.

На основе вычислительного эксперимента сформулированы следующие практические рекомендации. При отсутствии или слабой коррелированности признаков различия между методами невелики, и выбор может определяться вычислительной эффективностью. При умеренной мультиколлинеарности и необходимости устойчивых оценок без отбора признаков предпочтительна гребневая регрессия. Если основной задачей является интерпретация модели и выявление значимых признаков при умеренной коррелированности — целесообразно применение LASSO. При высокой мультиколлинеарности с группами связанных признаков наилучшие результаты по совокупности критериев показывает Elastic Net.

Полученные результаты согласуются с теоретическими свойствами рассмотренных методов и могут быть использованы при выборе модели регрессии в прикладных задачах анализа данных.

Список литературы:

1. Хасты Т., Тибширани Р., Фридман Дж. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование. — М.: Вильямс, 2020. — 768 с. ISBN 978-5-907144-42-2
2. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Вильямс, 2017. — 912 с.
3. Hoerl A. E., Kennard R. W. Ridge regression: biased estimation for nonorthogonal problems // *Technometrics*. — 1970. — Vol. 12, № 1. — P. 55–67.
4. Tibshirani R. Regression shrinkage and selection via the lasso // *Journal of the Royal Statistical Society. Series B*. — 1996. — Vol. 58, № 1. — P. 267–288.
5. Zou H., Hastie T. Regularization and variable selection via the elastic net // *Journal of the Royal Statistical Society. Series B*. — 2005. — Vol. 67, № 2. — P. 301–320.
6. Pedregosa F. et al. Scikit-learn: Machine learning in Python // *Journal of Machine Learning Research*. — 2011. — Vol. 12. — P. 2825–2830.

УДК 004.7:621.315

*Кудашкина М. В., Бочкин А. Ю., Гусаков А. Н., Сеськин В. П.
МГУ им. Н. П. Огарева, Саранск, Россия*

РАЗРАБОТКА СТРУКТУРИРОВАННОЙ КАБЕЛЬНОЙ СИСТЕМЫ В ОБЩЕОБРАЗОВАТЕЛЬНОМ УЧРЕЖДЕНИИ

Аннотация. В статье рассматриваются теоретические и практические аспекты проектирования структурированной кабельной системы (СКС) для общеобразовательных учреждений. Анализируются нормативно-правовые требования, предъявляемые к СКС в школах, основные архитектурные решения, требования к компонентам системы и этапы реализации проекта. Особое внимание уделяется специфике образовательных учреждений как объектов проектирования, включая требования к пропускной способности, пожарной безопасности и возможности масштабирования. Статья предназначена для специалистов в области проектирования инженерных систем, руководителей образовательных учреждений и студентов соответствующих специальностей.

Ключевые слова: структурированная кабельная система, общеобразовательное учреждение, горизонтальная подсистема, магистральная подсистема, телекоммуникационный шкаф, информационная розетка, стандарты СКС.

Введение

Современное общеобразовательное учреждение невозможно представить без развитой информационно-коммуникационной инфраструктуры. Внедрение электронных образовательных ресурсов, систем видеонаблюдения, контроля доступа, автоматизации управления школой и других цифровых сервисов требует создания надежной, масштабируемой и безопасной кабельной основы. Именно структурированная кабельная система (СКС) выступает тем фундаментом, на котором базируется вся цифровая экосистема школы.

Актуальность темы обусловлена несколькими факторами. Во-первых, происходит активное обновление нормативной базы, регламентирующей требования к ИТ-инфраструктуре образовательных организаций. Во-вторых, школы сталкиваются с необходимостью модернизации существующих кабельных сетей, которые часто не соответствуют современным требованиям по пропускной способности и безопасности. В-третьих, растет количество подключаемых устройств в рамках концепции «цифровая школа», что требует пересмотра подходов к проектированию.

Цель данной статьи – систематизировать знания о проектировании СКС для общеобразовательных учреждений, выделить ключевые требования и предложить практические рекомендации, основанные на актуальных нормативных документах.

Проектирование СКС в общеобразовательных учреждениях регулируется комплексом нормативных документов федерального и отраслевого уровней. Понимание этой базы является необходимым условием для разработки корректного и безопасного проекта.

Для образовательных организаций особое значение имеют отраслевые стандарты. Примером может служить Приложение 2 к приказу Департамента образования и науки города Москвы и Департамента информационных технологий города Москвы от 28 февраля 2024 г., которое устанавливает «Нормы и требования к ИТ-инфраструктуре ОО» в рамках Московской электронной школы. Данный документ детализирует требования к СКС, включая:

- состав подсистем (внешняя магистральная, внутренняя магистральная, горизонтальная);
- требования к кабельной продукции (категория не ниже 5е, цельно-медные жилы);
- требования к информационным розеткам и коммутационному оборудованию;
- требования к помещению серверной (аппаратной).

Стандартная СКС здания общеобразовательного учреждения в соответствии с ГОСТ Р 53246-2008 включает три основные подсистемы:

1) Внешняя магистральная подсистема предназначена для организации каналов связи передачи данных между зданием школы и внешними сетями (провайдер интернета, городская телефонная сеть, вышестоящие образовательные сети). Обычно эта подсистема реализуется с использованием оптоволоконного кабеля, обеспечивающего высокую пропускную способность и гальваническую развязку.

2) Внутренняя магистральная подсистема обеспечивает связь внутри здания между главным распределительным узлом (серверной) и этажными телекоммуникационными шкафами (ЭТШ). Для этой подсистемы также целесообразно использовать оптоволокно, особенно в многоэтажных школах, где длина кабельных трасс может превышать ограничения для медных линий (90 метров для горизонтальной подсистемы).

3) Горизонтальная подсистема соединяет этажные коммутационные шкафы с информационными розетками (ИР) в рабочих зонах – учебных кабинетах, административных помещениях, библиотеке и т.д. Каждая горизонтальная линия выполняется отдельным кабелем «витая пара» категории не ниже 5е (рекомендуется категория 6 или 6А), длина линии не должна превышать 90 метров от ИР до патч-панели.

Для общеобразовательных учреждений наиболее распространенной является иерархическая звездообразная топология. В рамках этой топологии:

- каждый рабочий терминал подключается отдельным кабелем к коммутатору этажного шкафа;

– этажные коммутаторы соединяются с центральным коммутатором (или коммутатором распределения) в серверной;

– центральное оборудование подключается к внешним каналам связи.

Такая топология обеспечивает высокую надежность (отказ одного сегмента не влияет на работу других), удобство администрирования (локализация неисправностей) и возможность масштабирования.

При большом количестве этажей или при наличии нескольких корпусов может применяться кольцевая топология для магистральных соединений, обеспечивающая резервирование каналов связи между коммутационными узлами.

Помещение серверной (аппаратной) является ключевым элементом СКС. К его размещению предъявляются следующие требования:

– помещение должно быть предназначено исключительно для обеспечения работы телекоммуникационных систем;

– запрещено размещать в нем оборудование посторонних инженерных систем (водоснабжения, водоотведения и т.п.);

– рекомендуется располагать серверную на первом этаже здания;

– необходимо предусмотреть гидроизоляцию при расположении ниже горизонта воды;

– требуется обеспечить достаточную вентиляцию и кондиционирование для поддержания температурного режима работы оборудования.

Этажные телекоммуникационные шкафы (ЭТШ) рекомендуется размещать в центре обслуживаемой зоны для минимизации длины горизонтальных кабелей. Шкафы должны быть оборудованы запирающимися дверцами для предотвращения несанкционированного доступа.

Выбор кабеля для СКС школы определяется несколькими факторами: требуемой пропускной способностью, длиной линий, уровнем электромагнитных помех и требованиями пожарной безопасности.

В учебных заведениях предпочтительна открытая прокладка кабеля в кабель-каналах, что обеспечивает доступ для обслуживания и возможность модернизации.

Тестирование должно выполняться в соответствии со стандартами TIA/EIA-568-C.2, IEEE 802.3, ISO 11801, ГОСТ Р 53245 и ГОСТ Р 53246. Результаты тестирования фиксируются в отчете и являются частью исполнительной документации. Тестирование включает проверку:

– целостности линий и правильности распиновки;

– затухания сигнала (Insertion Loss);

– обратных и перекрестных помех (NEXT, FEXT, PS NEXT);

– длины линии;

– импеданса.

Обязательным завершающим этапом является оформление исполнительной документации, включающей:

- принципиальные схемы СКС;
- поэтажные планы с указанием трасс прокладки кабеля;
- маркировочные таблицы (соответствие портов патч-панелей информационным розеткам);
- протоколы тестирования всех линий;
- спецификации использованного оборудования и материалов.

Заключение

Разработка структурированной кабельной системы для общеобразовательного учреждения представляет собой комплексную инженерную задачу, требующую учета многочисленных факторов: от соблюдения нормативных требований до прогнозирования перспективных потребностей в пропускной способности.

Анализ нормативной базы показывает, что сегодня в России сформирована достаточно полная система требований к СКС школ, включающая как общестроительные нормы (СП 134.13330.2012), так и специализированные отраслевые стандарты. Ключевыми требованиями являются: обеспечение пропускной способности не менее 1 Гбит/с на рабочее место, применение цельно-медных кабелей категории не ниже 5е (рекомендуется 6/6А), использование кабелей с пониженным дымо- и газовыделением (LSZH), соблюдение ограничений по длине горизонтальных линий (90 метров).

Архитектура СКС школы должна строиться на трехуровневой модели (внешняя магистраль – внутренняя магистраль – горизонтальная подсистема) с иерархической звездообразной топологией. Особое внимание следует уделять правильному размещению телекоммуникационных узлов и расчету пропускной способности с учетом пиковых нагрузок и перспектив масштабирования.

Практическая реализация проекта требует профессионального выполнения всех этапов: от предпроектного обследования до тестирования и сдачи исполнительной документации. Инвестиции в качественно спроектированную и смонтированную СКС окупаются долгим сроком службы (15–20 лет), надежностью и возможностью модернизации без замены кабельной инфраструктуры.

В условиях цифровой трансформации образования правильно спроектированная СКС становится не просто инженерной системой, а стратегическим активом школы, обеспечивающим основу для внедрения современных образовательных технологий и создания комфортной и безопасной цифровой среды.

Список литературы:

1. СП 134.13330.2012 «Системы электросвязи зданий и сооружений. Основные положения проектирования» (с изменениями №1, №2, №3). — М.: Минрегион России, 2012.

2. Приложение 2 к приказу Департамента образования и науки города Москвы и Департамента информационных технологий города Москвы от 28 февраля 2024 г. № 177/64-16-77/24 «Нормы и требования к ИТ-инфраструктуре ОО».

3. ГОСТ Р 53246-2008 «Информационная технология. Структурированные кабельные системы. Основные положения и требования». — М.: Стандартиформ, 2008.

4. ГОСТ Р 53245-2008 «Информационная технология. Структурированные кабельные системы. Методы испытания кабельной системы». — М.: Стандартиформ, 2008.

5. ГОСТ 31565-2012 «Кабельные изделия. Требования пожарной безопасности». — М.: Стандартиформ, 2012.

6. ISO/IEC 11801:2017 «Information technology — Generic cabling for customer premises». — ISO/IEC, 2017.

7. TIA/EIA-568-C.2 «Balanced Twisted-Pair Telecommunications Cabling and Components Standards». — TIA, 2009.

8. «Проектирование СКС в школе: требования, этапы и особенности» // ZSComp — информационный портал, 2023.

УДК 681.5.08:627.8

Ильин А. В.¹

Научный руководитель: Шахмаев Т. И.²

¹Саяно-Шушенский филиал Сибирского федерального университета, Черемушки, Россия

²Филиал ПАО «РусГидро – Саяно-Шушенская ГЭС им. П. С. Непорожнего», Черемушки, Россия

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРОТОТИПА УНИВЕРСАЛЬНОЙ СИСТЕМЫ АВТОМАТИЗАЦИИ ГИДРОТЕХНИЧЕСКИХ СООРУЖЕНИЙ

Аннотация: в статье представлена практическая реализация прототипа трёхуровневой универсальной системы автоматизации (АСУТП) для задач мониторинга технологических параметров. Реализована сквозная цепочка «датчики → контроллер → программный шлюз (фильтр-прибор) → СУБД», обеспечивающая сбор измерений в реальном времени, нормализацию разнородных данных, присвоение признака качества и долговременное хранение временных рядов в PostgreSQL. В качестве полевого уровня использованы датчики DHT11 (температура/влажность) и VL6180X (оптическое измерение дистанции), опрашиваемые контроллером Arduino Uno с периодом 1 с. Передача на верхний уровень выполнена по USB-Serial в формате однострочного JSON. Информационный уровень реализован на Python в виде постоянно работающего шлюза gateway.py, осуществляющего парсинг, фильтрацию, привязку полей к каналам измерений и вставку данных в таблицу измерений PostgreSQL. Приведены структура базы данных, формат сообщений, принцип контроля качества и способы проверки непрерывности записи.

Ключевые слова: АСУТП, мониторинг, Arduino, DHT11, VL6180X, JSON, gateway/ETL, PostgreSQL, временной ряд, качество данных.

Практическая реализация прототипа

Практическая часть работы выполнена как лабораторный прототип трёхуровневой системы автоматизации, предназначенный для демонстрации принципов универсального сбора, унификации и хранения измерений. Целью практической реализации является подтверждение работоспособности цепочки передачи данных от полевого уровня до СУБД с обеспечением непрерывной записи и возможностью последующего мониторинга и анализа. В качестве модельных параметров выбраны температура, относительная влажность и дистанция до объекта, что является типовым набором для задач технического контроля и мониторинга.

Уровень 1 (полевой): датчики и первичные измерения

Нижний уровень прототипа представлен датчиком температуры/влажности **DHT11** и оптическим датчиком расстояния **VL6180X**. DHT11 формирует цифровой сигнал по однопроводному интерфейсу и подключён к цифровому входу **D2** контроллера. Опрос

выполняется с периодом **1 с**, что соответствует техническим ограничениям данного датчика и достаточно для мониторинга микроклимата.

VL6180X подключён по интерфейсу **I2C** к линиям **SDA/SCL (A4/A5)** на **Arduino Uno**. Датчик возвращает измеренную дистанцию в миллиметрах, а также **код состояния** (status), который используется для определения корректности результата (например, наличие ошибки измерения, отсутствие объекта в зоне контроля и т.п.). Наличие статусного кода позволяет реализовать первичный контроль качества ещё до передачи данных на верхний уровень.

Уровень 2 (контроллерный): Arduino как агрегатор и унификатор данных

Контроллерный уровень реализован на **Arduino Uno**, которая выполняет функции опроса датчиков, первичной проверки и формирования унифицированного пакета телеметрии. На этапе отладки рассматривалась схема с несколькими платами и обменом между ними по I2C в конфигурации master–slave. Однако практические испытания показали, что в лабораторных условиях общая I2C-шина чувствительна к качеству соединений и особенностям подключения VL6180X: при некорректной коммутации датчик может приводить к «падению» шины и потере обмена. Для повышения воспроизводимости эксперимента и устойчивости системы была использована упрощённая конфигурация: **оба датчика опрашиваются одной платой Arduino**, а передача на информационный уровень выполняется по USB-Serial.

Выходной формат контроллера — **однострочный JSON** (одна строка = один кадр телеметрии), что упрощает разбор и обеспечивает расширяемость структуры без изменения базовой логики шлюза. Пример структуры кадра:

```
{ "seq":1, "t_c":25.3, "rh_pct":20.0, "dist_mm":6, "dht_ok":1, "vl_ok":1, "vl_status":0, "quality_flag":0 }
```

где:

seq — номер кадра (контроль последовательности и пропусков);

t_c, rh_pct, dist_mm — значения температуры, влажности и дистанции;

dht_ok, vl_ok — признаки успешного опроса каждого датчика;

vl_status — код состояния VL6180X;

quality_flag — агрегированный признак качества (используется далее на уровне БД/аналитики).

Частота формирования кадров составляет **1 Гц**. Это означает, что при разложении кадра на отдельные каналы измерений в БД формируется **три записи в секунду** (температура, влажность, дистанция).

Уровень 3 (информационный): Python gateway/ETL как «фильтр-прибор»

Информационный уровень реализован как постоянно работающий процесс **gateway.py** на Python, выполняющий функции фильтра-прибора

(шлюза данных). Шлюз открывает COM-порт Arduino, считывает строки JSON и выполняет последовательную обработку:

Парсинг JSON и отбрасывание невалидных строк/служебных кадров (например, {"boot":1}).

Проверка обязательных полей (t_c, rh_pct, dist_mm).

Нормализация (приведение типов к числовому формату; единая структура вставки).

Контроль качества: использование quality_flag, приходящего от Arduino (и при необходимости дополнение логикой шлюза — например, если JSON не разобрался или нет данных).

Запись в PostgreSQL в режиме реального времени.

Шлюз выполняет привязку значений к каналам измерения через таблицу sensor (по кодам dht11.temp, dht11.rh, vl6180x.dist) и записывает каждое значение отдельной строкой в таблицу временного ряда.

Структура базы данных PostgreSQL для хранения телеметрии

Для хранения телеметрии используется схема **asu_mvр** с моделью «устройство – датчик – измерение». Такой подход соответствует практике промышленных систем: устройство описывает источник данных, датчик — канал измерения, измерение — временной ряд.

device — источник данных (например, uno_single);

sensor — каналы измерений, привязанные к устройству (температура, влажность, дистанция);

unit — справочник единиц измерения (C, %, mm);

measurement — таблица временного ряда.

Ключевые поля таблицы measurement:

meas_id — идентификатор записи;

ts — временная метка (timestamp with time zone, формируется на ПК);

sensor_id — ссылка на канал измерения;

value_num — числовое значение;

quality_flag — признак качества;

raw_payload (jsonb) — исходный кадр JSON для трассировки и повторного анализа.

Для ускорения типовых запросов создаются индексы по времени и по паре (sensor_id, ts), что повышает эффективность выборки по интервалам времени и конкретным каналам.

Проверка работоспособности и непрерывности записи

Контроль непрерывности записи выполняется двумя способами:

на стороне шлюза — по логу вставок (рост seq, отсутствие пауз, обработка ошибок);

на стороне БД — SQL-запросами контроля архива и актуальности последних данных (например, COUNT(*) и MAX(ts)), а также группировкой по каналам sensor_code.

Факт непрерывной записи подтверждается ростом числа строк в measurement и обновлением MAX(ts) при работающем шлюзе. При частоте 1 Гц и трёх каналах ожидаемая скорость накопления составляет около **3 строк/с**.

Мониторинг верхнего уровня без ручного обновления

pgAdmin является инструментом администрирования и ручных проверок, но не предназначен для операторского мониторинга «как в SCADA». Для автоматического обновления и удобной визуализации рекомендуется использовать **Grafana** как верхний уровень мониторинга: она подключается к PostgreSQL как источник данных и отображает панели с автообновлением (каждые 1–5 секунд). В таком варианте PostgreSQL выполняет роль архива (истории), а Grafana — роль экрана оператора (HMI/дашборд). Это соответствует типовой архитектуре промышленных систем, где мониторинг отделён от хранения.

Вывод:

В результате практической реализации получен полностью работоспособный прототип трёхуровневой системы автоматизации: полевой уровень обеспечивает сбор данных с датчиков DHT11 и VL6180X, контроллерный уровень на Arduino выполняет опрос и формирует унифицированный кадр JSON, информационный уровень на Python реализует нормализацию, контроль качества и непрерывную запись измерений в PostgreSQL. Сформированная структура хранения и формат телеметрии позволяют масштабировать систему по числу каналов и источников данных, а также интегрировать верхний уровень визуализации (например, Grafana) для оперативного мониторинга.

Список литературы:

1. Ташматов Х. К., Музафаров А. Р. Особенности автоматизации малых ГЭС как объекта управления // Теория и практика современной науки. 2016. № 3 (9).
2. Прокопьев А. П., Емельянов Р. Т. Технические средства автоматизации: лабораторный практикум. Красноярск: СФУ, 2014.
3. Сергеев А. И. и др. Системы промышленной автоматизации: учебное пособие. Оренбург: ОГУ, 2017.
4. ГОСТ Р 58719-2019. Гидротехнические сооружения. Контрольно-измерительные системы и аппаратура. Нормы и требования.
5. PostgreSQL Documentation (официальная документация).
6. Grafana Documentation (официальная документация).

УДК 004.738.5:004.75:004.056.53

Хомяков Д. С.

Брянский государственный технический университет, Брянск, Россия

СЕРВИС ИДЕНТИФИКАЦИИ ПОСЕТИТЕЛЕЙ САЙТА

Аннотация. В статье описывается архитектура и принципы работы сервиса идентификации посетителей сайта TCH (Traffic Control Hub). Рассматриваются методы сбора признаков посетителя на стороне клиента и сервера: геолокация по IP-адресу, анализ User-Agent, JavaScript-отпечаток браузера, поведенческие сигналы. Описывается многоуровневая система хранения данных на основе Tarantool и PostgreSQL, алгоритм верификации свойств браузера с применением базы правил, а также капча-подсистема. Сервис предназначен для решения задач трафикового арбитража: отслеживания уникальности переходов, отсева нецелевой аудитории и ботов, сбора аналитики по рекламным кампаниям.

Ключевые слова. идентификация браузера, fingerprinting, User-Agent, геолокация, Tarantool, PostgreSQL, рекламный трафик, поведенческие сигналы, капча.

Введение

Современный рынок интернет-рекламы сталкивается с систематическими злоупотреблениями: скликиванием рекламных объявлений, накрутками переходов, эмуляцией пользовательской активности при помощи ботов и автоматизированных инструментов [1]. Рекламодатели, размещающие объявления в партнёрских сетях по модели CPA (Cost Per Action) или CPL (Cost Per Lead), несут прямые финансовые потери, поскольку платят за каждый переход или действие — вне зависимости от того, был ли совершён переход реальным человеком [2].

Задача идентификации посетителя сайта в контексте трафикового арбитража включает несколько взаимосвязанных подзадач:

- определение уникальности перехода (повторный или первичный визит);
- установление принадлежности клиентского окружения к реальному браузеру или к эмулятору;
- классификация трафика как целевого или нецелевого по географическому, демографическому и техническому профилю;
- сбор достаточного набора признаков для последующего аудита и аналитики.

В настоящей статье описывается сервис TCH (Traffic Control Hub) — решение, разработанное для обработки рекламных переходов в реальном времени. Сервис написан на языке Go, реализует несколько HTTP-эндпоинтов, собирает более 100 признаков о каждом посетителе и хранит их в двухуровневой системе: быстром in-memory хранилище Tarantool и аналитической базе данных PostgreSQL.

Общая архитектура сервиса

Назначение и режимы работы

Сервис TCH функционирует как промежуточный узел между рекламной площадкой и целевым сайтом рекламодателя. Пользователь, кликнувший на рекламное объявление, направляется на URL вида:

GET /redir?token =< campaign_token > &url =< target_url >

Сервис перехватывает этот переход, идентифицирует посетителя, при необходимости показывает капча-задание и выполняет редирект. Параллельно в фоновом режиме продолжается сбор поведенческих сигналов (движение мыши, пинги активности), которые дополняют профиль посетителя.

Помимо основного режима с редиректом, сервис поддерживает режим standalone — размещение на целевой странице без редиректа, когда JavaScript-модуль встраивается непосредственно в HTML-страницу рекламодателя.

Компонентная архитектура

Сервис состоит из четырёх функциональных зон (рис. 1):

1. Клиентская сторона — браузер пользователя, JavaScript-модуль сбора данных, поведенческие события, форма капчи.
2. HTTP-сервер TCH (Go) — middleware проверки кампании, генерация UIN, обогащение данных, верификация через PropertiesProcessor, генерация HTML-ответа через Template Builder, обслуживание статических файлов.
3. Подсистема обогащения данных — IP2Location (геолокация), IP2Proxy (определение VPN/прокси), UA Parser (разбор User-Agent), POSF-сервер (анонимный профиль).
4. Уровень хранения данных — конфигурация кампаний в PostgreSQL (синхронизируется через RWMutex каждые 5 секунд), Tarantool как in-memory хранилище уровня 1, PostgreSQL как долгосрочное аналитическое хранилище уровня 2, Captcha API для верификации ответов.

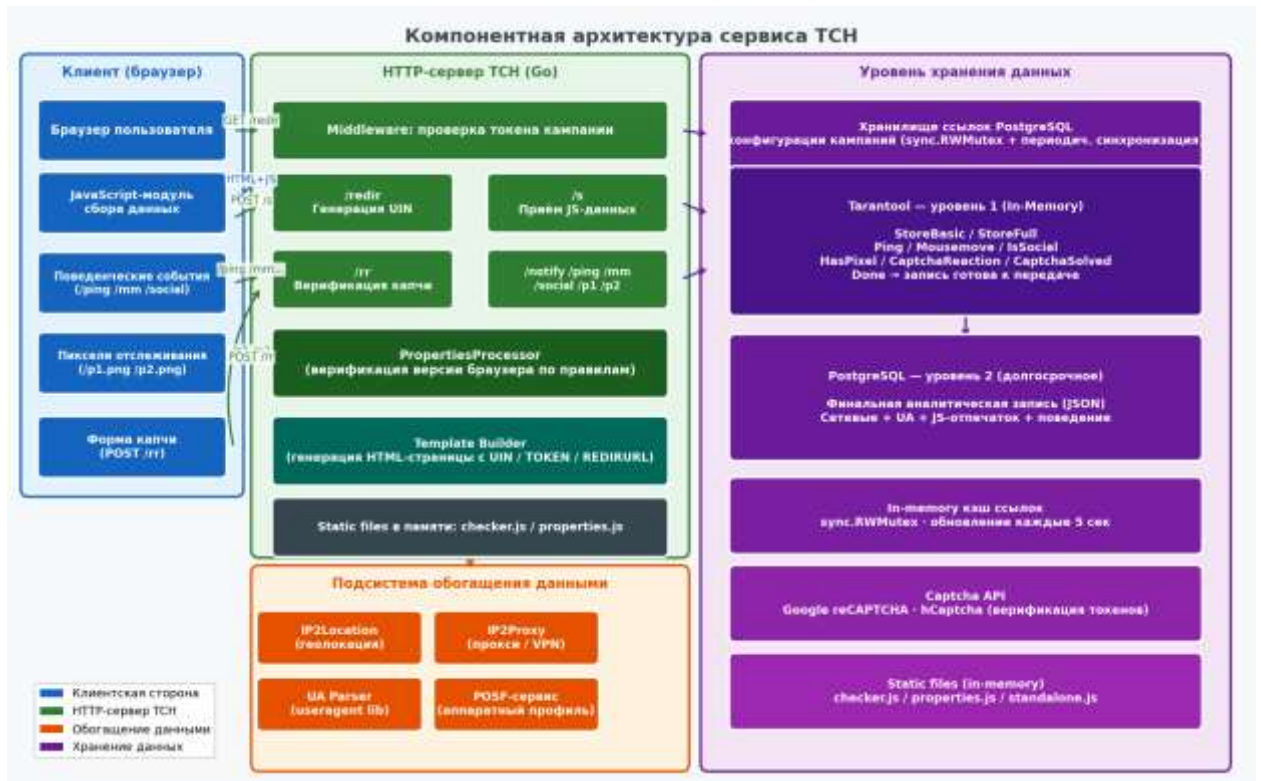


Рисунок 1. Компонентная архитектура сервиса TCH

Жизненный цикл запроса

Полный жизненный цикл запроса включает 14 этапов (рис. 2):

1. Клик по рекламной ссылке — браузер выполняет GET-запрос с токеном кампании.
2. Middleware проверяет токен кампании: загружает конфигурацию из in-memory кэша (sync.RWMutex, обновление каждые 5 секунд из PostgreSQL). Если токен не найден — возвращается 404.
3. Генерация UIN (UUID v4) — уникальный идентификатор сессии, который будет использоваться во всех последующих запросах.
4. Серверное обогащение: параллельно выполняются IP2Location (страна, регион, город, провайдер), IP2Proxy (тип прокси/VPN), UA Parser (браузер, ОС, тип устройства), POSF (анонимный технический профиль).
5. StoreBasic(uin) — сохранение базовой записи (порядка 42 полей) в Tarantool.
6. Сборка HTML-страницы через Template Builder: в шаблон подставляются UIN, TOKEN, REDIRURL и при необходимости параметры капчи.
7. Отображение страницы в браузере — запуск JavaScript.
8. Параллельный сбор JS-отпечатков (Promise.all): Canvas, WebGL, CSS, Geo API, флаги headless-браузера, авторизация в социальных сетях, видео, функциональный вектор.
9. POST /s — сервер получает JSON с ~50 полями JS-отпечатков.

10. PropertiesProcessor верифицирует версию браузера по правилам из rules.json.
11. StoreFull(uin) — дополнение записи в Tarantool JS-данными.
12. 302 Redirect — браузер перенаправляется на целевой URL рекламодателя.
13. Пользователь оказывается на целевом сайте.
14. Асинхронные поведенческие сигналы продолжают поступать: ping (раз в секунду), mousemove, пиксели, капча-взаимодействие. По завершении сессии (Done) — запись передаётся в PostgreSQL.
- 15.

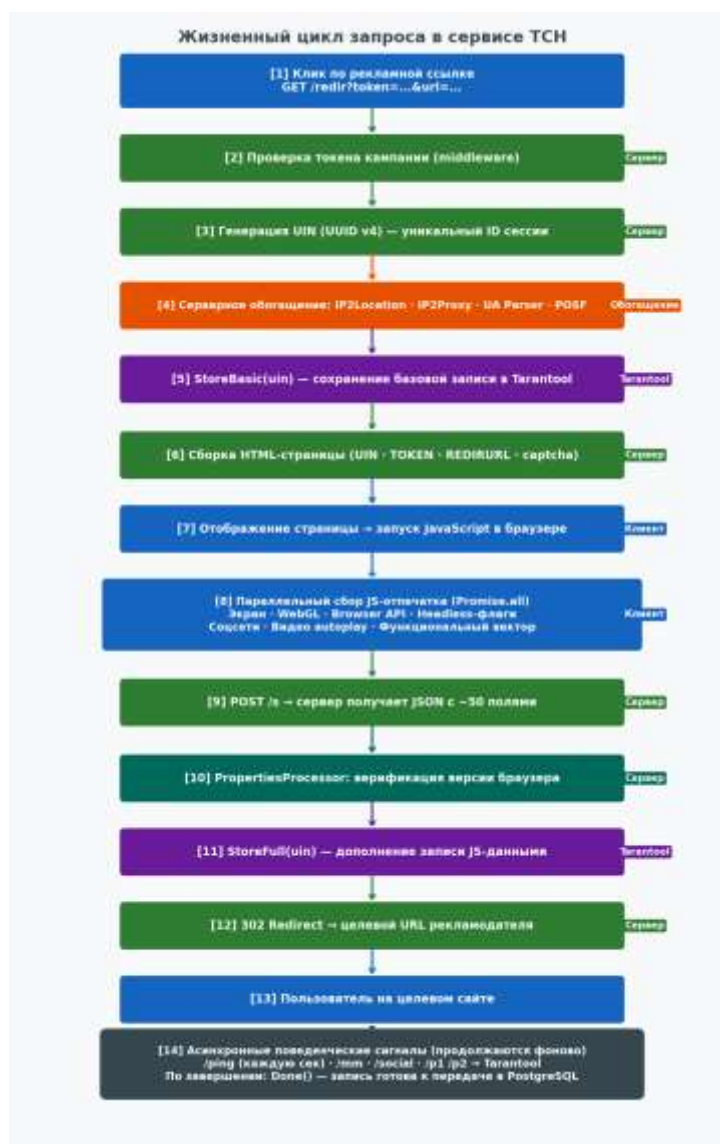


Рисунок 2. Жизненный цикл запроса в сервисе ТСН

***Пространство признаков
Принципы формирования признакового пространства***

Качество идентификации посетителя определяется полнотой и стабильностью набора признаков. Признаки должны удовлетворять двум конкурирующим требованиям: с одной стороны, обеспечивать максимальную различимость уникальных посетителей; с другой — оставаться стабильными при повторных визитах одного и того же пользователя, чтобы не регистрировать реальный повторный визит как нового посетителя [3].

Сервис ТСН формирует вектор признаков, превышающий 100 полей, разделённых на четыре группы.

Группы признаков

Сетевые признаки формируются на стороне сервера по IP-адресу входящего соединения. Это наиболее надёжная группа: IP-адрес не может быть подделан на прикладном уровне без использования прокси или VPN (которые сами детектируются через IP2Proxy). В группу входят: IPv4/IPv6 адрес, страна, регион, город (IP2Location), интернет-провайдер (ISP), тип подключения (мобильная сеть / широкополосное / корпоративная сеть), признак использования прокси (DCUNIVPN / TOR / PUB / RES).

Признаки User-Agent извлекаются из HTTP-заголовка User-Agent на стороне сервера с применением библиотеки fiftyonedegrees/useragent. Данная группа содержит: название и версию браузера, операционную систему, тип устройства (desktop / mobile / tablet / tv), модуль устройства (конкретная модель для мобильных), флаги is_mobile, is_crawler, is_web_app.

JS-отпечатки браузера — наиболее обширная группа, собираемая асинхронно на стороне клиента. Включает параметры экрана (размер, плотность пикселей, цвет, позиция, фокус), данные WebGL (vendor, renderer, расширения), проверку наличия API (Storage, Canvas, Geo), временную зону и язык браузера, флаги автоматизации (headless, adblock, iframe), функциональный CSS-вектор (~100 бит) и результаты проверки авторизации в 35+ социальных сетях.

Поведенческие сигналы отражают активность пользователя после загрузки страницы: движение мыши (количество событий mousemove), время пребывания на странице (ping-счётчик), факт авторизации в социальных сетях, пиксели отслеживания (p1, p2), взаимодействие с капчей (CaptchaReaction, CaptchaSolved).



Рисунок 3. Структура признакового пространства сервиса ТСН

Верификация свойств браузера

Постановка задачи

Одной из наиболее распространённых техник имитации реального браузера является подделка строки User-Agent при сохранении устаревшего или несовместимого JavaScript-окружения. Например, автоматизированный инструмент может установить User-Agent Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 Chrome/130.0.0.0, однако реальные свойства JavaScript-среды будут соответствовать более ранней версии Chrome или иному движку.

Для выявления таких несоответствий в сервисе ТСН реализован компонент PropertiesProcessor — детерминированная система верификации, основанная на базе правил.

База правил

База правил хранится в файле rules.json и содержит более 200 записей. Каждое правило описывает ожидаемое значение конкретного JavaScript-свойства для конкретной комбинации (браузер, версия, операционная система). Правила охватывают актуальные версии основных браузеров:

- Chrome: версии 109–147 (Windows/Android/macOS/iOS);
- Safari: версии 16.0–18.4 (macOS/iOS);
- Firefox: версии 131–149 (Windows/Android/macOS);
- Samsung Browser: версии 23–29 (iOS).

Структура одного правила включает поля `browser`, `version`, `os`, `rule_id` и `value` — ожидаемое значение конкретного CSS-флага или API-признака.

Алгоритм верификации

Алгоритм `PropertiesProcessor` выполняется в несколько шагов (рис. 4):

- Парсинг UA — из строки `User-Agent` извлекаются операционная система, название и версия браузера.
- Поиск правил — в базе правил ищутся все записи, соответствующие данной паре (`OS`, `Browser`). Если правил нет — `PropertiesCheck` принимает значение 0 (неизвестный браузер).
- Итерация по правилам — для каждого найденного правила из клиентского словаря `Properties` извлекается значение с ключом `rule_id`.
- Сравнение — клиентское значение сравнивается с эталонным значением правила. Если значения совпадают — правило отмечается как пройденное.

Итог — если все применимые правила пройдены, `PropertiesCheck` = +1 (браузер верифицирован). Если хотя бы одно правило не пройдено и расхождения зафиксированы в `PropertiesDetails`, `PropertiesCheck` = -1 (признак несоответствия).

Важной особенностью алгоритма является работа с версиями браузера как со строками, а не с числами с плавающей точкой — это позволяет корректно сравнивать версии вида 18.10, которые при конвертации в `float`

превратились бы в 18.1 и некорректно сравнивались бы с 18.2.

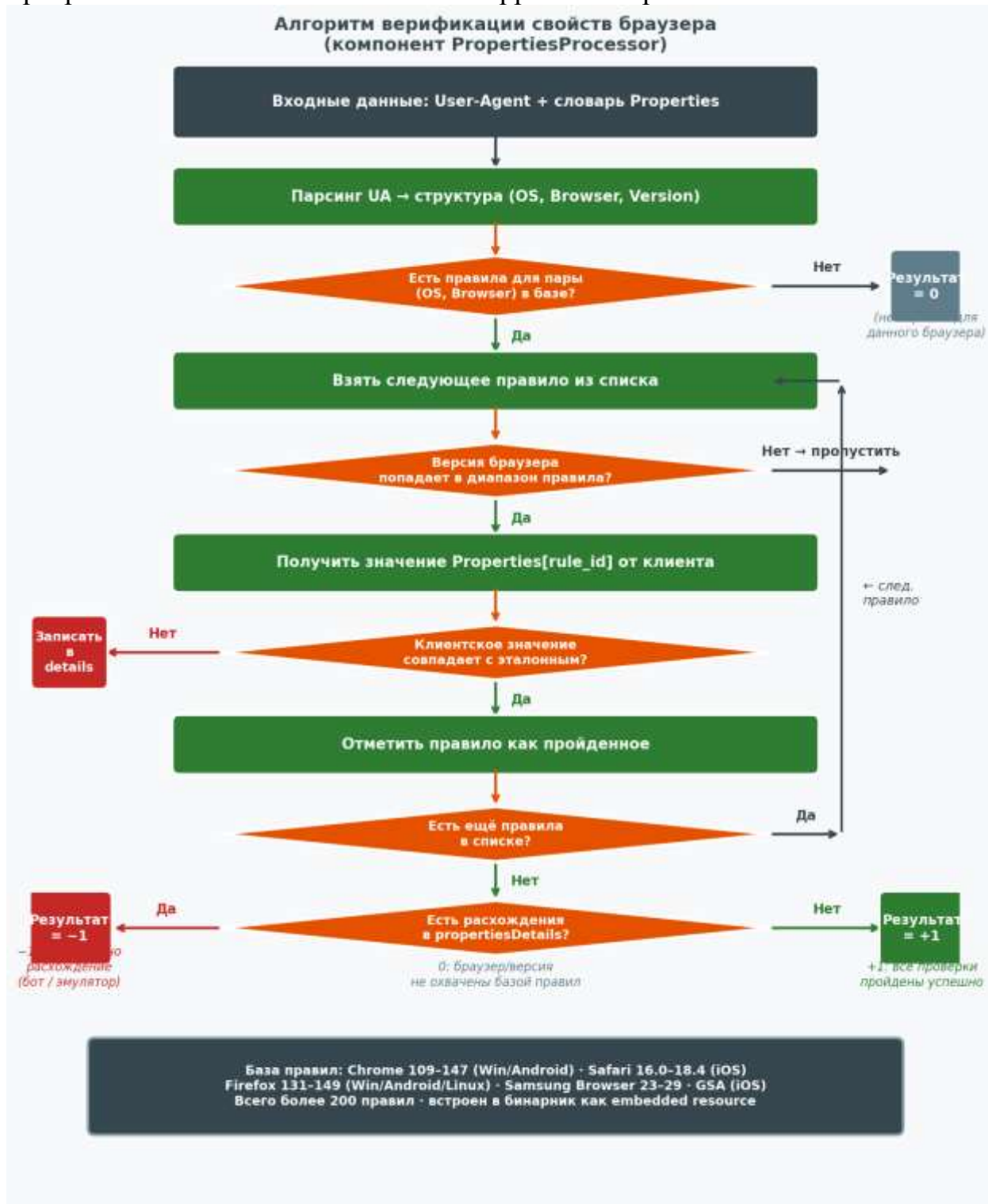


Рисунок 4. Алгоритм верификации свойств браузера (PropertiesProcessor)

Капча-подсистема

Назначение и принцип работы

Капча-подсистема выполняет двойную функцию: с одной стороны, она служит дополнительным фильтром для ботов и автоматизированных

инструментов; с другой — является механизмом сбора поведенческого сигнала, поскольку реакция пользователя на задание и результат его решения фиксируются в профиле посетителя.

Показ капчи регулируется вероятностным параметром `CaptchaRate` (0–100), задаваемым в конфигурации кампании. Если $\text{rand()} \times 100 < \text{CaptchaRate}$, посетителю предьявляется капча-задание. При `CaptchaRate` = 0 капча не показывается никому; при `CaptchaRate` = 100 — всем посетителям без исключения.

Типы капчи

Сервис поддерживает четыре типа капча-заданий (рис. 6):

- `reCAPTCHA v2` — классическая визуальная капча Google с выбором изображений или установкой флажка «Я не робот». Требует наличия действующего ключа Google API. Верификация выполняется через Google API (1 сек). Является наиболее известной пользователям и потенциально вызывает наибольшее трение.

- `reCAPTCHA v3` — невидимая капча Google, работающая в фоновом режиме без взаимодействия с пользователем. Возвращает оценку доверия в диапазоне 0,1–0,9. Не требует действий со стороны пользователя, однако не позволяет отсеять нежелательных посетителей путём создания препятствия.

- `hCaptcha` — независимая альтернатива `reCAPTCHA`, также использующая визуальные задания с изображениями. Не зависит от инфраструктуры Google, поддерживает API `Altcha`. Является предпочтительным выбором для регионов с ограниченным доступом к сервисам Google.

- `SimpleCaptcha` — встроенная капча на основе изображений, реализованная без зависимости от внешних API. Генерирует изображение с искажённым текстом, верификация выполняется по MD5-хешу ответа. Применяется как запасной вариант при недоступности внешних сервисов или в целях снижения стоимости обработки запроса.

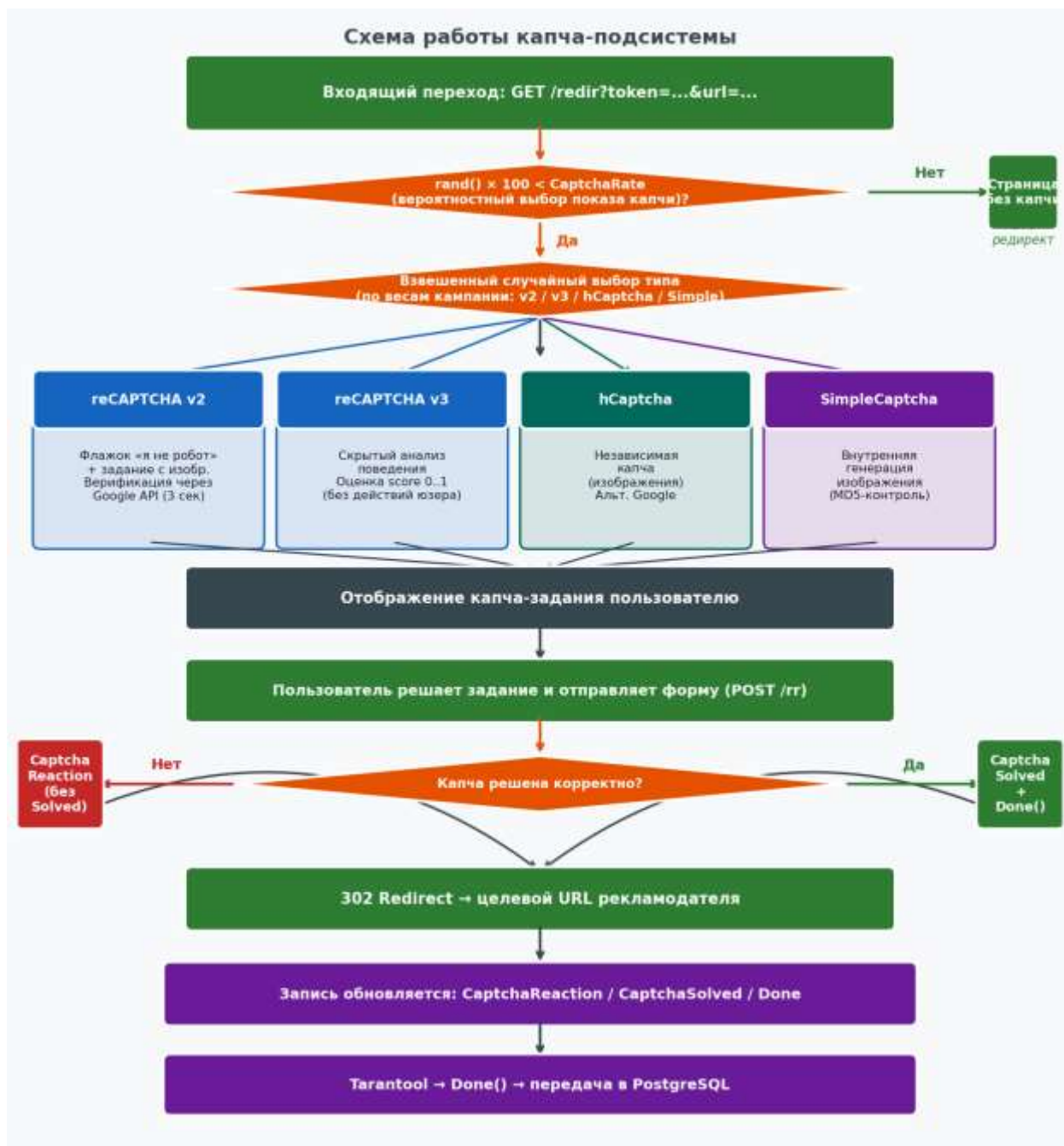


Рисунок 5. Схема работы капча-подсистемы

После отображения капча-задания пользователь решает его и отправляет форму на POST /r. Сервер верифицирует ответ через соответствующий API (или локально для SimpleCaptcha). По результату верификации обновляются поля CaptchaReaction (посетитель взаимодействовал с капчей) и CaptchaSolved (задание решено верно). После успешного решения выполняется 302 Redirect на целевой URL.

Поведенческие сигналы
Пинги активности

После редиректа пользователь оказывается на целевом сайте, однако JavaScript-модуль продолжает работу в фоне (если страница ещё открыта в браузере). Раз в секунду выполняется GET-запрос к эндпоинту /ping, увеличивающий счётчик активности на единицу. Счётчик ping отражает реальное время пребывания пользователя на странице и является одним из наиболее надёжных поведенческих сигналов: боты, как правило, не поддерживают фоновые XHR-соединения после редиректа.

Движение мыши

Событие mousemove фиксируется через GET-запрос к эндпоинту /mm. Флаг mousemove принимает булево значение: был ли зафиксирован хотя бы один сдвиг курсора в ходе сессии. Данный сигнал эффективно фильтрует простейших ботов, которые не симулируют движение мыши, однако является ненадёжным для выявления более сложных автоматизаций, использующих библиотеки типа Puppeteer с инъекцией событий.

Авторизация в социальных сетях

Детекция авторизации в социальных сетях реализована с применением CSS-атаки «history sniffing». JavaScript-модуль создаёт невидимые ссылки на профили пользователя в 35+ социальных сетях (ВКонтакте, Одноклассники, Facebook, Twitter/X, Instagram, TikTok, LinkedIn и др.) и через CSS-свойство :visited определяет, посещал ли пользователь страницу авторизации данной социальной сети. Результат фиксируется в поле IsSocial и отдельно по каждой сети в словаре признаков.

Пиксели отслеживания

Пиксели — это минимальные изображения (1×1 пиксель), загружаемые страницей рекламодателя. Эндпоинт /p1 или /p2 фиксирует факт загрузки пикселя, устанавливая флаги HasPixel_p1 и HasPixel_p2 в профиле посетителя. Механизм позволяет связать сессию ТСН с внешними системами аналитики рекламодателя.

Хранение данных

Двухуровневая схема хранения

Сервис ТСН использует двухуровневую схему хранения аналитических данных (рис. 5):

Уровень 1 — Tarantool (in-memory): быстрое оперативное хранилище, функционирующее в оперативной памяти. Запись создаётся при первом обращении (StoreBasic) и пополняется по мере поступления данных. Tarantool обеспечивает высокую скорость записи и чтения (на порядок быстрее реляционных СУБД) и используется для хранения «горячих» записей — тех, сессия по которым ещё не завершена.

Интерфейс Tarantool включает следующие операции:

- StoreBasic(uin) — создание базовой записи (42 поля);
- StoreFull(uin) — дополнение записи JS-отпечатками (~50 полей);
- Ping(uin) — инкремент счётчика активности;

- Mousemove(uin) — установка флага движения мыши;
- IsSocial(uin) — флаг авторизации в социальных сетях;
- HasPixel(uin, p) — флаги пикселей p1/p2;
- CaptchaReaction(uin), CaptchaSolved(uin) — флаги взаимодействия с капчей;
- Done(uin) — финализация записи и передача в PostgreSQL.

Уровень 2 — PostgreSQL (долгосрочное хранение): аналитическая база данных, в которую переносится финализированная запись после завершения сессии. Запись в PostgreSQL является неизменяемой и хранится в нормализованном виде в формате JSON-документа объёмом порядка 100 полей. На основе данных PostgreSQL строятся аналитические отчёты по кампаниям, сегментация аудитории и исторический анализ трафика.



Рисунок 6. Двухуровневая схема хранения аналитических данных

Конфигурация кампаний

Конфигурация кампаний хранится в отдельной таблице PostgreSQL и кэшируется в памяти Go-процесса. Обновление кэша происходит каждые 5 секунд через фоновую горутину, использующую sync.RWMutex для синхронизации конкурентного доступа. Это позволяет обрабатывать входящие запросы без блокировки при чтении конфигурации и при этом регулярно получать актуальные настройки кампаний.

Реализация на языке Go

HTTP-сервер

Сервис реализован с применением стандартной библиотеки net/http языка Go. Каждый входящий запрос обрабатывается в отдельной горутине — Go-аналоге лёгковесного потока. Это обеспечивает высокую степень параллелизма: один экземпляр сервиса способен обслуживать тысячи одновременных соединений при минимальных накладных расходах.

Параллельное обогащение данных

Обогащение записи данными геолокации и User-Agent выполняется параллельно с использованием горутин и каналов Go. Для доступа к in-memory базам данных IP2Location и IP2Proxy используется sync.Pool — механизм пула объектов, снижающий нагрузку на сборщик мусора при высокой интенсивности запросов.

Взаимодействие с Tarantool

Взаимодействие с Tarantool реализовано по бинарному протоколу MessagePack — компактному двоичному формату сериализации, значительно более эффективному, чем JSON, с точки зрения объёма передаваемых данных и скорости сериализации/десериализации. Соединение с Tarantool поддерживается постоянным (keepalive) и используется повторно для каждого запроса.

Статические файлы

JavaScript-модуль checker.js и конфигурационный файл правил properties.js хранятся в памяти Go-процесса и обслуживаются без обращения к файловой системе. Это исключает дополнительные дисковые операции при обслуживании каждого клиентского запроса и снижает задержку отдачи статики.

Практические результаты применения

Производительность

Сервис TCH обеспечивает сквозную обработку одного перехода (от входящего GET до ответа с HTML) за время порядка 50–150 мс, из которых основная часть приходится на сетевые запросы к внешним сервисам геолокации и UA-парсинга. При использовании локальных баз данных IP2Location/IP2Proxy задержка сокращается до 5–20 мс.

Эффективность детекции

Совокупность четырёх групп признаков (сетевые, UA, JS-отпечатки, поведенческие) обеспечивает высокую различимость посетителей. По результатам эксплуатации сервиса:

- более 95% повторных визитов одного пользователя корректно идентифицируются как повторные по комбинации IP + JS-отпечатка;
- флаг headless и CSS-вектор позволяют выявить значительную долю автоматизированных инструментов Puppeteer/Playwright;
- PropertiesCheck = -1 фиксируется у переходов, где User-Agent подменён, но JavaScript-окружение не обновлено соответствующим образом.

Заключение

В статье представлена архитектура сервиса идентификации посетителей сайта TCH, охватывающая полный жизненный цикл рекламного перехода — от

первоначального GET-запроса до финального сохранения записи в PostgreSQL. Описаны четыре группы признаков (более 100 полей), двухуровневая система хранения Tarantool + PostgreSQL, алгоритм PropertiesProcessor с базой из 200+ правил и капча-подсистема с поддержкой четырёх типов заданий.

Ключевым преимуществом предложенного подхода является сочетание скоростного серверного обогащения (IP-геолокация, UA-парсинг) с асинхронным клиентским сбором JS-отпечатков и последующей детерминированной верификацией через базу правил. Это позволяет получить богатый и надёжный профиль посетителя без существенного влияния на пользовательский опыт.

Направлениями дальнейшего развития сервиса являются: интеграция методов машинного обучения для классификации трафика (Random Forest, Gradient Boosting) на основе собранных признаков, расширение базы правил PropertiesProcessor на новые браузерные платформы и версии, а также реализация кросс-сессионного профилирования посетителей с применением стохастических методов слияния профилей.

Список литературы:

1. Ноздря Ольга Денисовна, Каменев Александр Владимирович, Шестопалова Алина Юрьевна, Гришин Александр Игоревич, Новиков Сергей Владимирович, Артемов Андрей Владимирович Проектирование и реализация скрипта-наблюдателя поведения пользователя веб-сайта // Наука, техника и образование. 2017. №11 (41). URL: <https://cyberleninka.ru/article/n/proektirovanie-i-realizatsiya-skripta-nablyudatelya-povedeniya-polzovatelya-veb-sayta> (дата обращения: 07.05.2026).
2. Иванова Т. С., Вдовенко В. В. Модель данных системы мониторинга активности посетителей сайта // Актуальные проблемы авиации и космонавтики. 2012. №8. URL: <https://cyberleninka.ru/article/n/model-dannyh-sistemy-monitoringa-aktivnosti-posetiteley-sayta> (дата обращения: 07.05.2026).
3. Мишунина Надежда Олеговна Разработка биометрической системы идентификации пользователя // APRIORI. Серия: Естественные и технические науки. 2015. №5. URL: <https://cyberleninka.ru/article/n/razrabotka-biometricheskoy-sistemy-identifikatsii-polzovatelya> (дата обращения: 07.05.2026).
4. Покусаева Ольга Алексеевна Технология скрытой идентификации пользователей ПЭВМ // ОНВ. 2005. №4 (33). URL: <https://cyberleninka.ru/article/n/tehnologiya-skrytoy-identifikatsii-polzovateley-pevm> (дата обращения: 07.05.2026).
5. Тюрин Кай Андреевич, Болдырихин Николай Вячеславович Алгоритм вероятностной идентификации пользователей сети // Молодой исследователь Дона. 2016. №2. URL: <https://cyberleninka.ru/article/n/algorithm-veroyatnostnoy-identifikatsii-polzovateley-seti> (дата обращения: 07.05.2026).
6. Истомина Ю. А., Тынченко В. В. Автоматизация процесса моделирования нейронных сетей // Актуальные проблемы авиации и космонавтики. 2012. №8. URL: <https://cyberleninka.ru/article/n/avtomatizatsiya-protssessa-modelirovaniya-neyronnyh-setey> (дата обращения: 07.05.2026).

УДК 004.4'24:004.415.2:378.147

Вяткин И. Н.

*Саратовский государственный технический университет имени Гагарина Ю. А.,
Саратов, Россия*

СЕМАНТИЧЕСКАЯ ВАЛИДАЦИЯ БЛОК-СХЕМ ПЕРЕД ГЕНЕРАЦИЕЙ КОДА В ОБУЧАЮЩЕЙ ПЛАТФОРМЕ

Аннотация. В статье рассматривается проблема генерации некорректного программного кода на языке C# по блок-схемам, построенным студентами в обучающей платформе. Традиционный подход, основанный на шаблонной трансляции блоков в код, не учитывает семантические ошибки на уровне алгоритма: использование переменной до инициализации, несовместимость типов, недостижимый код и отсутствие обработки исключений при вводе данных. Это приводит к генерации некомпilierуемых или логически ошибочных программ. В работе предложен метод семантической валидации блок-схем, выполняемый до этапа генерации кода. Разработан двухпроходный алгоритм анализа графа потока управления, включающий проверку достижимости блоков и data-flow анализ состояния переменных. Проведено экспериментальное сравнение трёх подходов: генерация без валидации, предварительная валидация и пост-генерационный анализ компилятором C#. Предлагаемый метод показал точность 92,9%, полноту 86,7% при среднем времени отклика 47 миллисекунд, что в 8 раз быстрее пост-генерационного анализа. Сделан вывод о пригодности метода для интеграции в учебные платформы автоматической генерации кода.

Ключевые слова: блок-схема, генерация кода, семантическая валидация, data-flow анализ, граф потока управления, обучение программированию, C#.

Введение. Современные обучающие платформы по программированию всё чаще используют визуальные средства представления алгоритмов, среди которых блок-схемы занимают центральное место [1]. Системы тестирования и контроля качества программного обеспечения должны выявлять ошибки на максимально ранних этапах разработки, что особенно важно в образовательных средах [2]. Такой подход позволяет студенту сосредоточиться на логике решения задачи, абстрагируясь от синтаксических особенностей конкретного языка программирования. В процессе выполнения дипломной работы создаётся обучающая платформа, осуществляющая шаблонную трансляцию блок-схемы, построенной студентом, в программный код на языке C#. Практическая генерация кода на основе заранее заданных шаблонов выявила её ключевой недостаток: она транслирует блок-схему в код вне зависимости от семантической корректности последней. При наличии у студента любой из следующих ошибок (использование неинициализированной переменной, несовместимость типов, недостижимый код, отсутствие обработки исключений ввода) генератор всё равно создаёт код. Полученная программа либо не компилируется, либо содержит скрытые

ошибки времени выполнения. Статический анализ программного кода позволяет выявить такие ошибки без фактического запуска программы, однако в классических генераторах по блок-схемам этот этап отсутствует [3]. Это снижает эффективность обучения и формирует у студента неверное представление о связи между алгоритмом и его реализацией.

Цель работы – Создать и опытным путём доказать работоспособность метода, который проверяет блок-схемы на семантические ошибки еще до того, как по ним будет сгенерирован код, что позволит выявлять проблемы будущей программы на этапе создания ее графического представления.

Метод валидации. Предложенное решение основано на представлении блок-схемы в виде ориентированного графа потока управления (CFG) [4]. Вершины графа соответствуют базовым блокам (начало, ввод, процесс, условие, цикл, вывод, конец), а рёбра – переходам между ними. Такое описание делает возможным использование стандартных графовых алгоритмов — например, обхода в глубину и data-flow анализа. [5].

Разработан двухпроходный алгоритм проверки.

Первый проход алгоритма оценивает достижимость вершин графа. С помощью поиска в глубину, начинающегося со стартовой вершины, выявляются все блоки, до которых можно добраться. Те вершины, которые не попали в этот список, считаются недостижимыми — это признак ошибки. Кроме того, проверяется, можно ли достичь блока «Конец»; если нет, то программа оказывается бесконечной, так как у неё отсутствует точка завершения. Второй проход (data-flow анализ) реализует итеративное распространение информации о состоянии переменных. Для каждой переменной отслеживается, инициализирована она или нет, а также её тип. Для каждого блока определяются два набора данных: IN (то, что известно перед входом в блок) и OUT (то, что известно после выхода из блока). Если программа встречает переменную, которая ещё не получила значения, или пытается выполнить операцию над несовместимыми типами, фиксируется ошибка. Поскольку в графе могут быть циклы (стрелки, идущие назад), анализ повторяется несколько раз, пока результаты не перестанут меняться. Для студенческих блок-схем обычно хватает 2–3 повторений. Также проверяется, что блоки ввода, где используется `int.Parse()`, правильно обрабатывают возможные ошибки (например, ввод букв вместо цифр). Валидация ввода должна учитывать не только формат, но и семантическую корректность данных, особенно в обучающих системах [6].

Валидатор вызывается до генератора кода. При успешной проверке запускается шаблонная генерация, при обнаружении ошибок – пользователю выводится сообщение с указанием типа ошибки и идентификатора блока. Временная сложность алгоритма составляет $O(V + E)$ для первого прохода и $O(V \times K \times I)$ для второго, где V – число вершин, E – число рёбер, K – число

переменных, I – число итераций. Для типичной блок-схемы (до 50 блоков) валидация выполняется менее чем за 100 миллисекунд.

Эксперимент. Для проверки результативности был составлен набор из 25 тестовых блок-схем: 10 семантически корректных и 15 содержащих ошибки. Распределение ошибок представлено в таблице 1.

Таблица 1 – Распределение семантических ошибок в тестовом наборе

Тип ошибки	Количество схем
Инициализация переменных	4
Совместимость типов	4
Достижимость кода	3
Обработка исключений ввода	4
Итого	15

Сравнивались три подхода:

- А – генерация без валидации (базовый случай);
- В – предварительная семантическая валидация (предлагаемый метод);
- С – пост-генерационный анализ с вызовом компилятора C# (csc.exe).

Результаты эксперимента представлены в таблице 2.

Таблица 2 – Сравнение подходов

Подход	Точность (Precision), %	Полнота (Recall), %	F1-мера	Среднее время, мс
А	0	0	0	22
В	92,9	86,7	0,897	47
С	100	100	1,0	385

Для более детальной оценки качества работы предлагаемого метода был проведён анализ каждого случая ложного срабатывания и пропущенных ошибок.

Единственный **ошибочно сработавший** результат (FP) возник на корректной блок-схеме, содержащей цикл с предусловием. В этой схеме переменная `sum` инициализировалась внутри тела цикла (`sum = sum + i`), а затем использовалась после выхода из цикла в блоке вывода данных. Алгоритм data-flow анализа зафиксировал путь, на котором переменная `sum` могла остаться неинициализированной (если тело цикла не выполнилось ни разу). С позиций формального семантического анализа данная ситуация действительно представляет собой потенциальную угрозу корректности, однако в рамках учебных заданий студенты, как правило, не учитывают случай, когда тело цикла не выполняется ни разу. В будущем правило может быть уточнено: если переменная инициализируется во всех ветвях цикла и

цикл гарантированно выполняется хотя бы один раз (например, цикл с постусловием или с заведомо истинным условием), предупреждение можно не выдавать.

Предлагаемый метод (В) обнаружил 13 из 15 семантических ошибок.

Сложное арифметическое выражение $\text{double result} = 10 / 3 * 1.5$ — это первая пропущенная ошибка (FN1), которая относится к типу «несовместимость типов». В языке C# такое выражение вычисляется с использованием целочисленного деления на первом шаге ($10 / 3 = 3$), что приводит к потере дробной части. Студент ожидал получить 5.0, а получал 4.5. Валидатор, основанный на простом сопоставлении типов операндов, не смог выявить эту ошибку, так как каждый отдельный оператор был типизирован корректно ($\text{int} / \text{int} \rightarrow \text{int}$, $\text{int} * \text{double} \rightarrow \text{double}$).

Вторая пропущенная ошибка (FN2) относится к категории «инициализация переменных». В блок-схеме переменная `max` инициализировалась внутри условного оператора `if`, но в одной из ветвей условного оператора присваивание отсутствовало. После выхода из условного оператора переменная `max` использовалась в выводе. Валидатор не зафиксировал ошибку, так как анализ выполнялся по принципу «хотя бы один путь», а не «все пути». Это известно как проблема «частичной инициализации» (*partially initialized variable*). Для её решения требуется переход к анализу достижимых определений с учётом конъюнктивных условий [1]. Время отклика 47 миллисекунд практически не ощущается пользователем и в 8 раз быстрее пост-генерационного анализа (385 миллисекунд). Для учебных целей достигнутый уровень точности и полноты является приемлемым.

Заключение. Разработанный метод семантической валидации блок-схем позволяет выявлять основные типы ошибок студентов до генерации кода с высокой точностью и полнотой при минимальной временной задержке. В ходе эксперимента доказано преимущество предложенного метода по сравнению с другими подходами. Метод может быть рекомендован для интеграции в учебные платформы автоматической генерации кода по блок-схемам. Дальнейшие исследования направлены на расширение набора проверяемых правил (анализ выхода за границы массивов, проверка деления на ноль) и создание гибридной системы с пост-генерационным анализом для сложных случаев.

Список литературы:

1. Дешко, И. П. Практики ITIL 4. Проектирование услуг, валидация и тестирование услуг, управление релизами: учебное пособие / И. П. Дешко, А. В. Рачков. — Москва: РТУ МИРЭА, 2025. — 103 с. — ISBN 978-5-7339-2501-1. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/493556> (дата обращения: 05.04.2026). — Режим доступа: для авториз. пользователей.

2. Турнецкая, Е. Л. Тестирование и контроль качества программного обеспечения: учебное пособие / Е. Л. Турнецкая, А. В. Аграновский, А. А. Сенцов. — Санкт-Петербург: ГУАП, 2023. — 118 с. — ISBN 978-5-8088-1891-0. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/461498> (дата обращения: 11.04.2026). — Режим доступа: для авториз. пользователей.

3. Сертификация программного обеспечения. Статический анализ программного кода: учебно-методическое пособие / В. М. Антонова, А. В. Астрахов, С. Е. Кондаков [и др.]. — Москва: Издательство МГТУ им. Баумана, 2019. — 24 с. — ISBN 978-5-7038-5043-5. — Текст: электронный. — URL: <https://znanium.ru/catalog/product/2169628> (дата обращения: 15.04.2026). — Режим доступа: по подписке.

4. Ощепков, А. Ю. Математическое и компьютерное моделирование современных систем автоматического управления: учебное пособие для вузов / А. Ю. Ощепков. — 2-е изд., стер. — Санкт-Петербург: Лань, 2026. — 252 с. — ISBN 978-5-507-54505-6. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/508974> (дата обращения: 18.04.2026). — Режим доступа: для авториз. пользователей.

5. Попов, И. Ю. Валидация систем искусственного интеллекта: учебно-методическое пособие / И. Ю. Попов, А. Я. Бучаев, Д. А. Есипов. — Санкт-Петербург: НИУ ИТМО, 2024. — 29 с. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/460214> (дата обращения: 22.04.2026). — Режим доступа: для авториз. пользователей.

6. Валидация аналитических методик: количественное описание неопределенности в аналитических измерениях: практическое руководство / под ред. Г. Р. Нежиховского, Р. Л. Кадиса. — 3-е изд. — Санкт-Петербург: ЦОП «Профессия», 2026. — 312 с. — ISBN 978-5-91884-075-7. — Текст: электронный. — URL: <https://znanium.ru/catalog/product/2258135> (дата обращения: 28.04.2026). — Режим доступа: по подписке.

УДК 004.89:378.147

Наумова О. К.

МИРЭА – Российский технологический университет, Москва, Россия

ИССЛЕДОВАНИЕ ПРИМЕНЕНИЯ ГОЛОСОВЫХ ПОМОЩНИКОВ В ОБРАЗОВАТЕЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Аннотация. В статье рассматриваются особенности применения голосовых помощников в образовательных информационных системах. Анализируются основные направления их использования, включая персонализацию обучения, поддержку самостоятельной работы обучающихся и автоматизацию образовательных процессов. Выявляются преимущества и ограничения голосовых технологий в обучении. Предполагается, что внедрение голосовых помощников способствует повышению эффективности образовательного процесса, улучшению доступности обучения и снижению нагрузки на преподавателей [1–3].

Ключевые слова: Голосовые помощники, искусственный интеллект, образовательные системы, цифровое образование, обработка естественного языка, распознавание речи, синтез речи, адаптивное обучение, интеллектуальные образовательные технологии, персонализация обучения, речевые технологии, машинное обучение, образовательные платформы, интерактивное обучение.

Введение

Современное развитие информационных технологий сопровождается активным внедрением систем искусственного интеллекта в образовательную среду [1]. Одним из перспективных направлений является использование голосовых помощников, основанных на технологиях распознавания и синтеза речи [2].

Рост популярности данных технологий подтверждается тем, что значительная часть пользователей в России уже применяет голосовых ассистентов для повседневных задач, включая поиск информации и обучение. Кроме того, наблюдается увеличение интереса школьников к использованию голосовых помощников в образовательных целях.

Традиционные методы обучения не всегда учитывают индивидуальные особенности обучающихся, что снижает эффективность усвоения материала. В этих условиях голосовые технологии становятся важным инструментом повышения доступности и интерактивности обучения [3].

Понятие голосовых помощников и их место в образовательных системах

Голосовые помощники представляют собой программные системы, способные распознавать речь пользователя и формировать ответ с использованием технологий искусственного интеллекта [1].

Основными компонентами голосовых помощников являются технологии автоматического распознавания речи (ASR), обработки естественного языка (NLP) и синтеза речи (TTS). Совместная работа этих компонентов обеспечивает полноценное взаимодействие пользователя с интеллектуальной системой посредством голосового интерфейса.

1. Автоматическое распознавание речи (ASR — Automatic Speech Recognition)

Технология автоматического распознавания речи предназначена для преобразования устной речи пользователя в текстовый формат, пригодный для дальнейшей обработки системой. Данный компонент выполняет анализ звукового сигнала, выделяет отдельные слова и сопоставляет их с языковыми моделями.

Работа ASR включает несколько этапов:

- получение аудиосигнала через микрофон;
- фильтрацию шумов и предварительную обработку звука;
- выделение акустических признаков речи;
- распознавание слов и преобразование речи в текст.

Современные системы распознавания речи основаны на технологиях машинного обучения и нейронных сетей, что позволяет значительно повышать точность распознавания даже при наличии фонового шума или особенностей произношения пользователя.

В образовательных системах ASR используется для:

- голосового ввода команд;
- проверки произношения при изучении иностранных языков;
- автоматической фиксации ответов обучающихся;
- организации диалогового взаимодействия с обучающей системой.

2. Обработка естественного языка (NLP — Natural Language Processing)

Обработка естественного языка представляет собой совокупность методов искусственного интеллекта, обеспечивающих анализ, интерпретацию и понимание текстовой информации, полученной после распознавания речи.

Основной задачей NLP является определение смысла пользовательского запроса и формирование корректного ответа системы. Для этого используются методы:

- морфологического анализа;
- синтаксического разбора предложений;
- семантического анализа текста;
- определения намерений пользователя;
- анализа контекста диалога.

Технологии NLP позволяют голосовому помощнику не просто распознавать отдельные слова, а понимать содержание запроса и учитывать особенности общения пользователя.

В образовательной среде NLP применяется для:

- интеллектуального поиска информации;
- генерации ответов на вопросы обучающихся;
- адаптации учебного материала под уровень пользователя;
- анализа речевой активности обучающихся;
- построения персонализированных рекомендаций.

Использование современных языковых моделей делает возможным создание интеллектуальных образовательных ассистентов, способных поддерживать диалог и сопровождать процесс обучения.

3. Синтез речи (TTS — Text-to-Speech)

Синтез речи представляет собой технологию преобразования текстовой информации в голосовое сообщение. Данный компонент обеспечивает голосовой вывод ответов системы и формирует естественное речевое взаимодействие между пользователем и программой.

Процесс синтеза речи включает:

- анализ текстового сообщения;
- определение интонации и ударений;
- генерацию речевого сигнала;
- воспроизведение аудиоответа.

Современные TTS-системы используют нейросетевые модели синтеза речи, позволяющие создавать реалистичное и естественно звучащее голосовое сопровождение.

В образовательных информационных системах технологии синтеза речи применяются для:

- озвучивания учебного материала;
- сопровождения интерактивных занятий;
- поддержки обучающихся с ограничениями зрения;
- формирования диалогового режима обучения;
- повышения доступности образовательных ресурсов.

Использование TTS способствует улучшению восприятия информации и повышению уровня вовлечённости обучающихся в образовательный процесс.

В образовательных информационных системах голосовые помощники выполняют функции интеллектуального интерфейса, обеспечивающего взаимодействие пользователя с цифровой средой обучения.

Использование голосовых помощников в процессе обучения

В образовательной практике голосовые помощники применяются для реализации адаптивного обучения и поддержки обучающихся [3].

Они позволяют:

- объяснять учебный материал;
- отвечать на вопросы в режиме реального времени;
- сопровождать процесс обучения;

- автоматизировать выполнение заданий.

Особую роль голосовые помощники играют в развитии речевых навыков. Исследования показывают, что такие технологии эффективно используются при обучении иностранным языкам и формировании коммуникативных компетенций.

Кроме того, голосовые помощники могут применяться в образовательной практике, включая физическое воспитание и организацию занятий, что расширяет область их использования.

Классификация подходов к использованию голосовых технологий **Классификация подходов к использованию голосовых технологий**

Методы применения голосовых помощников и речевых технологий в образовательных системах можно классифицировать по функциональному назначению и характеру взаимодействия с пользователем. Наиболее распространённой является классификация, включающая интерфейсные, аналитические и интеллектуальные методы. Каждая из указанных групп выполняет собственные задачи в организации цифрового образовательного процесса.

1. Интерфейсные методы

Интерфейсные методы ориентированы на обеспечение взаимодействия пользователя с информационной системой посредством голосового интерфейса. Основной задачей данной группы методов является упрощение доступа к образовательным ресурсам и повышение удобства использования цифровой среды обучения.

В рамках интерфейсного подхода голосовой помощник выполняет функции посредника между пользователем и системой. Обучающийся получает возможность управлять образовательной платформой с помощью голосовых команд без необходимости использования клавиатуры или графического интерфейса.

К интерфейсным методам относятся:

- голосовой ввод команд и запросов;
- навигация по образовательной системе;
- поиск учебной информации;
- управление электронными ресурсами;
- озвучивание учебного контента;
- голосовое сопровождение пользователя.

Применение интерфейсных методов особенно эффективно в мобильных образовательных приложениях и системах дистанционного обучения, где важна скорость взаимодействия и удобство доступа к информации.

Дополнительное значение данные методы имеют в инклюзивном образовании. Использование голосового интерфейса позволяет обеспечить доступ к образовательным ресурсам для пользователей с ограничениями зрения, моторики или другими особенностями восприятия информации.

Таким образом, интерфейсные методы обеспечивают естественное и удобное взаимодействие человека с образовательной системой, повышая доступность цифровой образовательной среды.

2. Аналитические методы

Аналитические методы направлены на обработку и анализ речевых данных обучающихся с целью получения информации об особенностях их учебной деятельности. В отличие от интерфейсных методов, данная группа ориентирована не только на взаимодействие, но и на исследование характеристик речи пользователя.

Основой аналитических методов являются технологии обработки естественного языка, распознавания речи и интеллектуального анализа данных. Система способна анализировать содержание ответов обучающихся, особенности произношения, речевую активность, эмоциональную окраску речи и другие параметры.

К аналитическим методам относятся:

- анализ правильности произношения;
- оценка речевых навыков;
- определение уровня владения иностранным языком;
- анализ активности обучающегося;
- выявление ошибок и затруднений;
- мониторинг вовлечённости в образовательный процесс.

В образовательной практике аналитические методы особенно востребованы при обучении иностранным языкам, где необходимо оценивать фонетику, интонацию и качество устной речи обучающихся.

Кроме того, речевая аналитика может использоваться для автоматизированной оценки знаний. Голосовой помощник способен анализировать ответы обучающегося и формировать предварительную оценку уровня усвоения материала.

Современные исследования также рассматривают возможность применения аналитических методов для определения эмоционального состояния обучающихся и уровня когнитивной нагрузки в процессе обучения.

Следовательно, аналитические методы позволяют образовательной системе получать дополнительную информацию о процессе обучения и повышать качество контроля знаний.

3. Интеллектуальные методы

Интеллектуальные методы являются наиболее сложным и перспективным направлением развития голосовых технологий в образовании. Они основаны на применении искусственного интеллекта, машинного обучения и нейросетевых моделей для адаптации образовательного процесса под индивидуальные особенности обучающегося.

Главной особенностью интеллектуальных методов является способность системы самостоятельно анализировать данные, обучаться на

основе пользовательского опыта и принимать решения без прямого участия человека.

К интеллектуальным методам относятся:

- адаптивное обучение;
- персонализация учебного контента;
- интеллектуальная генерация ответов;
- прогнозирование успеваемости;
- автоматическое формирование рекомендаций;
- построение индивидуальных образовательных траекторий.

Интеллектуальные голосовые помощники способны учитывать уровень подготовки обучающегося, скорость усвоения материала, предпочтительный формат обучения и результаты предыдущих занятий. На основе этих данных система может изменять сложность заданий, подбирать дополнительные материалы и корректировать образовательный процесс.

Важным направлением является использование больших языковых моделей и генеративного искусственного интеллекта. Такие технологии позволяют голосовым помощникам поддерживать диалог с обучающимся, объяснять сложные темы, генерировать учебные задания и сопровождать процесс самостоятельного обучения.

Интеллектуальные методы также применяются в системах образовательной аналитики, где ИИ используется для выявления закономерностей в учебной деятельности и прогнозирования результатов обучения.

Таким образом, интеллектуальные методы обеспечивают переход от простого голосового интерфейса к созданию полноценных интеллектуальных образовательных систем, способных адаптироваться к потребностям пользователя и повышать эффективность обучения.

Голосовые помощники рассматриваются как новая информационная технология, обладающая значительным потенциалом для развития образовательных систем [2].

Преимущества и ограничения применения

К основным преимуществам относятся:

- повышение доступности обучения;
- персонализация образовательного процесса;
- интерактивность взаимодействия;
- снижение нагрузки на преподавателей [1–3].

Исследования также показывают, что пользователи, особенно дети, склонны воспринимать голосовых помощников как субъектов общения, что повышает вовлечённость в процесс обучения.

Среди ограничений можно выделить:

- ошибки распознавания речи;
- зависимость от технологий;

- риски, связанные с обработкой данных;
- возможное влияние на когнитивные процессы обучающихся.

Перспективы развития голосовых помощников в образовании

Современные тенденции цифровизации образования показывают, что голосовые помощники постепенно становятся не только вспомогательным инструментом, но и полноценным элементом интеллектуальной образовательной среды. Развитие технологий искусственного интеллекта, машинного обучения и больших языковых моделей позволяет значительно расширять функциональные возможности таких систем.

Одним из перспективных направлений является интеграция голосовых помощников с системами адаптивного обучения. В этом случае помощник способен анализировать успеваемость обучающегося, определять уровень усвоения материала и автоматически подбирать индивидуальные рекомендации. Это обеспечивает формирование персонализированной образовательной траектории и повышает эффективность обучения.

Другим важным направлением является использование голосовых технологий в инклюзивном образовании. Голосовые интерфейсы позволяют упростить доступ к образовательным ресурсам для обучающихся с ограниченными возможностями здоровья, включая нарушения зрения, моторики и речи. Благодаря этому образовательная среда становится более доступной и универсальной.

Перспективным также считается применение голосовых помощников в системах дистанционного и смешанного обучения. В условиях удалённого взаимодействия голосовые интерфейсы обеспечивают более естественную коммуникацию между пользователем и образовательной платформой, повышая уровень вовлечённости обучающихся и удобство работы с цифровыми ресурсами.

Кроме того, развитие генеративного искусственного интеллекта позволяет создавать интеллектуальных голосовых ассистентов, способных не только отвечать на вопросы, но и формировать объяснения учебного материала, генерировать задания, проводить тестирование и оценивать результаты обучения. Такие системы могут выступать в роли виртуальных тьюторов и консультантов.

Заключение

Голосовые помощники являются важным элементом цифровой трансформации образования. Их применение способствует повышению эффективности обучения, развитию персонализированных образовательных траекторий и расширению возможностей образовательных систем.

Дальнейшее развитие технологий искусственного интеллекта и речевой обработки позволит интегрировать голосовых помощников в образовательную среду на более глубоком уровне.

Список литературы:

1. Дворянкин О. А. Голосовой помощник в интернете. Куда ведут нас информационные технологии // Молодой ученый. – 2021. – № 18.
2. Пчелинцева Н. В., Пчелинцев С. А., Скрипкина М. А. Популярные голосовые помощники: возможности и тенденции развития // Наука и образование. – 2023.
3. Верещагина Марина Владимировна, Бузоева Залина Султановна, Мулярчук Михаил Викторович НЕКОТОРЫЕ АСПЕКТЫ ПРИМЕНЕНИЯ НЕЙРОЛИНГВИСТИЧЕСКОГО ПРОГРАММИРОВАНИЯ В ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ // Проблемы современного педагогического образования. 2020. №67-3. URL: <https://cyberleninka.ru/article/n/nekotorye-aspekty-primeneniya-neyrolingvisticheskogo-programmirovaniya-v-obrazovatelnom-protsesse> (дата обращения: 24.05.2026).
4. Ушкин С. Г., Коваль Е., Яськин А. Жить с Алисой: как голосовые помощники трансформируют практики коммуникации // Журнал исследований социальной политики. – 2022.
5. Орлов В. А., Гавриленко А. В. Интеллектуальные голосовые помощники: архитектура и принципы построения // Успехи кибернетики. – 2025.
6. Сысоев П. В., Филатов Е. М., Сорокин Д. О. Чат-боты и голосовые помощники в обучении иностранным языкам. – 2024.

УДК 004.056.53:004.7

Дорофеева К. А., Касумова А. У.
РГУ нефти и газа (НИУ) имени И. М. Губкина, Москва, Россия

СРАВНЕНИЕ БЕЗОПАСНОСТИ LLDP И ВЕНДОРОЗАВИСИМЫХ ПРОТОКОЛОВ: АНАЛИЗ ВЕКТОРОВ АТАК И МЕТОДОВ ЗАЩИТЫ В МУЛЬТИ-ВЕНДОРНЫХ СРЕДАХ

Аннотация. В работе сопоставляется защищённость трёх протоколов канального уровня модели OSI, которые отвечают за обнаружение соседних узлов: LLDP, CDP и MNDP. Тестирование велось в изолированной лаборатории. Стенд включал маршрутизаторы Cisco, MikroTik и Eltex. Воспроизведены два сценария: пассивный сбор служебных кадров и активное переполнение плоскости управления поддельными кадрами LLDP. На всех трёх устройствах конфигурационные сведения раскрылись в полном объёме. Наибольшую уязвимость показал MikroTik, одновременно передававший максимальный набор данных по трём протоколам сразу. Активное воздействие заметно ухудшило отклик сети. Среднее время задержки выросло в 2,3 раза, в пиках достигало 11,4 раза. Загрузка процессора увеличилась до 22% на Cisco 1941, 15% на MikroTik CRS326 и 5% на Eltex ESR-15R. По итогам опытов предложен формализованный комплекс защитных мер, оформленный в виде матрицы политик безопасности. В него входят: отключение протоколов на недоверенных интерфейсах, ограничение служебного трафика плоскости управления и сегментация широковещательных доменов средствами виртуальных локальных сетей. Действенность комплекса подтверждена повторным экспериментом.

Ключевые слова: LLDP, CDP, MNDP, IEEE 802.1AB, сетевая безопасность, переполнение плоскости управления, пассивный перехват трафика, мульти-вендорная сеть.

Введение

Протоколы обнаружения соседних устройств на канальном уровне модели OSI, такие как IEEE 802.1AB (Link Layer Discovery Protocol, LLDP), проприетарный Cisco Discovery Protocol (CDP) и MikroTik Neighbor Discovery Protocol (MNDP), широко используются в корпоративных сетях для инвентаризации и диагностики. Все три протокола не имеют механизмов аутентификации и шифрования, что делает их потенциальными векторами атак.

Актуальность работы обусловлена распространённостью мульти-вендорных сетей, где протоколы обнаружения функционируют параллельно, расширяя поверхность атаки. Комплексная сравнительная количественная оценка уязвимостей этих протоколов с применением доверительных интервалов в реальных мульти-вендорных конфигурациях остаётся недостаточно изученной.

Объект исследования — протоколы LLDP, CDP и MNDP. Предмет — механизмы безопасности данных протоколов, уязвимости их реализаций, количественные показатели деградации плоскости управления при атаках и методы минимизации рисков на оборудовании Cisco, Eltex и MikroTik.

Цель работы — сравнительная оценка безопасности указанных протоколов, экспериментальное воспроизведение атак с количественной оценкой деградации и выработка рекомендаций по защите мульти-вендорной инфраструктуры.

Для достижения цели поставлены следующие задачи: 1) выполнить обзор научной литературы, нормативно-технических документов и отраслевых стандартов по тематике безопасности протоколов обнаружения; 2) развернуть лабораторный стенд с маршрутизаторами трёх вендоров и атакующим узлом; 3) измерить легитимный фоновый трафик протоколов обнаружения для обоснования порогов срабатывания политики ограничения; 4) воспроизвести атаку пассивного перехвата трафика и оценить объём раскрываемых данных; 5) воспроизвести атаку переполнения плоскости управления при пяти уровнях интенсивности и зафиксировать показатели деградации; 6) разработать формализованный комплекс защитных мер и провести его экспериментальную верификацию; 7) сформулировать рекомендации по конфигурированию пограничных портов.

Литературный обзор

Базовые принципы работы протоколов обнаружения изложены в практикуме А.Г. Уймина [14]. Роль CDP в построении карты сети раскрыта в [13], где обоснована необходимость отключения протокола на пограничных интерфейсах из-за раскрываемых сведений о версиях ПО. Возможности LLDP для определения топологии через SNMP-запросы рассмотрены в [9]. Комплексный анализ угроз на канальном уровне приведён в [12]; технические аспекты сбора и разбора трафика описаны в [11]. Подходы к защите систематизированы в обзоре [16]: отсутствие аутентификации сглаживается лишь организационными решениями, такими как сегментация виртуальных локальных сетей и фильтрация по MAC-адресу. Современные методы противодействия атакам типа «отказ в обслуживании» описаны в [15].

Существующие исследования преимущественно посвящены сетям одного производителя, тогда как на практике устройства Cisco, MikroTik и Eltex чаще работают в одном сегменте, обрабатывая LLDP, CDP и MNDP параллельно, что существенно расширяет поверхность атаки. В работах [4, 5, 8] обсуждаются пассивный сбор трафика как часть подготовительной фазы целевых атак, ограничения программной обработки служебных кадров и риски смешанных конфигураций без единого шаблона настройки. Нормативно-техническая база опирается на ГОСТ Р 57580.1-2017 [7], ГОСТ Р 52633.0-2006 [6], методический документ ФСТЭК России [10] и рекомендации NIST SP 800-115 [3]. В отличие от перечисленных работ, здесь выполнен сравнительный экспериментальный

анализ встроенных механизмов защиты Cisco, MikroTik и Eltex без специализированных систем обнаружения вторжений, а меры защиты формализованы в виде матрицы политик.

Методология и топология исследования

Эксперимент выполнялся в изолированной лабораторной среде на оборудовании трёх производителей: Cisco, MikroTik и Eltex. Все устройства подключены к одному коммутатору на канальном уровне. Так мы воспроизводим типичную ситуацию корпоративной сети, где несколько маршрутизаторов разных вендоров соединены через агрегирующий коммутатор. Топология представлена на рисунке 1.

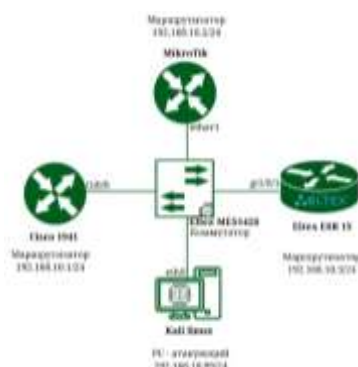


Рисунок 1. Лабораторная топология

Лабораторная топология объединяет пять узлов через коммутатор Eltex MES1428, работающий на втором уровне модели OSI. В качестве маршрутизаторов выступают Cisco 1941, MikroTik CRS326 и Eltex ESR-15R, а роль атакующего исполняет узел на базе Kali Linux. Все устройства расположены в одной IP-подсети 192.168.10.0/24, что имитирует типичный агрегирующий сегмент корпоративной сети. Базовая настройка IP-адресов и интерфейсов приведена в таблице 1.

Таблица 1. Назначение IP-адресов и интерфейсов

Устройство	Интерфейс	IP-адрес	Роль
Cisco 1941	Gi0/0	192.168.10.1/24	Маршрутизатор
MikroTik CRS326	ether1	192.168.10.2/24	Маршрутизатор
Eltex ESR-15R	gi1/0/1	192.168.10.3/24	Маршрутизатор
Kali Linux	eth0	192.168.10.99/24	Атакующий узел
Eltex MES1428	—	—	Коммутатор канального уровня

Методика измерений

Время отклика измерялось командой ping с узла атакующего. В каждой серии отправлялось 10 пакетов, с числом повторных серий $n = 12$ для каждого уровня интенсивности атаки. Этого объёма выборки достаточно для статистической мощноти при ожидаемых нелинейных и асимметричных распределениях метрик. 95% доверительные интервалы рассчитывались по распределению Стьюдента ($t(0,975; df = 11) = 2,201$), значимость различий оценивалась непараметрическим критерием Манна–Уитни. Загрузка процессора фиксировалась штатными средствами каждого производителя, анализ трафика выполнялся анализатором сетевых протоколов с фильтрами по LLDP, CDP и MNDP. Перед основной серией опытов в течение часа замерялся легитимный фоновый трафик: средняя интенсивность составила 0,3–0,4 пакета/с, при пиковом значении в 1,2 пакета/с. Эти цифры в дальнейшем использованы для обоснования порогов срабатывания политики ограничения. На каждом устройстве выполнена базовая настройка интерфейсов и протоколов обнаружения, обеспечивающая одновременную работу LLDP, CDP и MNDP в одном широковещательном домене. На Cisco 1941 и Eltex ESR-15R протокол LLDP активируется командами `lldp run` и `lldp transmit/receive`, который по умолчанию выключен, на MikroTik CRS326 все три рассматриваемых протокола включены сразу из коробки и без дополнительной настройки анонсируют максимальный объём сведений. Подобная конфигурация даёт полную связность на канальном уровне между узлами до начала экспериментов.

Воспроизведение атак

В работе воспроизведены два вектора: пассивный перехват и анализ служебного трафика, а также активное переполнение плоскости управления поддельными LLDP-кадрами. Перехват проводился с атакующего узла при помощи анализатора сетевых протоколов и соответствует методике, изложенной в работе [11]. Из перехваченных кадров LLDP, CDP и MNDP удалось извлечь полные сведения о каждом маршрутизаторе: версию операционной системы, модель оборудования, адрес управления и идентификатор интерфейса. Пассивная атака не создаёт нагрузки на оборудование. Пример захваченного кадра приведён на рисунке 2.

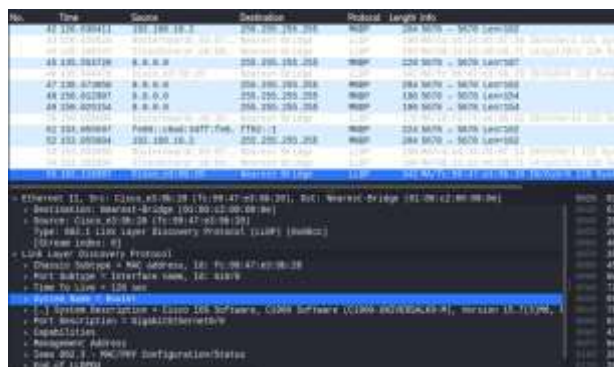


Рисунок 2. Перехваченный кадр LLDP с информацией о Cisco 1941

Активная атака переполнения плоскости управления выполнялась инструментом генерации служебных кадров. Каждый поддельный кадр содержал уникальный адрес источника и идентификатор шасси, что вынуждало маршрутизатор создавать новую запись с собственным таймером удержания. Дополнительно проверялась устойчивость стека к некорректно сформированным служебным полям и инъекции вредоносных значений: все три устройства корректно отбрасывали такие кадры на стадии разбора без переполнения буфера, однако формирование таблицы соседей оставалось уязвимым к описанному ниже переполнению.

В качестве базовой интенсивности выбрана скорость 2500 пакетов/с как умеренно агрессивный сценарий, который превышает типичные лимиты политики ограничения трафика плоскости управления (около 1000 пакетов/с), но не насыщает физический интерфейс. Серия тестов проводилась при пяти уровнях интенсивности: 0 (фон), 500, 1000, 2500 и 5000 пакетов/с. Сводные результаты собраны в таблице 2.

Таблица 2. Влияние интенсивности переполнения плоскости управления на время отклика и загрузку процессора (95% доверительные интервалы, n = 12)

Интенсивность, пакетов/с	Среднее время отклика, мс	Макс. время отклика, мс	Вариация задержки, мс	Cisco, %	MikroTik, %	Eltex, %
0 (фон)	2,1 ± 0,3	3,4	0,3 ± 0,1	3 ± 1	2 ± 1	1 ± 0,5
500	2,6 ± 0,4	6,1	1,2 ± 0,4	8 ± 2	5 ± 1	2 ± 0,5
1000	3,4 ± 0,8	14,2	4,3 ± 1,3	13 ± 3	9 ± 2	3 ± 1
2500	4,8 ± 1,4	38,6	12,4 ± 3,8	20 ± 5	13 ± 3	5 ± 1
5000	8,1 ± 2,2	124,3	38,7 ± 6,1	41 ± 7	28 ± 5	9 ± 2

Деградация имеет нелинейный характер: при 500–1000 пакетов/с изменения остаются умеренными, а на уровне 2500 пакетов/с среднее время отклика растёт уже в 2,3 раза, максимальное — в 11,4 раза, вариация задержки — в 41 раз. Применение критерия Манна–Уитни к выборкам времени отклика при 0 и 2500 пакетов/с подтвердило статистическую значимость различий ($p < 0,001$) для всех трёх устройств. При интенсивности 5000 пакетов/с загрузка процессора Cisco достигает 41%, MikroTik — 28%, появляются потери на втором уровне модели OSI. Показатели Eltex ESR-15R на уровне 5–9% подтверждают высокую устойчивость многоядерной архитектуры с аппаратным выделением ядра плоскости управления. Резкий рост вариации задержки (с 0,3 до 12,4 мс) указывает на состояние процессорного голодания.

Дополнительно проводился тест на исчерпание ёмкости таблицы соседей при непрерывном переполнении плоскости управления интенсивностью 2500 пакетов/с: фиксировались время до отказа, число записей и причина остановки служебной плоскости. Типовой вид роста таблицы на Eltex ESR-15R показан на рисунке 4.

```

esr-15r# show lldp neighbors
Local Interface  Chassis Id      Port Info      System Name
-----
g11/0/1         c4:ad:34:83:6f:2a ether1         Mikrotik
g11/0/1         fc:99:47:e3:9b:28 G10/0         Router
g11/0/1         port0           id
g11/0/1         port2           id
g11/0/1         port4           id
g11/0/1         port10          id
g11/0/1         port11          id
g11/0/1         port12          id
g11/0/1         port13          id
g11/0/1         port14          id
g11/0/1         port15          id
g11/0/1         port16          id
g11/0/1         port18          id
g11/0/1         port21          id
g11/0/1         port22          id
More? Enter - next line; Space - next page; Q - quit; R - show the rest.

```

Рисунок 4. Рост таблицы соседей Eltex ESR-15R при переполнении плоскости управления кадрами LLDP

Количественные показатели роста таблиц соседей и точки отказа каждого устройства приведены в таблице 3.

Таблица 3. Скорость вставки записей и точки отказа таблиц соседей при переполнении плоскости управления интенсивностью 2500 пакетов/с

Устройство	Скорость вставки, зап/с	Точка отказа (время / записей)	Причина отказа плоскости управления
Cisco 1941	310	~100 с / ~1000 записей	Срабатывание встроенного сторожевого механизма; новые объявления отбрасываются
MikroTik CRS326	420	~160 с / ~2048 записей	Деградация хэш-функции, зависание интерфейса управления
Eltex ESR-15R	190	~170 с / ~1000 записей	Достижение программного лимита демона плоскости управления

Скорость вставки записей колеблется в пределах 190–420 в секунду. Это существенно ниже генерируемой интенсивности 2500 пакетов/с. Остальные 80–90% кадров отбрасываются сетевым стекком, что в свою очередь и составляет реальную пропускную способность плоскости управления. Для Cisco 1941 и Eltex ESR-15R предел в районе 1000 записей задают превентивные алгоритмы защиты ОС. На MikroTik CRS326 теоретический лимит RouterOS составляет

8192 записи для 512 МБ ОЗУ, однако фактический отказ наступает уже на 2048, потому что одноядерный процессор ARM 800 МГц не успевает пересчитывать связанные списки и хэш-коллизии.

Таблица 4. Сравнение результатов по устройствам и векторам атаки

Параметр	Cisco 1941	MikroTik CRS326	Eltex ESR-15R
Протоколы по умолчанию	CDP вкл., LLDP выкл.	MNDP, CDP, LLDP вкл.	LLDP выкл.
Раскрываемые пассивно данные	Версия IOS, модель, IP	Версия RouterOS, модель, IP	Версия ESR, модель, IP
Рост загрузки процессора при 2500 пакетов/с, %	20 ± 5	13 ± 3	5 ± 1
Ёмкость таблицы до отказа, записей	~1000	~2048	~1000
Защита по умолчанию	Частичная	Отсутствует	Безопасная конфигурация по умолчанию

Меры защиты

Уязвимость LLDP, CDP и MNDP заключается в том, что служебная информация передаётся в открытом виде, без аутентификации. Поэтому защита строится организационными и конфигурационными методами, согласованными с требованиями ГОСТ Р 57580.1-2017 [7] и методического документа ФСТЭК России [10]. Авторы предлагают формализованный комплекс из трёх взаимодополняющих мер. Первой является отключение протоколов обнаружения на недоверенных интерфейсах, чтобы полностью исключить разведку и пополнение плоскости управления на конкретном порту: на Cisco применяются команды `no cdp enable` и `no lldp transmit/receive`, на MikroTik — ограничение области рассылки командой `/ip neighbor discovery-settings set discover-interface-list=LAN`, на Eltex — `no lldp transmit/receive` на пограничном интерфейсе. Второй мерой является применение политики ограничения трафика плоскости управления с порогом 1000 пакетов/с по совокупности служебных протоколов. Замер фонового трафика показал, что средняя интенсивность не превышает 1,2 пакета/с, поэтому выбранный порог обеспечивает запас на три порядка относительно нормального уровня и гарантированно отсекает атакующие интенсивности от 2500 пакетов/с и выше. Третья мера — микросегментация широковещательных доменов средствами виртуальных локальных сетей; поскольку протоколы обнаружения соседних устройств не пересекают границы третьего уровня модели OSI, разделение сети на

виртуальные сегменты гарантирует, что шторм служебного трафика из клиентского сегмента не достигнет интерфейсов агрегирующих маршрутизаторов. Формализованная матрица политик безопасности приведена в таблице 5. Каждой строке соответствует тип интерфейса с присвоенным уровнем доверия, а столбцы определяют профиль ограничения трафика, правило сегментации и набор разрешённых протоколов обнаружения.

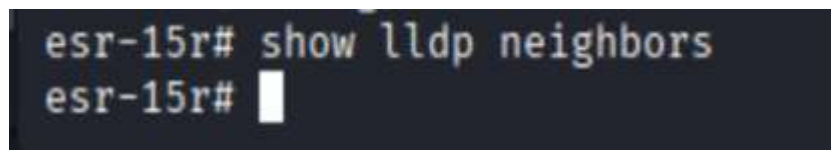
Таблица 5 – Матрица политик безопасности для протоколов обнаружения соседних устройств

Тип интерфейса	Уровень доверия	Разрешённые протоколы	Профиль ограничения трафика	Правило сегментации
Магистральный (uplink) между маршрутизаторами	Доверенный	LLDP (приём и передача)	1000 пакетов/с	Управляющая виртуальная локальная сеть, изолированная от пользовательского сегмента
Доступа (access) для рабочих станций	Недоверенный	Все протоколы обнаружения отключены	0 пакетов/с (полный отбрасывающий фильтр)	Пользовательская виртуальная локальная сеть с привязкой к порту
Гостевой (guest) или открытый	Враждебный	Все протоколы обнаружения отключены	0 пакетов/с (полный отбрасывающий фильтр на канальном уровне)	Изолированная виртуальная локальная сеть без маршрутизации в корпоративный сегмент

Согласно архитектурным принципам защиты плоскости управления [1, 15], политика ограничения формализуется как тройка (Классификатор, Профиль, Действие). Классификатор сопоставляет заголовки канального уровня с групповыми адресами LLDP (01:80:C2:00:00:0E), CDP (01:00:0C:CC:CC:CC) или MNDP (UDP-порт 5678) при условии, что входной интерфейс относится к недоверенным. Профиль задаётся как ограничитель по среднему числу пакетов в секунду со значением всплеска: для доверенных интерфейсов 1000 пакетов/с при всплеске 1500; для недоверенных 0 пакетов/с, что является полным запрещением. При превышении лимита кадр отбрасывается, счётчик увеличивается, событие записывается в журнал безопасности; при сохранении превышения в течение 60 секунд формируется уведомление по протоколу системного журналирования.

Логика принятия решений по настройке интерфейса задаётся набором правил «ЕСЛИ–ТО». R1: для магистрального доверенного интерфейса включить LLDP, ограничение 1000 пакетов/с, управляющую виртуальную сеть; CDP и MNDP отключить. R2: для интерфейса доступа отключить все протоколы обнаружения, применить полный отбрасывающий фильтр, назначить пользовательскую виртуальную сеть. R3: для гостевого или враждебного интерфейса применить R2 и дополнительно отбрасывать кадры с групповыми адресами 01:80:C2:00:00:0E и 01:00:0C:CC:CC:CC. R4: при превышении 60 событий отбрасывания за 60 секунд формировать инцидент в системе сбора и корреляции событий безопасности.

Для проверки эффективности комплекса проведено повторное тестирование атаки переполнения с интенсивностями 2500 и 5000 пакетов/с после применения матрицы политик на всех трёх маршрутизаторах. Таблицы соседей остаются пустыми, IP-связность между доверенными узлами сохраняется без потерь, а служебные кадры на пограничных интерфейсах не фиксируются. Среднее время отклика возвращается к фоновому уровню (2,1–2,2 мс), загрузка процессора снижается до 1–4%, потери на втором уровне отсутствуют даже при интенсивности в размере 5000 пакетов/с. Критерий Манна–Уитни для выборок до и после защиты подтверждает значимость различий на всех маршрутизаторах. Результат на Eltex ESR-15R приведён на рисунке 5.



```
esr-15r# show lldp neighbors
esr-15r#
```

Рисунок 5. Чистая таблица соседей на Eltex ESR-15R

Заключение

Протоколы автоматического обнаружения соседних устройств канального уровня создают реальную угрозу безопасности в гетерогенных корпоративных сетях. Пассивная разведка успешно реализуется на всех трёх платформах: MikroTik CRS326 в исходной конфигурации передаёт максимальный объём системных сведений сразу по трём протоколам; Cisco IOS безопаснее в части LLDP, однако включённый CDP даёт столь же полное раскрытие данных; Eltex ESR-15R, демонстрирует подход безопасной конфигурации по умолчанию. Атака переполнения плоскости управления интенсивностью 2500 пакетов/с вызвала достоверный рост среднего времени отклика более чем в два раза и загрузки процессоров в диапазоне 5–20%, отказ таблиц соседей наступал в пределах 1000–2048 записей. Лучшую устойчивость показал Eltex ESR-15R благодаря многоядерной архитектуре с аппаратным выделением ядра плоскости управления.

Для безопасности сетей с разнородным оборудованием необходимы три согласованные меры: отключение протоколов на граничных интерфейсах, политика ограничения трафика плоскости управления с порогом 1000 пакетов/с и сегментация широковещательных доменов средствами виртуальных локальных сетей. Комплекс формализован в виде матрицы политик и набора правил «ЕСЛИ–ТО», что обеспечивает машинную интерпретацию в средствах автоматизации управления безопасностью. Эффективность подтверждена повторным экспериментом: метрики деградации возвращаются к фоновым значениям, таблицы соседей остаются пустыми даже при 5000 пакетов/с. Стандартный LLDP и проприетарные протоколы одинаково уязвимы к пассивному перехвату, поэтому эффективная защита достигается только организационно-техническими мерами, согласованными с требованиями ГОСТ Р 57580.1-2017 и методических документов ФСТЭК России.

Список литературы:

1. Cisco IOS Quality of Service Solutions Configuration Guide. Control Plane Policing / Cisco Systems, Inc. — 2023. — URL: https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/qos_plcshp/configuration/15-mt/qos-plcshp-15-mt-book/qos-plcshp-ctrl-pln-plc.html (дата обращения: 29.03.2026).
2. Parsons, G. IEEE Standard for Local and Metropolitan Area Networks — Station and Media Access Control Connectivity Discovery / G. Parsons. — NY.: IEEE, 2016. — 146 p.
3. Scarfone, K., Souppaya, M., Cody, A., Orebaugh A. NIST Special Publication 800-115. Technical Guide to Information Security Testing and Assessment / K. Scarfone, M. Souppaya, A. Cody, A. Orebaugh. — National Institute of Standards and Technology: U.S. Department of Commerce. — 2008. — 80 p.
4. Бутаков, Л. О., Морозов Т. А. Методы тестирования и защиты от атаки VLAN Hopping / Л. О. Бутаков, Т. А. Морозов // Теория и практика современной науки. — 2025. — №6 (120). — С. 104–115.
5. Вартамян, А. А. Угрозы и атаки сетевой безопасности / А. А. Вартамян // E-Scio. — 2023.
6. ГОСТ Р 52633.0-2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. — М.: Стандартинформ. — 2007. — 20 с.
7. ГОСТ Р 57580.1-2017. Безопасность финансовых (банковских) операций. Защита информации финансовых организаций. Базовый состав организационных и технических мер. — М.: Стандартинформ. — 2017. — 92 с.
8. Зуфарова, А. С., Кошелева, А. Д. Что такое MITRE ATT&CK: разбор популярной тактической модели / А. С. Зуфарова, А. Д. Кошелева // Информатика. Экономика. Управление. — 2025. — №4 (1). — С. 2027–2037.
9. Лагутин, И. А. Определение топологии с помощью протокола LLDP в сетях Juniper / А. М. Лагутин // NVisionGroup. — URL: <https://cyberleninka.ru/article/n/opredelenie-topologii-s-pomoschyu-protokola-ldp-v-setyah-juniper> (дата обращения: 29.03.2026).
10. Меры защиты информации в государственных информационных системах: методический документ / ФСТЭК России. — Утверждён 11 февраля 2014 г. — URL: <https://fstec.ru/dokumenty/vse-dokumenty/spetsialnye-normativnye-dokumenty/metodicheskij-dokument-ot-11-fevralya-2014-g> (дата обращения: 29.03.2026).

11. Мешкова, Е. В. Перехват и анализ сетевого трафика с помощью «Wireshark» / Е. В. Мешкова // Контентус. — 2016.
12. Муханова, А. А., Ревнивых, А. В., Федотов, А. М. Классификация угроз и уязвимостей информационной безопасности в корпоративных системах / А. А. Муханова, А. В. Ревнивых, А. М. Федотов // Вестник Новосибирского государственного университета. — 2013. — № 2 (11) — С. 55–76. — URL: <https://cyberleninka.ru/article/n/klassifikatsiya-ugroz-i-uyazvimostey-informatsionnoy-bezopasnosti-v-korporativnyh-sistemah> (дата обращения: 29.03.2026).
13. Рубашенков, А. М. Протокол CDP / А. М. Рубашенков, А. В. Бобров // Academy. — 2018. — № 6(33). — URL: <https://cyberleninka.ru/article/n/protokol-cdp> (дата обращения: 29.03.2026).
14. Уймин, А. Г., Толмачев И. М. Применение отечественного сетевого оборудования Eltex и EсоRouter в рамках специальности 09.02.06 "Сетевое и системное администрирование". Вопросы импортозамещения и подготовки квалифицированных кадров в сетевом оборудовании / А. Г. Уймин, И. М. Толмачев // Автоматизация и информатизация ТЭК. – 2025. – № 11(628). – С. 58-62.
15. Ширин, К. О. Современные подходы к решению проблемы защиты от сетевых атак «Отказ в обслуживании»: системы автоматического предотвращения вторжений / К. О. Ширин // Т-Comm - Телекоммуникации и Транспорт. — 2011. — С. 161–163.
16. Щерба, Е. В., Никонов, В. И., Литвинов, Г. А. Обеспечение безопасности протоколов маршрутизации для телекоммуникационных сетей с динамической топологией / Е. В. Щерба, В. И. Никонов, Г. А. Литвинов // Доклады ТУСУР. — 2018. — №3 (21). — С. 19–29. — URL: <https://cyberleninka.ru/article/n/obespechenie-bezopasnosti-protokolov-marshrutizatsii-dlya-telekommunikatsionnyh-setey-s-dinamicheskoy-topologiey> (дата обращения: 29.03.2026).

УДК 621.3.049.774:629.78:539.1

*Соловьев В. А., Чаплыгин С. Н., Алабов С. С., Сапунов Д. М.
МГТУ им. Н. Э. Баумана, Москва, Россия*

РАДИАЦИОННАЯ СТОЙКОСТЬ ИНТЕГРАЛЬНЫХ МИКРОСХЕМ В УСЛОВИЯХ КОСМИЧЕСКОГО ПРОСТРАНСТВА: МЕХАНИЗМЫ ПОВРЕЖДЕНИЯ И МЕТОДЫ ЗАЩИТЫ

Аннотация. В настоящей работе проведен обзор механизмов радиационных повреждений интегральных микросхем, эксплуатируемых в условиях космического пространства, систематизированы методы повышения их радиационной стойкости. Приведена классификация радиационных эффектов. Подробно рассмотрены физические механизмы каждого класса эффектов в кремниевых МОП- и биполярных структурах, включая эффект пониженной мощности дозы (ELDRS) в биполярных приборах, образование точечных дефектов и кластеров в кристаллической решётке, ионизационно-индуцированный заряд критической величины в чувствительных узлах ИМС. Рассмотрены технологические, конструктивные, схемотехнические и программные методы повышения радиационной стойкости, а также вопросы экранирования. Обсуждается применение коммерческих компонентов (COTS) в условиях с увеличением числа миссий в рамках концепции «Нового космоса».

Ключевые слова: радиационная стойкость, интегральные микросхемы, эффект суммарной дозы, одиночные радиационные эффекты, смещение атомов.

1 Введение

Изучение влияния радиационного излучения на электронные компоненты и устройства началось в середине XX века в связи с потребностью в электронике, способной выдерживать воздействие ядерного взрыва, а также с освоением космоса и открытием радиационных поясов Земли. На сегодняшний день оба запроса сохраняются и растут.

Стремительный рост числа космических миссий, прежде всего в рамках концепции «Нового космоса», определяет необходимость применения электронной компонентной базы в условиях жёсткого радиационного воздействия. Если ранее задача обеспечения радиационной стойкости решалась преимущественно для ограниченного класса государственных и оборонных программ, то к середине 2020-х годов сотни коммерческих спутников ежегодно выводятся на различные орбиты компаниями со всего мира, причём среди их разработчиков велика доля тех, кто не располагает опытом работы с радиационно-стойкими компонентами [1].

Применительно к орбитальным аппаратам ключевыми источниками ионизирующих излучений являются: галактические космические лучи — высокоэнергетические ядра элементов, проникающие из-за пределов Солнечной системы; солнечные энергичные частицы, интенсивность потока которых резко возрастает в периоды вспышек и корональных выбросов масс; радиационные

пояса Земли — внутренний пояс, заполненный преимущественно протонами, и внешний пояс, заполненный электронами [2]. Радиационная нагрузка существенно зависит от орбиты космического аппарата: воздействие значительно выше на средних и геостационарных орбитах по сравнению с низкой околоземной [3].

Целью работы является обзор механизмов радиационного повреждения интегральных микросхем (ИМС) в условиях космического пространства и систематизация современных методов повышения их радиационной стойкости. В разделе 2 приведена общая классификация радиационных эффектов; в разделах 3–5 рассмотрены эффект суммарной дозы, эффект смещения атомов и одиночные радиационные эффекты; раздел 6 посвящён обзору методов защиты — технологических, конструктивных, схемотехнических, программных, а также экранированию и перспективным направлениям.

2 Классификация и механизмы радиационных эффектов

Радиационные эффекты в ИМС принято классифицировать по характеру взаимодействия частиц или квантов с веществом полупроводника, по типу вызываемых повреждений и по обратимости их последствий [4, 9].

По характеру взаимодействия выделяют ионизационные эффекты, обусловленные образованием электронно-дырочных пар в полупроводнике или подзатворном диэлектрике, и эффекты упругого и неупругого рассеяния, приводящие к смещению атомов кристаллической решётки. По характеру накопления повреждений радиационные эффекты разделяют на дозовые (накопительные), при которых параметры приборов изменяются постепенно по мере накопления поглощённой дозы, и одиночные (случайные), при которых повреждение возникает в результате воздействия отдельной заряженной частицы и проявляется случайным образом во времени [9].

К дозовым эффектам относятся: эффект суммарной поглощённой дозы (Total Ionizing Dose, TID; ЭСД), связанный с накоплением заряда в диэлектрических слоях и образованием поверхностных состояний на границе раздела полупроводник–диэлектрик; эффект смещения атомов (Displacement Damage, DD; структурные повреждения или СА), связанный с образованием дефектов в кристаллической решётке. К одиночным радиационным эффектам (ОПЭ, Single Event Effects, SEE) относятся обратимые сбои (SEU, SET), не приводящие к физическому разрушению структуры, и потенциально разрушающие эффекты (SEL, SEB, SEGR), способные привести к катастрофическому отказу прибора [8, 21].

Деление эффектов на обратимые и необратимые тесно связано с их физической природой. Накопление заряда в подзатворных оксидах МОП-транзисторов при ЭСД частично снимается термическим отжигом и поэтому формально является частично обратимым на больших временных масштабах [6, 13]. Накопление структурных дефектов при СА в практически значимых условиях необратимо [19]. Сбои SEU и SET являются обратимыми по своей

природе (информация в ячейке может быть восстановлена), тогда как SEL может перейти в катастрофический отказ из-за разрушения металлизации, а SEB и SEGR практически всегда приводят к необратимому повреждению структуры прибора [9, 11]. Полная классификация радиационных эффектов в ИМС представлена в таблице 1.

Таблица 1 — Классификация радиационных эффектов в ИМС

Эффект	Тип эффекта	Природа взаимодействия	Обратимость	Уязвимая группа приборов
ЭСД (TID)	Накопительный	Ионизация в диэлектриках	Частично (отжиг)	МОП и биполярные ИМС
ELDRS	Накопительный	Ионизация при низкой мощн. дозы	Частично	Биполярные линейные ИС
СА (DD)	Накопительный	Упругое и неупругое смещение	Необратим	Биполярные ИС, приборы с зарядовой связью, солнечные элементы
SEU	Одиночный	Ионизация	Обратим (перезапись)	Память, регистры, защёлки
SET	Одиночный	Ионизация	Обратим (фильтрация)	Комбинационная логика
MCU / MBU	Одиночный	Ионизация (один трек — неск. ячеек)	Зависит от кода коррекции	ОЗУ высокой плотности
SEL	Одиночный	Включение паразитного тиристора	Обратим сбросом питания; возможен необратимый отказ	КМОП ИМС с объёмной подложкой
SEFI	Одиночный	Отказ управляющей логики	Обратим (перезапуск)	СБИС, FPGA, микроконтроллеры
SEB	Одиночный	Вторичный пробой структуры	Необратим	Силовые MOSFET, биполярные транзисторы
SEGR	Одиночный	Пробой подзатворного диэлектрика	Необратим	Силовые MOSFET, ЭСПЗУ

Чувствительность конкретного прибора к тому или иному эффекту существенно зависит от технологии его изготовления, конструктивных особенностей и режимов работы [22]. В частности, переход к субмикронным и нанометровым техпроцессам резко снизил вклад накопления заряда в подзатворных оксидах в общую радиационную деградацию параметров, одновременно повысив роль одиночных эффектов и накопления заряда в боковой изоляции STI [14, 27]. Поэтому современное проектирование радиационно-стойких ИМС требует одновременного учёта всех перечисленных классов эффектов на этапах выбора технологии, разработки топологии и схемотехнической реализации.

3 Эффект суммарной дозы (ЭСД, TID)

Эффект суммарной дозы характеризует постепенное изменение электрических параметров полупроводниковых приборов вследствие накопления поглощённой ионизирующей дозы. В качестве количественной меры используется поглощённая доза в кремнии, выраженная в радах ($1 \text{ рад} = 0,01 \text{ Гр}$) или в Грехах. Типичные уровни накопленной дозы за время эксплуатации космического аппарата составляют единицы — десятки крад на низкой околоземной орбите и от сотен крад до Мрад на средних и геостационарных орбитах [3, 35].

3.1 Механизм ЭСД в МОП-структурах

Модель ЭСД в МОП-структурах [6, 13] выделяет четыре последовательных физических процесса в подзатворном диэлектрике (SiO_2):

1) образование электронно-дырочных пар при поглощении энергии ионизирующего излучения. Для SiO_2 средняя энергия образования пары составляет около 17 эВ; число пар прямо пропорционально поглощённой дозе;

2) быстрая рекомбинация части пар. Доля электронов и дырок, избежавших начальной рекомбинации, определяется напряжённостью электрического поля в оксиде и видом ионизирующего излучения. При больших полях и слабоионизирующих частицах (γ -кванты, электроны) выход неравновесных носителей близок к единице, при сильноионизирующих частицах он существенно ниже из-за плотного трекового образования;

3) транспорт носителей в оксиде. Электроны быстро покидают оксид благодаря высокой подвижности, тогда как дырки переносятся медленным прыжковым механизмом по локализованным состояниям. Время дрейфа дырок к границе раздела SiO_2/Si может составлять от пикосекунд до часов в зависимости от температуры и поля;

4) захват дырок на ловушках в объёме оксида и образование поверхностных состояний на границе раздела Si/SiO_2

Накопленный в оксиде положительный заряд и плотность поверхностных состояний приводят к смещению порогового напряжения МОП-транзисторов (ΔV_{th}), к увеличению токов утечки и снижению крутизны. Знак смещения противоположен для n- и p-канальных транзисторов: у n-МОП ΔV_{th} носит

отрицательный характер, поэтому при больших дозах транзистор может перейти в открытое состояние при нулевом напряжении на затворе — это наблюдается экспериментально как радиационно-индуцированный ток утечки и в предельном случае ведёт к функциональному отказу схемы [7, 17].

В современных субмикронных технологиях толщина подзатворного оксида уменьшена до 1–2 нм, и накопление заряда в этом тонком слое перестало быть доминирующим механизмом ЭСД-деградации [14, 27]. Основной вклад в радиационный отклик современных КМОП-приборов вносит накопление заряда в боковой изоляции (Shallow Trench Isolation, STI), толщина которой остаётся сравнительно большой (сотни нанометров). Заряд, захваченный в STI, индуцирует паразитные каналы утечки вдоль границы транзистора, что увеличивает токи в выключенном состоянии и может привести к функциональному отказу схемы при дозах в десятки крэд. Для систематизации наблюдаемых явлений в [17] предложена классификация микродозовых эффектов в нанoeлектронных структурах, в которой паразитные каналы в STI рассматриваются как самостоятельный класс эффектов суммарной дозы.

3.2 ЭСД в биполярных приборах. ELDRS

В биполярных приборах ЭСД проявляется через образование заряженных ловушек и поверхностных состояний в пассивирующих оксидах, покрывающих переходы база–эмиттер и база–коллектор [5]. Накопленные поверхностные состояния увеличивают скорость поверхностной рекомбинации неосновных носителей в базе, что приводит к росту входного тока базы и снижению коэффициента усиления транзисторов. Особенно чувствительны к этому эффекту входные каскады прецизионных операционных усилителей и компараторов, в которых деградация ведёт к росту входных токов смещения и сдвигу напряжения смещения нуля.

Серьёзной практической проблемой биполярной радиационно-стойкой электроники является эффект пониженной мощности дозы (Enhanced Low Dose Rate Sensitivity, ELDRS): деградация параметров биполярных приборов при облучении с малой мощностью дозы, характерной для космических миссий (10^{-4} – 10^{-2} рад/с), оказывается заметно сильнее, чем при той же суммарной дозе и характерной для наземных испытаний высокой мощности (50–300 рад/с) [15]. Впервые эффект был идентифицирован для дискретных биполярных транзисторов в 1991 г. и подтверждён для биполярных линейных микросхем в 1994 г.; в дальнейшем в компендиумах NASA были идентифицированы десятки распространённых типов линейных микросхем, чувствительных к ELDRS [15].

Физическая природа ELDRS связана с конкуренцией процессов захвата дырок и водородных частиц на ловушках в объёме оксида и на границе раздела. При высокой мощности дозы значительная часть дырок успевает локализоваться в объёме оксида и впоследствии отжигается; при низкой мощности дозы доминирует медленный процесс формирования стабильных поверхностных состояний, обуславливающий усиленную деградацию [15, 16].

В отечественной литературе обобщённая модель ELDRS, учитывающая преобразование объёмных ловушек в поверхностные состояния, рассмотрена в работе [16]. Учёт ELDRS критически важен при испытаниях биполярных линейных микросхем для космических применений: испытания при стандартной мощности дозы могут существенно недооценивать фактическую деградацию.

3.3 Параметрический отклик и моделирование

С точки зрения проектировщика аппаратуры основными контролируемыми параметрами при оценке ЭСД-стойкости являются: смещение порога ΔV_{th} и подпороговый ток для МОП-транзисторов; ток базы и коэффициент усиления для биполярных транзисторов; смещения входных токов и напряжений, а также изменение скорости нарастания для аналоговых ИМС; пороговое напряжение функционального отказа для цифровых ИМС [11]. Прогнозирование ЭСД-отклика выполняется на основе аналитических моделей и приборно-технологического моделирования (TCAD), а экспериментальная аттестация — по международным (MIL-STD-883 Test Method 1019, ESCC 22900) и отечественным стандартам [12].

4 Смещение атомов (CA, Displacement Damage)

Эффект смещения атомов представляет собой кумулятивные радиационные повреждения, обусловленные передачей энергии от налетающих частиц атомам кристаллической решётки полупроводника. Если переданная энергия превышает пороговое значение энергии смещения (для кремния около 21 эВ), атом покидает узел решётки, и образуется пара Френкеля — вакансия и междоузельный атом [18, 19]. При энергиях налетающей частицы, существенно превышающих порог, образованные первичные смещённые атомы (Primary Knock-on Atom, PKA) сами выступают как ионизирующие частицы и инициируют каскады последующих смещений, в результате чего формируются кластеры дефектов. Тяжёлые ионы и быстрые нейтроны производят преимущественно кластерные дефекты, лёгкие электроны — точечные. Протоны промежуточных энергий (десятки МэВ — ГэВ), доминирующие в радиационных поясах Земли, занимают промежуточное положение [19].

Для количественного описания DD используется концепция неионизирующих потерь энергии (Non-Ionizing Energy Loss, NIEL). Модель постулирует, что радиационная деградация параметров приборов, обусловленная DD, пропорциональна интегральному NIEL-флюенсу частиц. Это позволяет пересчитывать эквивалентные дозы DD между различными видами частиц и энергиями, что является основой методик испытаний DD-эффектов. Модель NIEL хорошо работает для протонов и нейтронов в широком диапазоне энергий, но даёт существенные отклонения для тяжёлых ионов из-за внутрикаскадной рекомбинации точечных дефектов в плотных треках [19].

Образованные при DD дефекты служат центрами рекомбинации и захвата неосновных носителей, что приводит к ряду параметрических деградаций. В

биполярных транзисторах снижается время жизни неосновных носителей в базе и, как следствие, снижается коэффициент усиления β ; этот эффект делает биполярные технологии существенно более чувствительными к DD по сравнению с МОП [18]. В ПЗС и КМОП-фотоприёмниках наблюдается рост темнового тока, появление «горячих» пикселей, шум случайных телеграфных сигналов (Random Telegraph Signal, RTS) [19]; эти эффекты ограничивают характеристики оптических приборов наблюдения Земли и научной аппаратуры дальних космических миссий. В солнечных элементах деградация диффузионной длины носителей приводит к снижению тока короткого замыкания и КПД: на геостационарной орбите накопленный за десятилетний срок миссии DD-флюенс протонов и электронов вызывает падение мощности солнечных батарей на десятки процентов; этот фактор является одним из основных при проектировании энергоустановок космических аппаратов. Наконец, деградация наблюдается и в светодиодах и лазерных диодах, что значимо для оптоэлектронных компонентов космической связи и оптронных развязок [19].

В МОП-структурах вклад DD в общую радиационную деградацию обычно мал по сравнению с TID, поскольку проводимость канала формируется основными носителями в тонком приповерхностном слое, на параметры которого слабо влияют объёмные дефекты решётки. Однако в широкозонных полупроводниках (GaN, SiC), применяемых в перспективной силовой электронике космических аппаратов, относительная роль DD возрастает, и для них накопление структурных дефектов исследуется как самостоятельный вид отказа [20].

5 Одинокные радиационные эффекты (ОРЭ, SEE)

Одинокные радиационные эффекты возникают в результате прохождения через чувствительный объём прибора отдельной высокоэнергетической заряженной частицы — иона или протона. На своём пути частица оставляет плотный трек электронно-дырочных пар; собранный заряд может изменить логическое состояние ячейки памяти, вызвать кратковременный импульс в аналоговой или комбинационной схеме, инициировать включение паразитных структур или привести к локальному пробоему [21].

Количественной характеристикой чувствительности прибора к ОРЭ является сечение эффекта σ , измеряемое в см^2 и зависящее от линейной передачи энергии (ЛПЭ, Linear Energy Transfer, LET) налетающей частицы. Зависимость $\sigma(\text{LET})$ имеет характерную S-образную форму: ниже порогового значения LET_{th} эффект практически не наблюдается, выше — сечение растёт и при достаточно высоких ЛПЭ выходит на насыщение [23, 25]. Для космических применений сечение интегрируется по реальному спектру ЛПЭ в условиях миссии и пересчитывается в ожидаемую частоту сбоев (single event rate) [24, 38]. Принято выделять следующие основные типы ОРЭ.

5.1 Одинокные сбои (SEU, Single Event Upset)

SEU представляет собой обратимое изменение логического состояния ячейки памяти, регистра или защёлки под действием одиночной ионизирующей частицы. Механизм возникновения следующий: ионизация полупроводниковой подложки приводит к образованию заряда в чувствительной области стока МОП-транзистора в ячейке памяти; собранный заряд Q разряжает узел через паразитную ёмкость, и если он превышает критический заряд ячейки Q_c , происходит инверсия логического состояния [21]. Критический заряд тем меньше, чем меньше топологические размеры и напряжение питания, поэтому SEU-чувствительность ИМС резко растёт с переходом к новым технологическим узлам.

С масштабированием технологии в одном треке частицы стали оказываться сразу несколько чувствительных узлов, что приводит к множественным сбоям в одной ячейке (MBU, Multiple Bit Upset) или одновременным сбоям нескольких смежных ячеек (MCU, Multiple Cell Upset) [22]. Кроме того, появился отдельный класс одиночных функциональных отказов SEFI (Single Event Functional Interrupt) - сбоя в управляющей логике сложных СБИС (FPGA, контроллеры памяти), приводящих к временной потере работоспособности устройства до перезагрузки [11].

5.2 Одиночные кратковременные импульсы (SET, Single Event Transient)

SET представляют собой кратковременные импульсы напряжения или тока, возникающие в комбинационных цифровых или аналоговых цепях в результате сбора ионизационно-индуцированного заряда. В аналоговых ИМС (операционных усилителях, компараторах) SET может вызвать ошибочное срабатывание; в комбинационной цифровой логике SET распространяется по последующим вентилям и при попадании в окно отбора защёлки или триггера фиксируется как ошибочный SEU [21]. С увеличением рабочих частот ИМС вероятность фиксации SET растёт, и для современных нанометровых КМОП-схем SET становится не менее значимой проблемой, чем классический SEU.

Особенностью SET по сравнению с SEU является зависимость от длительности и формы импульса, а не только от собранного заряда. В работах [8, 23] исследуется влияние импульсного характера ионизирующего излучения на параметры чувствительности ИС к ОРЭ; показано, что классическое описание через критический заряд и пороговое значение ЛПЭ требует уточнения для коротких треков и сильноионизирующих частиц.

5.3 Одиночное «защёлкивание» (SEL, Single Event Latchup)

SEL представляет собой включение паразитного тиристора, образованного четырьмя последовательными областями n-p-n-p в КМОП-структуре с объёмной подложкой. В нормальном режиме паразитный тиристор остаётся закрытым; однако ионизационный ток одиночной частицы способен сместить переход база–эмиттер одного из паразитных биполярных транзисторов в прямом направлении и инициировать самоподдерживающееся включение [21].

После включения через тиристор течёт большой ток, ограничиваемый только сопротивлением шин и источника питания; устранение состояния возможно только полным сбросом питания. Если ограничение тока недостаточно, локальный перегрев способен разрушить металлизацию и привести к необратимому отказу ИМС [11].

Важно отметить, что SEL принципиально невозможен в технологиях «кремний на диэлектрике» (КНИ, SOI) и «кремний на сапфире» (КНС, SOS), поскольку в этих структурах между активными областями отсутствует общая подложка и паразитный тиристор не формируется. Этот фактор является одним из главных аргументов в пользу применения КНИ-технологий для радиационно-стойких ИМС [9, 30]. Сравнение структур объемного КМОП и КНИ применительно к механизму SEL показано на рисунке 1.

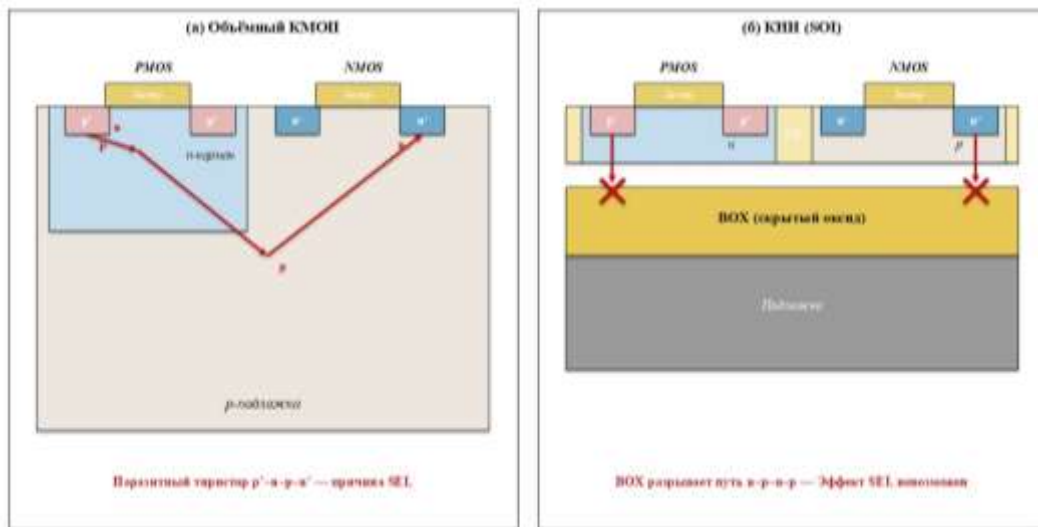


Рисунок 1 – Структура а)объемного КМОП с паразитным тиристором б)КНИ, скрытый оксид BOX предотвращает тиристорный эффект

5.4 Одиночный пробой (SEB, Single Event Burnout)

SEB представляет собой необратимый отказ силового MOSFET или биполярного транзистора, возникающий при попадании тяжёлой ионизирующей частицы в высоковольтную область прибора. Механизм заключается в индуцированном частицей включении паразитного биполярного транзистора, образованного истоком n^+ , p -карманом и стоком n^- силового MOSFET. Включение паразитного транзистора при высоком напряжении сток–исток вызывает лавинный пробой и вторичный тепловой пробой структуры, что приводит к необратимому повреждению кристалла [34].

SEB особенно значим для мощной электроники космических аппаратов: преобразователей напряжения, регуляторов, силовых ключей. Для широкозонных полупроводников (SiC, GaN) механизмы SEB исследуются как отдельная актуальная задача: эти материалы обладают высокой пороговой энергией дефектообразования и большим запасом по электрическому пробую,

однако демонстрируют существенно более низкие пороги SEB по сравнению с теоретически ожидаемыми [20, 34].

5.5 Пробой подзатворного диэлектрика (SEGR, Single Event Gate Rupture)

SEGR представляет собой необратимый локальный пробой подзатворного диэлектрика силового MOSFET или ячейки энергонезависимой памяти (ЭСПЗУ, Flash) под действием одиночной тяжёлой частицы. Механизм заключается в одновременном существовании двух условий: высокого приложенного напряжения на затворе и плотного ионизационного трека в подложке непосредственно под подзатворным диэлектриком. Заряд трека вызывает локальное усиление поля в диэлектрике, превышающее пробивное, и формирует проводящий канал - место необратимого пробоя [34]. SEGR характеризуется выраженной зависимостью от напряжения на затворе: с его ростом пороговое значение ЛПЭ для пробоя резко падает.

5.6 Методы оценки чувствительности к ОРЭ

Стандартные методы испытаний ОРЭ-стойкости включают облучение на ускорителях тяжёлых ионов с диапазоном ЛПЭ от единиц до сотен МэВ·см²/мг и облучение протонами на синхротронных пучках. Для предварительной оценки и серийного контроля применяются лазерные имитационные методы, основанные на воздействии сфокусированного импульса фемтосекундного или пикосекундного лазера на чип, демонтированный с верхней разводки [23]. Хотя лазерные методы не воспроизводят полностью трек тяжёлой частицы, они позволяют получить количественные оценки порогов SEE и идентифицировать чувствительные узлы топологии при существенно меньших затратах, чем испытания на ускорителях.

Расчётное прогнозирование ОРЭ-стойкости основано на интегрировании измеренного сечения по ожидаемому спектру ЛПЭ космического излучения и проводится в соответствии с международными и отечественными стандартами [12, 24]. В работе [24] рассмотрены модели оценки соответствия ИМС требованиям по стойкости к воздействию тяжёлых заряженных частиц.

6 Методы повышения радиационной стойкости

Повышение радиационной стойкости ИМС является комплексной задачей, охватывающей все уровни проектирования, от выбора технологии и материала подложки до системного программного обеспечения. Каждый уровень имеет свою область эффективности, и в реальных проектах применяется комбинация нескольких методов, выбираемая исходя из задач космического аппарата, допустимой массы и доступной элементной базы [9, 12]. Ниже рассмотрены технологические, конструктивные, схемотехнические и программные методы, экранирование, а также перспективные направления.

6.1 Технологические методы

Технологические методы направлены на создание изначально радиационно-стойкого прибора путём изменения структуры подложки, профиля

легирования или диэлектрических слоёв. Основным технологическим решением, применяемым для ИМС, предназначенных для эксплуатации в космической среде является переход от объёмного кремния к структурам «кремний на изоляторе» (КНИ, SOI). В КНИ-приборе активная область транзистора отделена от подложки слоем скрытого оксида (Buried Oxide, BOX), что исключает образование паразитного n-p-n-p-тиристора между смежными транзисторами и тем самым полностью устраняет SEL [9, 30]. Дополнительно, за счёт уменьшения чувствительного объёма транзистора (фактически только тонкого приповерхностного кремниевого слоя), собранный одиночной частицей заряд оказывается меньше критического в большем диапазоне условий, что снижает SEU-чувствительность.

Развитием концепции КНИ являются полностью обеднённые КНИ-приборы (Fully Depleted SOI, FDSOI), в которых тонкая (несколько нанометров) кремниевая плёнка обедняется при нормальной работе. FDSOI-технологии 28 нм и 22 нм демонстрируют сочетание высокой удельной производительности и пониженной чувствительности к одиночным эффектам, что делает их перспективной платформой для радиационно-стойких ИМС следующего поколения [27]. Аналогичные перспективы характерны для FinFET-структур: трёхзатворная архитектура снижает чувствительный объём и собранный заряд при одиночных эффектах, однако накопление заряда в боковой изоляции остаётся актуальной проблемой [26, 27].

«Кремний на сапфире» (КНС, SOS) представляет собой предшественника технологии «кремний на изоляторе» с диэлектрической подложкой из монокристаллического сапфира. КНС обеспечивает полное устранение SEL и высокую ЭСД-стойкость, однако ограничена в плотности интеграции и применяется главным образом в специализированных радиационно-стойких СБИС [30].

Специализированные радиационно-стойкие техпроцессы (radiation-hardened foundries) предусматривают модификации стандартных процессов: оптимизацию профилей легирования карманов для подавления паразитных биполярных транзисторов, использование изоляции типа LOCOS вместо STI в наиболее ответственных областях, специальные пассивирующие слои с пониженной плотностью ловушек. Радиационная стойкость серийных микросхем, изготовленных по таким техпроцессам, может достигать нескольких Мрад по TID при гарантированном отсутствии SEL [30].

Перспективным направлением для силовой и СВЧ-электроники космических аппаратов является переход к широкозонным полупроводникам - карбиду кремния (SiC) и нитриду галлия (GaN). Высокая пороговая энергия дефектообразования (около 35 эВ для GaN против 21 эВ для Si) обуславливает повышенную DD-стойкость этих материалов, а большее напряжение пробоя позволяет реализовать силовые приборы с меньшими потерями. Однако пороги SEB и SEGR для коммерческих GaN- и SiC-приборов, как показывают

современные исследования, оказываются ниже теоретически ожидаемых, и квалификация таких приборов для космических применений остаётся актуальной задачей [20, 34].

6.2 Конструктивные методы

Конструктивные (топологические) методы повышения радиационной стойкости (Radiation Hardening By Design) заключаются в специальной разработке топологии стандартных коммерческих технологических процессов, обеспечивающей подавление либо ЭСД-, либо SEE-эффектов без изменения самого процесса. Это позволяет использовать дешёвые коммерческие ИМС для производства радиационно-стойких ИМС.

Наиболее известным конструктивным решением является использование транзисторов с кольцевой топологией (Enclosed Layout Transistor, ELT; кольцевые или замкнутые транзисторы). В такой топологии исток и сток разделены не прямолинейной, а замкнутой структурой затвора, охватывающего одну из областей; при этом периметр транзистора с боковой изоляцией STI исключается, и образующиеся под действием TID паразитные каналы утечки в STI оказываются «обойдѐнными» каналом транзистора. Программная работа [28] продемонстрировала применение ELT и охранных колец в коммерческом 0,5 мкм КМОП-процессе для изготовления считывающей микросхемы пиксельного детектора; деградация характеристик микросхемы начиналась только после суммарной дозы 600 крад – 1,7 Мрад в зависимости от типа излучения. Этот результат стал фактическим прецедентом RHBD для последующих чипов в технологиях 0,25 мкм и меньше, применяемых в детекторах ЛНС.

К конструктивным методам также относится использование охранных колец (guard rings) - областей сильно легированного полупроводника, окружающих чувствительные транзисторы и собирающих ионизационно-индуцированный заряд до того, как он попадѐт в чувствительный узел. Охранные кольца эффективно подавляют SEL в КМОП ИМС с объёмной подложкой, поскольку прерывают цепь паразитного тиристора, и снижают распространение заряда к смежным транзисторам, уменьшая MCU/MBU [29]. Дополнительно применяются увеличенные расстояния между чувствительными ячейками, перемежение разрядов одного слова в физическом массиве памяти (для перевода MCU в исправляемые EDAC одиночные сбои), резервирование транзисторов в критических ячейках.

6.3 Схемотехнические методы

Схемотехнические методы направлены на парирование радиационных эффектов на уровне электрической схемы, без изменения топологии транзисторов. Базовым приёмом против SEU является введение избыточности.

Тройное модульное резервирование (Triple Modular Redundancy, TMR) предусматривает реализацию каждой критической функции в виде трёх параллельно работающих модулей, выходы которых поступают на элемент, формирующий результат по принципу «2 из 3». При одиночном сбое в любом из

трёх каналов управляющее устройство автоматически замаскирует ошибку, и система продолжит правильную работу. TMR широко применяется в радиационно-стойких FPGA на базе SRAM (например, в семействе Xilinx Virtex для космоса), где совместно с периодической перезаписью конфигурационной памяти образует основной механизм борьбы с SEU [32]. Особенностью TMR в SRAM-FPGA является необходимость учитывать сбои в самой конфигурационной памяти и в принимающих решения элементах.

Альтернативным схемотехническим решением для борьбы с SEU в защёлках и регистрах являются ячейки DICE (Dual Interlocked Storage Cell), предложенные в работе [31]. В ячейке DICE состояние хранится в четырёх взаимосвязанных узлах, причём изменение одного из них не приводит к изменению логического выхода, так как для переключения ячейки необходимо одновременно изменить как минимум два диагональных узла. Такая структура обеспечивает иммунитет к одиночному сбою при увеличении площади ячейки и потребляемой мощности примерно вдвое. Развитием DICE стали более сложные ячейки (изолированная DICE, ячейки с увеличенными расстояниями между чувствительными узлами для парирования MCU).

Для борьбы с SET в комбинационной логике применяют временное резервирование, а именно выборка сигнала в моменты времени, отстоящие на величину, превышающую длительность SET-импульса, с последующим применением мажоритарной логики. В аналоговых ИМС снижение SET-восприимчивости достигается фильтрацией высокочастотных составляющих сигнала и использованием дифференциальных систем, в которых одиночный импульс смещения проявляется как синфазная помеха и подавляется. На уровне платы применяются аппаратные коды коррекции ошибок (Код Хэмминга и др.) для защиты содержимого ОЗУ и интерфейсов, а также сторожевые таймеры для парирования SEFI [11].

6.4 Программные методы

Программные методы защиты дополняют аппаратные средства и реализуются на уровне прошивки или системного программного обеспечения, не требуя модификации элементной базы. Наиболее распространённой является программная реализация кодов обнаружения и коррекции ошибок (Software-Implemented EDAC, SI-EDAC), при которой содержимое ОЗУ хранится с избыточностью, обеспечивающей исправление одиночных и обнаружение многократных битовых ошибок. Эффективность SI-EDAC применительно к SEU подробно проанализирована в работе [33]; показано, что комбинация программного кода Хэмминга и периодической перезаписи памяти позволяет на порядок снизить вероятность катастрофического отказа при дополнительных затратах процессорного времени порядка единиц процентов.

Близким приёмом является фоновое периодическое считывание и перезапись содержимого памяти (в первую очередь конфигурационной памяти SRAM-FPGA и оперативной памяти бортовых вычислителей), исправляющее

одинокые сбои до того, как они смогут накопиться и перейти в неисправляемые многократные. Период перезаписи выбирается с учётом ожидаемой частоты SEU на конкретной орбите и характеристик кода ошибок [32].

К программным методам также относятся: программное TMR (последовательное выполнение задачи три раза с мажоритарной логикой), регулярные точки сохранения состояния и откат при обнаружении сбоя, контроль выполнения программы по контрольным точкам, программные сторожевые таймеры. Дополнительно при проектировании бортового ПО применяется защитное программирование: введение явных проверок диапазонов значений переменных, ограничение допустимых переходов в конечных состояниях, дублирование критических переменных. Программные методы особенно ценны для микро- и наноспутников, использующих коммерческие компоненты с заведомо ограниченной радиационной стойкостью, и для перепрограммирования аппаратуры после запуска [35, 36].

6.5 Экранирование

Экранирование — физическое уменьшение потока ионизирующего излучения, доходящего до полупроводниковых приборов, посредством размещения вокруг них слоя поглощающего материала. Эффективность экранирования сильно различается для разных компонент космического излучения. Электроны радиационных поясов и низкоэнергетичные протоны эффективно поглощаются сравнительно тонкими (единицы миллиметров) слоями алюминия; для высокоэнергетичных протонов внутреннего пояса и солнечных энергичных частиц требуются существенно большие толщины, причём возникает дополнительный поток вторичного тормозного излучения, генерируемого тормозящимися электронами. Защита от тяжёлых ионов галактических космических лучей с помощью пассивных экранов практически нецелесообразна: пробег ионов высокой энергии в любом материале превышает приемлемые толщины и массы, а вторичные частицы фрагментации могут даже увеличить радиационный риск [3, 4].

При проектировании реальных космических аппаратов экранирование обычно реализуется как комбинация общей экранирующей оболочки корпуса (алюминиевая обшивка толщиной 1–3 мм) и локального экранирования наиболее чувствительных компонент (spot shielding) — установки на отдельные микросхемы дополнительных пластин или колпачков из плотных материалов (тантал, вольфрам, специальные сплавы). Локальное экранирование позволяет существенно снизить накопленную дозу на критическом компоненте при незначительных дополнительных затратах массы. Для современных малых спутников исследуются новые материалы экранирования, в том числе многослойные композиты «полиэтилен — металл», использующие водородсодержащие компоненты для эффективного замедления протонов и нейтронов.

6.6 Перспективные методы и системный подход

Помимо упомянутых в предыдущих разделах перспективных технологий (FDSOI, FinFET, GaN, SiC), активно развиваются следующие направления.

Во-первых, всё более широкое применение коммерческих компонентов (Commercial Off-The-Shelf, COTS) в малых космических аппаратах с компенсацией их ограниченной радиационной стойкости на схемотехническом и программном уровнях. Систематические исследования TID-стойкости типовых COTS-компонентов для CubeSat показали значительный разброс параметров: например, промышленные SD-карты Delkin сохраняли работоспособность до суммарной дозы около 24 крад, тогда как карты SanDisk отказывали при дозе порядка 8 крад [35]. Такие данные служат базой для отбора компонентов и проектирования резервированных архитектур в условиях «Нового космоса» [36].

Во-вторых, развитие методик ускоренных испытаний и отбраковки COTS-партий, включая лазерные методы, тестирование на ускорителях частиц и продвинутую статистическую обработку результатов [23, 24].

В-третьих, стандартизация и расширение применимости моделей радиационной обстановки позволяет с лучшей точностью прогнозировать накопленную дозу и частоту ОРЭ для конкретной миссии.

В-четвёртых, развитие приборно-технологического моделирования (TCAD) и Monte Carlo-моделирования с учётом квантовых эффектов в нанометровых технологиях и более точного описания трекового образования заряда тяжёлыми ионами [22].

Существенно отметить, что эффективная радиационная стойкость космического прибора достигается не отдельным методом, а согласованной системой технических решений на всех уровнях. При планировании полета выбираются орбита и срок эксплуатации, определяющие требования к стойкости. На уровне аппаратуры — выбираются технология (промышленная COTS, RHBD, специализированная радиационно-стойкая), степень аппаратного резервирования и стратегия программной защиты. На уровне испытаний — формируется программа испытаний, охватывающая ЭСД, DD и весь спектр ОРЭ. Такой системный подход является основой современного проектирования космической электроники [9, 11, 12].

Заклучение

Радиационная стойкость интегральных микросхем является ключевым требованием к электронной компонентной базе космической техники. В обзоре систематизированы три основных класса радиационных эффектов, определяющих радиационный отклик ИМС в космосе: эффект суммарной дозы, обусловленный накоплением заряда в диэлектрических слоях и поверхностных состояний на границах раздела; эффект смещения атомов, связанный с образованием дефектов кристаллической решётки; одиночные радиационные эффекты, вызываемые отдельными высокоэнергетическими частицами и проявляющиеся как обратимые сбои либо как катастрофические отказы.

Современный инструментарий повышения радиационной стойкости охватывает все уровни проектирования: технологический (КНИ, КНС, FDSOI, FinFET, широкозонные полупроводники), конструктивный (RHBD, кольцевые транзисторы, охранные кольца), схемотехнический (DICE, TMR, ECC), программный (SI-EDAC, TMR) и физический (экранирование). В реальных проектах применяется их согласованная комбинация, выбираемая исходя из требований миссии и доступной элементной базы.

Развитие концепции «Нового космоса» ставит перед сообществом разработчиков новые задачи: испытание и применение коммерческих компонентов в условиях короткого срока эксплуатации, разработку упрощённых процедур радиационных испытаний для малых спутников, развитие открытых баз данных по радиационной стойкости COTS-компонентов. Параллельно прогресс в нанометровых технологиях, широкозонных материалах и реконфигурируемой электронике обеспечивает технологический фундамент для космических полетов следующего поколения — от малых спутников связи и наблюдения Земли до научных аппаратов дальнего космоса.

Ожидаемые направления дальнейших исследований включают: систематизацию данных по радиационной стойкости новых техпроцессов (5 нм и ниже, FinFET, GAAFET); разработку методик испытаний многокристалльных и трехмерных ИМС; исследование радиационной стойкости специализированных ИМС с применением искусственного интеллекта на основе нейроморфных архитектур; уточнение моделей радиационной обстановки в условиях нестандартных орбит (окололунные, межпланетные миссии). Накопленный за десятилетия фундамент знаний в сочетании с современными методами проектирования и испытаний обеспечивает реалистичную основу для решения этих задач.

Список литературы:

1. Пермяков Р. В. «Новый космос»: глобальный ландшафт и модели коммерциализации // Экономика космоса. — 2023. — № 6. — URL: <https://cyberleninka.ru/article/n/novyy-kosmos-globalnyu-landshaft-i-modeli-kommertsializatsii> (дата обращения: 13.05.2026).
2. Сокуров В. Ф. Физика космических лучей: космическая радиация. — Ростов н/Д.: Феникс, 2005. — 188 с. — (Высшее образование).
3. Rahman M. M., Shankar D., Santra S. Analysis of Radiation Environment and its Effect on Spacecraft in Different Orbits // 68th International Astronautical Congress (IAC 2017). — Adelaide, Australia, 2017.
4. Соловьёв Н. О., Клименков К. О. Влияние радиации на электронику // Сборник материалов конференции. — М., 2018.
5. Pease R. L. Total Ionizing Dose Effects in Bipolar Devices and Circuits // IEEE Transactions on Nuclear Science. — 2003. — Vol. 50, No. 3. — P. 539–551. — DOI: 10.1109/TNS.2003.813133.
6. Oldham T. R., McLean F. B. Total Ionizing Dose Effects in MOS Oxides and Devices // IEEE Transactions on Nuclear Science. — 2003. — Vol. 50, No. 3. — P. 483–499. — DOI: 10.1109/TNS.2003.812927.

7. Андреев В. В., Жалнин В. П., Столяров А. А. Физические основы микро- и наноэлектроники. Библиотека «Приборостроение». Том 3. — М.: МГТУ им. Н. Э. Баумана, 2018.
8. Чумаков А. И., Бобровский Д. В., Соловьёв С. А. Влияние импульсного характера излучения на параметры чувствительности интегральных схем к одиночным радиационным эффектам // Микроэлектроника. — 2017.
9. Чумаков А. И. Радиационные эффекты в интегральных схемах. — М.: Техносфера, 2024. — 384 с. — (Серия «Мир электроники»). — ISBN 978-5-94836-707-1.
10. Никифоров А. Ю., Телец В. А., Чумаков А. И. Радиационные эффекты в КМОП ИС. — М.: Радио и связь, 1994. — 164 с.
11. Радиационная стойкость изделий ЭКБ: научное издание / под ред. А. И. Чумакова. — М.: НИЯУ МИФИ, 2010. — 148 с.
12. Schwank J. R., Shaneyfelt M. R., Dodd P. E. Radiation Hardness Assurance Testing of Microelectronic Devices and Integrated Circuits: Radiation Environments, Physical Mechanisms, and Foundations for Hardness Assurance // IEEE Transactions on Nuclear Science. — 2013. — Vol. 60, No. 3. — P. 2074–2100. — DOI: 10.1109/TNS.2013.2254722.
13. Schwank J. R., Shaneyfelt M. R., Fleetwood D. M., Felix J. A., Dodd P. E., Paillet P., Ferlet-Cavrois V. Radiation Effects in MOS Oxides // IEEE Transactions on Nuclear Science. — 2008. — Vol. 55, No. 4. — P. 1833–1853. — DOI: 10.1109/TNS.2008.2001040.
14. Barnaby H. J. Total-Ionizing-Dose Effects in Modern CMOS Technologies // IEEE Transactions on Nuclear Science. — 2006. — Vol. 53, No. 6. — P. 3103–3121. — DOI: 10.1109/TNS.2006.885952.
15. Pease R. L., Schrimpf R. D., Fleetwood D. M. ELDRS in Bipolar Linear Circuits: A Review // IEEE Transactions on Nuclear Science. — 2009. — Vol. 56, No. 4. — P. 1894–1908. — DOI: 10.1109/TNS.2008.2011485.
16. Першенков В. С., Бакеренков А. С., Фелицын В. А., Родин А. С. Обобщённая модель эффекта низкой интенсивности в биполярных приборах // Труды НИИСИ РАН. — 2018. — Т. 8, № 3. — С. 94–98.
17. Зебрев Г. И. Радиационные эффекты в кремниевых интегральных схемах высокой степени интеграции. — М.: НИЯУ МИФИ, 2010. — 148 с.
18. Srour J. R., Marshall C. J., Marshall P. W. Review of Displacement Damage Effects in Silicon Devices // IEEE Transactions on Nuclear Science. — 2003. — Vol. 50, No. 3. — P. 653–670. — DOI: 10.1109/TNS.2003.813197.
19. Srour J. R., Palko J. W. Displacement Damage Effects in Irradiated Semiconductor Devices // IEEE Transactions on Nuclear Science. — 2013. — Vol. 60, No. 3. — P. 1740–1766. — DOI: 10.1109/TNS.2013.2261316.
20. Pearton S. J., Aitkaliyeva A., Xian M., Ren F., Khachatryan A., Ildefonso A., Islam Z., Rasel M. A. J., Haque A., Polyakov A. Y., Kim J. Review — Radiation Damage in Wide and Ultra-Wide Bandgap Semiconductors // Journal of Vacuum Science & Technology B. — 2023. — Vol. 41, No. 3. — P. 030802. — DOI: 10.1116/6.0002628.
21. Dodd P. E., Massengill L. W. Basic Mechanisms and Modeling of Single-Event Upset in Digital Microelectronics // IEEE Transactions on Nuclear Science. — 2003. — Vol. 50, No. 3. — P. 583–602. — DOI: 10.1109/TNS.2003.813129.
22. Dodd P. E., Shaneyfelt M. R., Schwank J. R., Felix J. A. Current and Future Challenges in Radiation Effects on CMOS Electronics // IEEE Transactions on Nuclear Science. — 2010. — Vol. 57, No. 4. — P. 1747–1763. — DOI: 10.1109/TNS.2010.2042613.
23. Чумаков А. И. Лазерная методика оценки параметров чувствительности БИС к эффектам воздействия отдельных заряженных частиц // Микроэлектроника. — 2018. — Т. 47, № 3. — С. 198–204.

24. Согоян А. В., Смолин А. А., Чумаков А. И. Оценка соответствия интегральных схем требованиям по стойкости к воздействию тяжёлых заряженных частиц // *Безопасность информационных технологий*. — 2020. — Т. 27, № 1. — С. 68–82. — DOI: 10.26583/bit.2020.1.06.
25. Чумаков А. И. Оценка чувствительности интегральных схем к одиночным радиационным эффектам для точечной области собирания заряда // *Микроэлектроника*. — 2015. — Т. 44, № 1. — С. 34–40.
26. Gaillardin M., Paillet P., Ferlet-Cavrois V., Faynot O., Jahan C., Cristoloveanu S. Total Ionizing Dose Effects on Triple-Gate FETs // *IEEE Transactions on Nuclear Science*. — 2006. — Vol. 53, No. 6. — P. 3158–3165. — DOI: 10.1109/TNS.2006.886219.
27. Fleetwood D. M. Evolution of Total Ionizing Dose Effects in MOS Devices With Moore’s Law Scaling // *IEEE Transactions on Nuclear Science*. — 2018. — Vol. 65, No. 8. — P. 1465–1481. — DOI: 10.1109/TNS.2017.2786140.
28. Snoeys W., Faccio F., Burns M., Campbell M. et al. Layout Techniques to Enhance the Radiation Tolerance of Standard CMOS Technologies Demonstrated on a Pixel Detector Readout Chip // *Nuclear Instruments and Methods in Physics Research A*. — 2000. — Vol. 439. — P. 349–360. — DOI: 10.1016/S0168-9002(99)00899-2.
29. Мироненко Л., Юдинцев В. Повышение радиационной стойкости интегральных схем. Конструктивные методы на базе промышленной технологии // *Электроника: Наука, Технология, Бизнес*. — 2008. — № 8. — С. 74–79.
30. Юдинцев В. Радиационно-стойкие интегральные схемы. Надёжность в космосе и на земле // *Электроника: Наука, Технология, Бизнес*. — 2007. — № 5. — С. 72–77.
31. Calin T., Nicolaidis M., Velazco R. Upset Hardened Memory Design for Submicron CMOS Technology // *IEEE Transactions on Nuclear Science*. — 1996. — Vol. 43, No. 6. — P. 2874–2878. — DOI: 10.1109/23.556880.
32. Kastensmidt F. L., Sterpone L., Carro L., Sonza Reorda M. On the Optimal Design of Triple Modular Redundancy Logic for SRAM-based FPGAs // *Proceedings of Design, Automation and Test in Europe (DATE)*. — 2005. — P. 1290–1295. — DOI: 10.1109/DATE.2005.34.
33. Shirvani P. P., Saxena N., McCluskey E. J. Software-Implemented EDAC Protection Against SEUs // *IEEE Transactions on Reliability*. — 2000. — Vol. 49, No. 3. — P. 273–284. — DOI: 10.1109/24.914544.
34. Lauenstein J.-M. Wide Bandgap Power — SiC, GaN — Radiation Reliability. — Presentation at IEEE Nuclear and Space Radiation Effects Conference (NSREC) Short Course. — NASA NEPP Program, 2020. — NTRS document 20205009739.
35. Kingsbury R., Schmidt F., Cahoy K., Sklair D., Blackwell W., Osaretin I., Legge R. TID Tolerance of Popular CubeSat Components // *Proceedings of the IEEE Radiation Effects Data Workshop (REDW)*. — 2013. — P. 1–4. — DOI: 10.1109/REDW.2013.6658220.
36. Sinclair D., Dyer J. Radiation Effects and COTS Parts in SmallSats // *Proceedings of the AIAA/USU Conference on Small Satellites*. — 2013. — SSC13-IV-3.
37. Ginet G. P., O’Brien T. P., Huston S. L., Johnston W. R. et al. AE9, AP9 and SPM: New Models for Specifying the Trapped Energetic Particle and Space Plasma Environment // *Space Science Reviews*. — 2013. — Vol. 179. — P. 579–615. — DOI: 10.1007/s11214-013-9964-y.
38. Чумаков А. И., Ужegov В. М., Ахметов А. О., Бойченко Д. В., Яненко А. В., Рясной Н. В. Оценка показателей стойкости интегральных схем при воздействии тяжёлых заряженных частиц с использованием различных моделей // *Безопасность информационных технологий*. — 2017. — Т. 24, № 1. — С. 73–84. — DOI: 10.26583/bit.2017.1.09.

УДК 658.5:519.246:519.237

Титов Е. А.

Уфимский государственный нефтяной технический университет, Уфа, Россия

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И РЕДУКЦИЯ РАЗМЕРНОСТИ ТЕХНИКО-ЭКОНОМИЧЕСКИХ ИНДИКАТОРОВ ПРОИЗВОДСТВА КОРПУСНЫХ ИЗДЕЛИЙ

Аннотация. Представлены результаты количественной обработки помесечных наблюдений за работой отечественного завода, выпускающего электротехнические оболочки и шкафы для систем промышленной автоматизации. Выборка охватывает тринадцать показателей за сорок восемь календарных месяцев. По объёму выпуска проведена аддитивная сезонно-трендовая декомпозиция STL: на сезонную составляющую приходится около 75,8 % дисперсии, на трендовую — порядка 18,6 %. Кросс-корреляционный анализ выявил запаздывающие связи между внешними возмущениями и индикаторами производства, в том числе зависимость себестоимости от стоимости листового проката с лагом два месяца. Метод главных компонент свёл тринадцатимерное признаковое пространство к четырём компонентам, объясняющим суммарно 86,1 % дисперсии.

Ключевые слова: временные ряды, STL-декомпозиция, кросс-корреляция, метод главных компонент, индикаторы производства, корпусные изделия, промышленная автоматизация.

Введение

Российский рынок корпусных изделий для аппаратуры промышленной автоматизации последние годы растёт — и одновременно колеблется. Цены на металлопрокат меняются непредсказуемо, логистические схемы перестраиваются, часть импортных комплектующих приходится замещать. Спрос на распределительные шкафы и электротехнические оболочки в целом продолжает увеличиваться, однако нормативные значения и ретроспективные отчёты, на которые традиционно опираются плановые службы, такой темп изменений уже не догоняют [1]. Отсюда — устойчивый запрос на инструменты, которые умеют доставать из накопленной производственной статистики неочевидные зависимости и переводить их в форму, пригодную для прогнозирования [2].

Цель публикации — изложить результаты статистической обработки помесечных показателей одного из российских заводов, выпускающих корпусную электротехническую продукцию. Задач, по сути, три. Разложить временной ряд выпуска на тренд, сезонность и шум. Найти запаздывающие связи между внешними возмущениями и производственными индикаторами. И, наконец, ужать тринадцатимерное признаковое пространство до экономной размерности, не потеряв при этом значимой доли дисперсии.

Методически работа опирается на классические подходы прикладной статистики и анализа временных рядов [2; 3]. Объём выборки — 48 точек по 13 показателям — для перечисленных методов уверенно достаточен: оценки получаются и статистически значимыми, и пригодными для содержательной интерпретации.

1. Объект исследования и исходные данные

Исходные сведения предоставлены заводом, который выпускает электротехнические корпуса: распределительные шкафы, настенные боксы, пульта управления, серверные и климатические стойки. Учёт по 13 параметрам ведётся ежемесячно, охватываемый отрезок — с января 2021 г. по декабрь 2024 г. (48 наблюдений; всего 624 значения). Описательная статистика основных переменных собрана в таблице 1.

Уже первичное описание в таблице 1 даёт несколько примечательных деталей. Цена листа металла за наблюдаемый период подросла приблизительно с 55 до 79 тыс. руб./т — устойчивое и неприятное давление на затратную часть. Средний месячный выпуск (763 шт.) почти в точности совпадает со средним портфелем заказов (763 шт.); расхождение — на уровне случайной составляющей. По сути, завод работает «с колёс». Относительный разброс шире всего у загрузки оборудования (СКО $\approx 9,4\%$ при среднем 73,2%) — прямой след выраженной сезонности спроса на корпусную продукцию [6].

Таблица 1. Сводная описательная статистика производственных индикаторов (48 мес.)

Показатель	Мин.	Макс.	Среднее	СКО	Ед.
Выпуск	502	1019	763	151	шт./мес.
Заказы	500	1024	763	150	шт./мес.
Цена металла	55 234	87 881	75 161	9 586	руб./т
Себестоимость	26 055	36 645	32 399	3 474	руб./шт.
ОЕЕ	68,9	75,0	72,0	1,5	%
Загрузка оборудования	56,7	89,7	73,2	9,4	%
Уровень брака	3,9	6,3	4,7	0,6	%
ОТІF	92,0	97,1	94,5	1,4	%
Рентабельность	27,3	66,1	41,8	14,2	%
Запасы	12,0	36,2	24,5	6,0	дни

2. Сезонно-трендовая декомпозиция объёма выпуска

Для разложения ряда выпуска $Y(t)$ на трендовую $T(t)$, сезонную $S(t)$ и остаточную $\varepsilon(t)$ составляющие выбрана аддитивная процедура STL (Seasonal-

Trend decomposition using LOESS) [3]. Сезонный период — 12 месяцев, режим оценивания — устойчивый (robust). Распределение дисперсии между компонентами показано в таблице 2.

Картина из таблицы 2 однозначная — главный двигатель поведения выпуска именно сезонная составляющая, её амплитуда около $\pm 24\%$ от среднегодового уровня. Максимум сезонного индекса приходится на апрель (+182 шт., +24% к среднему), минимум — на январь (−164 шт., −21%). Цифры хорошо согласуются с тем, что говорят на самом производстве: пики заказов — весна (март–май) и осень (сентябрь–ноябрь), провалы — зима и лето. Корпуса для шкафов автоматизации идут вслед за строительной активностью, и месячные данные это подтверждают.

Таблица 2. Распределение дисперсии между составляющими временного ряда объёма выпуска

Компонента	Среднее	Доля дисперсии, %	Содержательное толкование
Тренд $T(t)$	766 шт./мес.	18,6	Умеренный рост сегмента
Сезонность $S(t)$	± 182 шт./мес.	75,8	Циклы строительной активности
Остаток $\varepsilon(t)$	0 ($\sigma \approx 30$)	4,1	Случайные сбои, шум спроса

Расширенный тест Дики–Фуллера однозначно отверг гипотезу о стационарности исходного ряда $Y(t)$; после однократного дифференцирования стационарность достигается — порядок интегрирования $d = 1$ [4]. На практике это значит, что в SARIMA-моделях для выпуска параметр d следует фиксировать единицей - этим результатом воспользуемся при сборке гибридной прогнозной системы.

3. Кросс-корреляционный анализ запаздывающих связей

Кросс-корреляционная функция $R_{xy}(\tau)$ — естественный инструмент для поиска лага τ , на котором связь фактора с индикатором максимальна [2]. Для каждой пары рассчитаны значения $R_{xy}(\tau)$ на интервале $\tau \in [0; 6]$ мес.; значимыми считались связи с $|R| > 0,45$ при $p < 0,05$. Систематизированные результаты — в таблице 3.

Наибольший интерес для прогнозного моделирования представляет связка «цена металла — себестоимость». Коэффициент $R_{xy}(2) = +0,99$ при лаге в два месяца — это уже не совпадение, а долгосрочная равновесная (коинтегрированная) зависимость [4]. С практической стороны вывод прост: цена листа, которую видим сегодня, через два месяца почти гарантированно отзовется на себестоимости. Логично поэтому включить ценовой ряд в модель SARIMAX как экзогенный регрессор [4; 7].

Таблица 3. Значимые запаздывающие связи между факторами и индикаторами производства

Фактор X(t)	Индикатор Y(t+τ)	Лаг мес. τ,	R _{xy} (τ)	Содержание связи
Цена металла	Себестоимость	2	+0,99	Сильный лагированный эффект
Цена металла	Рентабельность	2	-0,87	Сжатие маржи
Заказы	Выпуск	0	+0,99	Синхронная реакция
Заказы	Загрузка	0	+0,97	Прямой отклик мощностей
Загрузка	ОТIF	0	-0,93	Срыв сроков при перегрузке
Загрузка	Брак	0	+0,84	Дефицит контроля на пиках
Загрузка	Запасы	0	-0,85	Расход запасов на пиках спроса
Выпуск	ОТIF	0	-0,89	Снижение ритмичности

А вот связи группы «загрузка → производственные показатели» работают почти мгновенно ($\tau = 0$). Объяснение лежит на поверхности: стоит загрузке оборудования перейти отметку порядка 85 %, и брак растёт, а ОТIF проседает буквально в тот же месяц. Ценовые сигналы доходят дольше — через цепочку «закупка — склад — потребление в цеху», что само по себе занимает время [6].

4. Снижение размерности методом главных компонент

Чтобы избавиться от мультиколлинеарности переменных и получить компактное пространство признаков, применён метод главных компонент (РСА) [5]. Сначала данные нормированы — центрирование и приведение к единичной дисперсии. Затем сделано собственное разложение ковариационной матрицы. Результаты по компонентам собраны в таблице 4.

Формально критерий Кайзера ($\lambda > 1$) оставил бы три главные компоненты [5]. Но четвёртую мы всё-таки включили: её собственное число $\lambda \approx 0,97$ буквально упирается в порог, а высокие нагрузки приходятся на ОЕЕ и затраты на ТОиР — именно те переменные, через которые потом будет работать управление техническим состоянием.

Таблица 4. Структура главных компонент 13-мерного признакового пространства

Компонента	Собств. число λ	Доля диспер., %	Накопл., %	Содержательная интерпретация
PC1	5,99	46,1	46,1	Рыночная нагрузка (выпуск, заказы, загрузка, брак)
PC2	2,94	22,6	68,7	Ценовое давление (цена металла, себестоимость, рентабельность)
PC3	1,29	9,9	78,6	Логистический контур (ОТИФ, запасы)
PC4	0,97	7,5	86,1	Техническое состояние (ОЕЕ, затраты на ТОиР)

Четыре компоненты, объясняющие 86,1 % суммарной дисперсии, читаются содержательно почти без натяжек. Первая — совместное движение объёмов и нагрузочных характеристик; можно называть её «рыночной». Вторая — связка «цена сырья — себестоимость — рентабельность», то есть канал внешнего ценового давления. Третья группирует логистические показатели, четвёртая — параметры технического состояния оборудования. Совпадение с экспертным членением индикаторов промышленной деятельности на пять групп [6] здесь не случайное; оно лишний раз подтверждает, что состав исходных переменных подобран верно.

Заключение

Подведем итог. Статистическая обработка помесечных показателей за четыре года дала несколько прикладных ориентиров для дальнейшей работы с производством корпусных изделий промышленной автоматизации.

Во-первых, выпуск подчиняется отчётливому годовому циклу с пиками в апреле и октябре. На сезонную составляющую приходится 75,8 % общей дисперсии ряда, на тренд — умеренные 18,6 %. Это однозначно склоняет выбор в пользу сезонных моделей прогноза.

Во-вторых, кросс-корреляционный анализ зафиксировал коинтеграционную связку «цена металла → себестоимость» с лагом два месяца ($R = +0,99$). Отдельно выделена группа мгновенных откликов — индикаторы качества и логистики реагируют на превышение оптимальной загрузки в тот же месяц.

В-третьих, четыре главные компоненты по матрице тринадцати нормированных переменных объясняют 86,1 % суммарной дисперсии и читаются ровно по тем группам, в которые индикаторы попадают и содержательно: рыночные, ценовые, логистические, технические.

Полученные численные характеристики удобно использовать как готовый информационный задел для гибридных прогнозных моделей и систем

поддержки принятия решений, рассчитанных на работу при подвижной внешней среде [7].

Список литературы:

1. Зайцев, Н. Л. Экономика, организация и управление предприятием : учебное пособие для вузов по специальности "Менеджмент организации" / Н. Л. Зайцев, Гос. ун-т управления (ГУУ). – 2-е изд., доп. – М. : ИНФРА-М, 2012. – 455 с. – (Высшее образование). – ISBN 978-5-16-002841-5.
2. Орлов, А. И. Прикладная статистика : краткий курс / А. И. Орлов. - Москва : ИНТУИТ, 2016. - 647 с. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2156508> (дата обращения: 21.04.2026). – Режим доступа: по подписке.
3. Cleveland R. B., Cleveland W. S., McRae J. E., Terpenning I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess // Journal of Official Statistics. 1990. Vol. 6. P. 3–73.
4. Box G. E. P., Jenkins G. M., Reinsel G. C. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken : Wiley, 2015. 712 p.
5. Jolliffe I. T. Principal Component Analysis. 2nd ed. New York : Springer, 2002. 487 p.
6. ГОСТ Р ИСО 22400-2-2019. Ключевые показатели эффективности для управления производственными операциями. М. : Стандартиформ, 2019. 36 с.
7. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. OTexts, 2021. [Электронный ресурс]. URL: <https://otexts.com/fpp3/> (дата обращения: 15.03.2025).
8. Айвазян, С. А. Методы эконометрики : учебник / С. А. Айвазян. — Москва : Магистр : ИНФРА-М, 2022. — 512 с. - ISBN 978-5-9776-0153-5. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1840468> (дата обращения: 21.04.2026). – Режим доступа: по подписке.

УДК 004.89:81'322

Лысенко А. Ф., Титаренко Д. С.

Донской государственный технический университет, Ростов-на-Дону, Россия

АВТОМАТИЗАЦИЯ ПРОЦЕССА ИДЕНТИФИКАЦИИ И ПЕРЕВОДА ТЕРМИНОВ С КЛАССИФИКАЦИЕЙ ТЕМАТИКИ ТЕКСТА

Аннотация: в работе исследуется возможность автоматизации идентификации терминов, их перевода и классификации тематики текста в многоязычных корпусах. Необходимость исследования связана с ростом объемов научной и технической документации, обработка которой вручную становится затруднительной. Объектом исследования являются многоязычные текстовые данные специализированных предметных областей, представленные в виде научных публикаций, технических описаний и нормативных документов. Методологическая база работы включает анализ классических статистических методов и современных нейросетевых подходов к автоматическому извлечению терминов, машинному переводу и классификации текстов. В качестве базовой архитектуры рассматривается многоязычная трансформерная модель XLM-RoBERTa, способная формировать единое семантическое пространство для различных языков. В статье описывается конвейерный подход к построению системы, объединяющей модули идентификации терминов, их перевода и определения тематики текста. Полученные результаты могут быть использованы в системах автоматического перевода и анализа текстов.

Ключевые слова: автоматическое извлечение терминов, машинный перевод, тематическая классификация, многоязычные модели, XLM-RoBERTa, обработка естественного языка.

Введение. В последние годы объем текстовых данных, особенно в научно-технической сфере, значительно увеличился. К таким данным относятся научные публикации, технические описания, патенты, эксплуатационная и нормативная документация, активно используемые в промышленности, науке и сфере высоких технологий. При этом значительная часть таких данных представлена на разных языках, что усложняет их обработку.

Ручная идентификация терминов, их перевод и определение тематики текстов требуют значительных временных затрат и высокой квалификации специалистов. При этом подобный подход не обеспечивает стабильного качества обработки и не гарантирует терминологическую согласованность, особенно при работе с большими массивами данных. Участие человека в данных процессах также повышает риск возникновения смысловых и терминологических ошибок.

Автоматизация задач идентификации и перевода терминов, а также классификации тематики текста рассматривается как один из ключевых факторов повышения эффективности обработки многоязычных корпусов [1]. Использование методов обработки естественного языка позволяет ускорить

анализ документов, обеспечить единый подход к интерпретации терминологии и повысить качество последующего поиска и анализа информации. Существенный прогресс в данной области в последние годы связан с развитием архитектур глубокого обучения, прежде всего трансформерных моделей, демонстрирующих высокую эффективность на широком спектре лингвистических задач [2].

Целью настоящей обзорной работы является анализ существующих подходов к автоматизации идентификации и перевода терминов с одновременной классификацией тематики текста, а также обоснование целесообразности применения многоязычной модели XLM-RoBERTa для решения данного комплекса задач.

Основная часть. Ранние методы автоматического извлечения терминов основывались преимущественно на статистическом анализе текстов и количественных характеристиках употребления слов. К таким методам относятся частотный анализ, TF-IDF, различные меры ассоциации и устойчивости словосочетаний. Эти подходы позволяют выявлять часто употребляемые лексические единицы и потенциальные терминологические кандидаты, однако они слабо учитывают контекст употребления терминов и не позволяют надёжно различать общезыковую лексику и специализированные понятия [3]. Кроме того, статистические методы демонстрируют низкую эффективность при обработке многокомпонентных терминов, аббревиатур и сложных терминологических конструкций.

Развитие методов машинного обучения привело к использованию моделей последовательной разметки, включая условные случайные поля, рекуррентные нейронные сети. Данные подходы позволили учитывать контекст на уровне предложения и ближайшего окружения слова, однако их эффективность во многом зависела от объёма и качества размеченных данных.

Кроме того, такие модели плохо масштабировались и испытывали сложности при переносе между языками и предметными областями.

Появление трансформерных моделей значительно улучшило качество обработки текстов. Они позволяют лучше учитывать контекст и взаимосвязи между словами. [4]. Благодаря этому трансформеры стали активно применяться для задач извлечения терминов, классификации текстов и машинного перевода.

Перевод терминов отличается от общего машинного перевода тем, что требует строгой точности и единообразия. В отличие от общего перевода, терминологический перевод требует строгой однозначности, сохранения смысловой целостности специализированных понятий. На практике для решения данной задачи применяются гибридные подходы, сочетающие автоматическое извлечение терминов, использование специализированных терминологических словарей и нейронные модели перевода, адаптированные под конкретные предметные области.

Тематическая классификация текстов также претерпела значительную эволюцию. Классические алгоритмы машинного обучения, такие как байесовский классификатор и метод опорных векторов, постепенно уступили место глубоким нейросетевым моделям, способным учитывать сложные семантические зависимости и контекст документа в целом. Использование предобученных языковых моделей позволяет эффективно решать задачу классификации даже при ограниченном объеме размеченных данных, что особенно актуально для специализированных корпусов [5].

Модель XLM-RoBERTa относится к семейству многоязычных трансформеров и обучается на задаче маскированного языкового моделирования с использованием масштабных корпусов данных более чем на ста языках. В результате предобучения создается общее семантическое пространство, в котором понятия на различных языках возможно сопоставить. Это делает модель особенно эффективной для кросс-лингвистических задач, включая автоматическое извлечение терминов, перевод терминологических единиц и классификацию тематики текста [6].

На рис.1 представлена схема обработки текста моделью XML-ROBERTa и результаты работы модели.

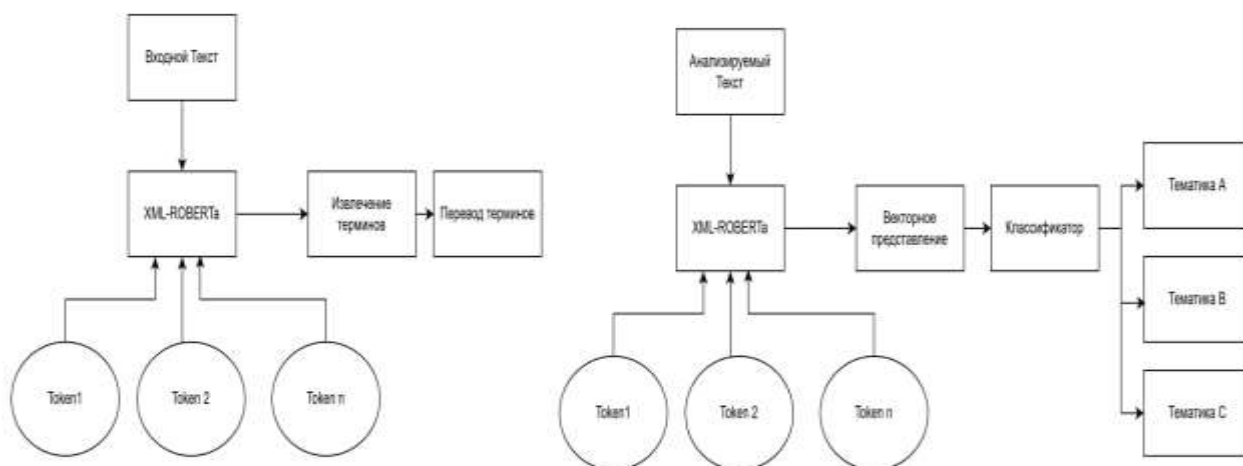


Рис.1 Архитектура перевода терминов и определения тематик модели XML-ROBERTa

Предлагаемая методика основана на архитектуре, включающей три взаимосвязанных модуля: модуль идентификации терминов, модуль перевода терминологических единиц и модуль классификации тематики текста. XLM-RoBERTa используется, в качестве базового энкодера в модуле идентификации терминов, при этом выходные слои и обучающие данные адаптируются под конкретные задачи. Такой подход обеспечивает согласованность обработки данных и позволяет интегрировать результаты отдельных этапов в рамках единой системы.

Использование ВЮ-структуры для извлечения терминов. Одним из ключевых подходов к автоматическому извлечению терминов является

формализация данной задачи как задачи последовательной разметки (sequence labeling). В рамках данного подхода каждому токену текста присваивается метка, отражающая его принадлежность к терминологической единице. Наиболее распространённой схемой разметки является формат BIO (Begin–Inside–Outside), в котором используются три типа тегов: начало термина (B), продолжение термина (I) и отсутствие принадлежности к термину (O).

BIO-разметка часто используется в задачах NER, поскольку позволяет выделять границы терминов в тексте. BIO-разметка обеспечивает удобное представление данных для обучения моделей машинного обучения, так как преобразует задачу извлечения терминов в задачу классификации на уровне токенов.

Исследования показывают, что такой подход хорошо работает вместе с нейросетевыми моделями. В работе по извлечению информации из документов NER традиционно рассматривается как задача предсказания BIO-тегов для последовательности токенов.

Интеграция BIO-разметки и XLM-RoBERTa. Комбинирование BIO-разметки с трансформерными моделями представляет собой современный стандарт в задачах автоматического извлечения терминов. В данной архитектуре входной текст токенизируется и подаётся в модель XLM-RoBERTa, которая генерирует контекстные векторные представления для каждого токена. Далее поверх этих представлений обучается классификатор, предсказывающий BIO-теги. Такой подход позволяет: учитывать контекст слова в предложении, корректно обрабатывать многословные термины, переносить знания между языками, работать с различными предметными областями.

Заключение. В статье рассмотрены основные подходы к автоматизации обработки терминов в многоязычных текстах. Традиционные методы имеют ограничения, особенно при работе с контекстом. Использование многоязычных трансформерных моделей позволяет существенно повысить качество обработки специализированных текстов и обеспечить перенос знаний между языками.

Модель XLM-RoBERTa является перспективной архитектурой для построения комплексных систем анализа текста, объединяющих задачи извлечения терминов, их перевода и тематической классификации. Практическая значимость исследования заключается в возможности внедрения предложенного подхода в системы автоматизированного перевода, информационно-поисковые и аналитические платформы. Перспективы дальнейших исследований связаны с развитием многозадачного обучения, оптимизацией вычислительных затрат и расширением специализированных терминологических ресурсов.

В итоге сочетание BIO-разметки и XLM-RoBERTa показывает хорошие результаты при извлечении терминов. Данный подход обеспечивает высокую

точность, масштабируемость и универсальность для дальнейшего перевода, что делает его перспективным для интеграции в комплексные системы обработки текстов.

Список литературы:

1. Jurafsky D., Martin J. Speech and Language Processing. [Электронный ресурс]. URL: <https://web.stanford.edu/~jurafsky/slp3/> (дата обращения: 21.01.2026)
2. Vaswani A., Shazeer N., Parmar N., et al. Attention Is All You Need. NeurIPS, 2017. [Электронный ресурс]. URL: <https://arxiv.org/abs/1706.03762> (дата обращения: 15.01.2026)
3. Manning C., Raghavan P., Schütze H. Introduction to Information Retrieval. [Электронный ресурс]. URL: <https://nlp.stanford.edu/IR-book/> (дата обращения: 28.01.2026)
4. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL, 2019. [Электронный ресурс]. URL: <https://arxiv.org/abs/1810.04805> (дата обращения: 08.02.2026)
5. Aggarwal C. Machine Learning for Text. Springer, 2018. [Электронный ресурс]. URL: <https://link.springer.com/book/10.1007/978-3-319-73531-3> (дата обращения: 12.02.2026)
6. Conneau A., Khandelwal K., Goyal N., et al. Unsupervised Cross-lingual Representation Learning at Scale. ACL, 2020. [Электронный ресурс]. URL: <https://arxiv.org/abs/1911.02116> (дата обращения: 15.02.2026)

УДК 004.056:004.738.5

Фаттахов Э. Н., Тюрнев А. С.

Иркутский национальный исследовательский технический университет, Иркутск, Россия

АНАЛИЗ КИБЕРУГРОЗ ПРОМЫШЛЕННОЙ СЕТИ ИНТЕРНЕТА ВЕЩЕЙ И ПРИМЕНЕНИЕ ZERO TRUST ARCHITECTURE ДЛЯ ПРЕДОТВРАЩЕНИЯ УЩЕРБА

Аннотация. В статье рассматриваются актуальные киберугрозы промышленным сетям интернета вещей (IIoT), возникающие при цифровизации производства. Анализируются DDoS-атаки, ботнеты, уязвимости прошивок, слабая аутентификация и риски компрометации SCADA-систем и HMI-интерфейсов. Показано, что периметровая модель защиты недостаточна для распределённых инфраструктур, где устройства постоянно обмениваются данными. В качестве решения рассматривается Zero Trust Architecture, основанная на отсутствии доверия по умолчанию, минимальных привилегиях, микросегментации и непрерывной проверке доступа. Сделан вывод, что внедрение ZTA помогает локализовать инциденты и снизить ущерб от атак на IIoT-среду.

Ключевые слова: промышленный интернет вещей, IIoT, IoT, Zero Trust Architecture, кибербезопасность, ботнет, DDoS-атака, микросегментация.

Введение

Промышленный интернет вещей (Industrial Internet of Things, IIoT) представляет собой совокупность взаимосвязанных датчиков, контроллеров, исполнительных механизмов и аналитических платформ, обеспечивающих автоматизацию производственных процессов и передачу данных в режиме реального времени [1]. Внедрение IIoT позволяет предприятиям существенно повысить производительность, сократить эксплуатационные расходы и повысить эффективность мониторинга технологических процессов [4].

Одновременно с ростом количества подключённых устройств увеличивается число угроз информационной безопасности. Многие IoT-устройства используют устаревшее программное обеспечение, слабые механизмы аутентификации и не поддерживают своевременное обновление безопасности [2]. Это делает промышленную инфраструктуру привлекательной целью для злоумышленников.

Традиционные механизмы защиты, основанные исключительно на периметровой безопасности, становятся недостаточными в условиях распределённых сетей и удалённого доступа. В связи с этим всё большее распространение получает концепция Zero Trust Architecture (ZTA), предполагающая отказ от модели доверия по умолчанию [1].

1. Основные киберугрозы промышленным IoT-сетям

Одной из наиболее опасных угроз промышленным IoT-сетям являются распределённые атаки типа «отказ в обслуживании» (DDoS). Их целью является перегрузка сетевых ресурсов большим количеством запросов, что приводит к недоступности сервисов и нарушению работы инфраструктуры [3]. Для реализации подобных атак злоумышленники активно используют ботнеты — сети заражённых IoT-устройств.

Наиболее известным примером является ботнет Mirai, использовавший устройства с заводскими логинами и паролями по умолчанию. После заражения устройства становились частью распределённой сети, применяемой для проведения масштабных DDoS-атак [3]. Особую опасность представляет то, что многие IoT-устройства не оснащены встроенными средствами защиты, что существенно упрощает их компрометацию [5].

Исследования компании HP показали, что около 70 % IoT-устройств не используют шифрование при передаче данных, а более 60 % имеют уязвимые веб-интерфейсы [2]. Кроме того, значительная часть устройств использует простые или стандартные пароли, что существенно облегчает несанкционированный доступ.

Дополнительную угрозу представляют уязвимости встроенного программного обеспечения. Многие устройства работают на устаревших версиях операционных систем и не поддерживают автоматическое обновление прошивок. Это позволяет злоумышленникам использовать известные уязвимости для удалённого выполнения кода и получения полного контроля над устройствами [5].

Особую опасность представляют атаки на SCADA-системы, PLC-контроллеры и HMI-интерфейсы. Компрометация подобных систем может привести к остановке производственных процессов, авариям и серьёзным финансовым потерям [6]. В критической инфраструктуре последствия подобных атак могут представлять угрозу для жизни и здоровья сотрудников предприятия.

2. Концепция Zero Trust Architecture

Zero Trust Architecture представляет собой современную модель информационной безопасности, основанную на принципе отсутствия доверия по умолчанию к любому пользователю, устройству или сервису [1]. Согласно данной концепции, каждый запрос на доступ должен проходить обязательную аутентификацию и авторизацию независимо от расположения субъекта внутри или вне корпоративной сети.

Основными принципами ZTA являются:

- минимальные привилегии доступа;
- непрерывная аутентификация;
- микросегментация сети;
- постоянный мониторинг устройств;
- динамические политики безопасности.

Согласно стандарту NIST SP 800-207 архитектура Zero Trust включает три основных логических компонента: Policy Engine (PE), Policy Administrator (PA) и Policy Enforcement Point (PEP) [1]. Policy Engine отвечает за принятие решений о предоставлении доступа на основе анализа контекста подключения, уровня риска и политик безопасности. Policy Enforcement Point обеспечивает принудительное применение политик доступа и блокировку несанкционированных соединений.

Важной особенностью Zero Trust является постоянный контроль состояния устройств. Анализируется версия программного обеспечения, наличие обновлений безопасности, корректность конфигурации и признаки аномального поведения [1]. При выявлении подозрительной активности доступ устройства к критическим ресурсам может быть автоматически ограничен.

3. Применение ZTA в промышленной инфраструктуре

Одним из наиболее эффективных механизмов Zero Trust Architecture является микросегментация сети. Данный подход предполагает разделение инфраструктуры на небольшие изолированные сегменты, между которыми действует строгий контроль доступа [1]. Использование микросегментации позволяет локализовать атаки и предотвратить распространение вредоносного программного обеспечения внутри промышленной сети.

Даже при компрометации одного устройства злоумышленник не сможет получить полный контроль над всей инфраструктурой предприятия. Это особенно важно для IoT-среды, где большое количество устройств обменивается данными в режиме реального времени [4].

Непрерывная аутентификация также играет важную роль в обеспечении безопасности промышленной инфраструктуры. Для повышения безопасности используются технологии многофакторной аутентификации, цифровые сертификаты, VPN и системы управления идентификацией пользователей [1].

Для эффективной реализации Zero Trust Architecture в промышленной среде рекомендуется использовать:

- SIEM-системы;
- IDS/IPS;
- VPN, ZTNA;
- системы централизованного мониторинга;
- механизмы анализа сетевого трафика;
- контроль доступа на основе ролей.

Дополнительно рекомендуется регулярно проводить аудит безопасности IoT-устройств, обновлять встроенное программное обеспечение и запрещать использование стандартных учётных записей [5]. Комплексное применение данных мер позволяет существенно снизить вероятность успешной атаки и минимизировать последствия компрометации устройств.

Заключение

Рост количества устройств промышленного интернета вещей сопровождается увеличением числа киберугроз и уязвимостей. Традиционные методы защиты, основанные исключительно на периметровой безопасности, уже не способны эффективно противодействовать современным атакам в условиях распределённых инфраструктур [6].

Одним из наиболее эффективных подходов к обеспечению безопасности IoT является концепция Zero Trust Architecture. Использование принципов минимальных привилегий, микросегментации, непрерывной аутентификации и постоянного мониторинга позволяет существенно снизить вероятность компрометации промышленной сети и минимизировать потенциальный ущерб [1].

Таким образом, внедрение архитектуры Zero Trust является важным направлением развития систем информационной безопасности промышленного интернета вещей и необходимым условием обеспечения устойчивости критической инфраструктуры.

Список литературы:

1. Zero Trust Architecture : NIST Special Publication 800-207. – Gaithersburg : National Institute of Standards and Technology, 2020. – 59 p. – DOI: 10.6028/NIST.SP.800-207.
2. Шевцов, В. Ю. Анализ угроз и уязвимостей концепций ИОТ и ИОТ / В. Ю. Шевцов, Н. П. Касимовский // НБИ технологии. – 2020. – Т. 14, № 3. – С. 28-35. – DOI 10.15688/NBIT.jvolsu.2020.3.5. – EDN OHSEBR.
3. Лобова, А. И. Обзор DDOS-атак на ИОТ устройства / А. И. Лобова, Е. В. Вершинин, В. О. Федоров // Нацбезопасность. – 2022. – № 1(3). – С. 8-13. – EDN JXBPSA.
4. Cisco Annual Internet Report (2018–2023) White Paper. — Текст : электронный // Cisco : [сайт]. — URL: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (дата обращения: 10.05.2026).
5. Kaspersky unveils an overview of IoT-related threats in 2023. — Текст : электронный // Kaspersky : [сайт]. — URL: <https://www.kaspersky.com/about/press-releases/kaspersky-unveils-an-overview-of-iot-related-threats-in-2023> (дата обращения: 05.05.2026).
6. ГОСТ Р МЭК 62443-2-1-2015. Сети коммуникационные промышленные. Защищённость (кибербезопасность) сети и системы. Часть 2-1. Создание программы обеспечения защищённости промышленной автоматизации и системы управления. – Введ. 2016-08-01. – Москва : Стандартинформ, 2016.

УДК 004.85:519.237

Мацук В. А.

Донской государственный технический университет, Ростов-на-Дону, Россия

ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ИНКРЕМЕНТАЛЬНОГО МЕТОДА ГЛАВНЫХ КОМПОНЕНТ ПРИ ОБРАБОТКЕ ПОТОКОВЫХ ДАННЫХ

Аннотация: в статье рассматривается применение инкрементального метода главных компонент для обработки потоковых данных. Обосновывается актуальность использования алгоритмов, способных обновлять модель по мере поступления новых наблюдений без полного переобучения на всей накопленной выборке. Рассматриваются ограничения классического метода главных компонент при работе с данными большого объема и непрерывного поступления. Приводится экспериментальное сравнение классического PCA и Incremental PCA на синтетическом наборе данных с оценкой доли объяснённой дисперсии и ошибки восстановления. Сделан вывод о целесообразности применения инкрементального подхода в задачах предварительной обработки потоковых данных.

Ключевые слова: потоковые данные, метод главных компонент, Incremental PCA, снижение размерности, машинное обучение, анализ данных.

В настоящее время объёмы данных, генерируемых информационными системами, постоянно увеличиваются. При этом многие наборы данных поступают не как заранее сформированная статическая выборка, а в виде непрерывного потока наблюдений. К таким данным можно отнести телеметрию датчиков, сетевой трафик, финансовые транзакции, логи веб-сервисов и данные пользовательской активности. В подобных условиях возрастает значение методов, которые позволяют выполнять обработку информации по мере её поступления.

Потоковые данные отличаются от статических наборов несколькими особенностями. Во-первых, их объём может быть практически неограниченным, поэтому хранение всей истории наблюдений в оперативной памяти часто является затруднительным. Во-вторых, данные поступают последовательно, а значит, модель должна иметь возможность обновляться без полного пересчёта всех параметров. В-третьих, в реальных системах возможно изменение распределения данных во времени, из-за чего особенно важными становятся адаптивность и вычислительная эффективность используемых алгоритмов [1].

Одной из распространённых задач предварительной обработки данных является снижение размерности признакового пространства. При большом количестве признаков возрастает вычислительная сложность последующих алгоритмов машинного обучения, а также усложняется интерпретация структуры данных. Метод главных компонент (Principal Component Analysis,

РСА) позволяет перейти от исходных признаков к новому набору ортогональных компонент, упорядоченных по величине объясняемой дисперсии [2; 6]. За счёт этого можно сократить количество признаков, сохранив значительную часть информации, содержащейся в исходной выборке.

Классический РСА эффективен при работе со статическими наборами данных, однако его применение в потоковом сценарии имеет ограничение: для построения компонент обычно требуется доступ ко всей выборке. Если новые наблюдения поступают непрерывно, то при изменении выборки возникает необходимость повторного пересчёта модели. При больших объёмах данных такой подход может приводить к повышенным затратам памяти и времени обработки. Поэтому для потоковых данных целесообразно использовать модификации метода, способные выполнять обновление модели частями.

Инкрементальный метод главных компонент (Incremental PCA) направлен на решение указанной проблемы. В отличие от классического подхода, он обучается не на всей выборке сразу, а на последовательных фрагментах данных, которые часто называют батчами. После поступления очередного батча модель уточняет оценки главных компонент и продолжает использовать уже накопленную информацию. Такой принцип делает Incremental PCA применимым в случаях, когда полный набор данных не помещается в память или поступает постепенно [3].

С практической точки зрения Incremental PCA может использоваться как этап предварительной обработки перед построением моделей классификации, регрессии, кластеризации или визуального анализа данных. Например, в системах мониторинга данные от датчиков могут поступать постоянно, а алгоритм снижения размерности позволяет выделять наиболее информативные направления изменения признаков. В библиотеке scikit-learn данный подход реализован в классе IncrementalPCA, который поддерживает поэтапное обучение через механизм частичного обновления модели [4; 5].

Методика экспериментального исследования

Для демонстрации работы подхода был сформирован синтетический набор данных с латентной структурой. Набор содержал 30000 наблюдений и 40 признаков. Данные были сгенерированы таким образом, чтобы основная часть информации определялась шестью скрытыми факторами, после чего к ним был добавлен небольшой шум. Такая постановка позволяет проверить, насколько классический РСА и Incremental PCA способны восстановить исходную низкоразмерную структуру.

Перед применением алгоритмов признаки были стандартизированы. Для сравнения использовались шесть главных компонент. Классический РСА обучался на всей выборке сразу. Incremental PCA обучался последовательно: выборка была разделена на батчи по 1500 наблюдений, после чего модель обновлялась на каждом батче. Качество оценивалось по суммарной доле

объяснённой дисперсии и среднеквадратичной ошибке восстановления данных после обратного преобразования.

Таблица 1. Результаты сравнения PCA и Incremental PCA

Метод	Способ обучения	Доля объяснённой дисперсии	Ошибка восстановления MSE	Особенность применения
PCA	обучение на всей выборке	0,9954	0,004612	требует обработки полной выборки
Incremental PCA	обучение батчами	0,9954	0,004612	позволяет обновлять модель по мере поступления данных

Полученные результаты показывают, что при заданных параметрах оба метода сохранили практически одинаковую долю исходной информации. Значение ошибки восстановления также оказалось одинаковым с точностью до используемого округления. Это объясняется тем, что синтетический набор данных действительно имел выраженную низкоразмерную структуру, которую удалось выделить с помощью шести компонент.

При этом интерпретировать результаты следует с учётом различий в сценариях применения методов. Классический PCA удобен в ситуации, когда вся выборка заранее доступна и помещается в память. Incremental PCA особенно полезен тогда, когда данные поступают частями, а модель необходимо обновлять без полного переобучения. Поэтому главное преимущество инкрементального подхода заключается не только в значении метрик на отдельном эксперименте, но и в возможности организации поэтапной обработки данных.

В ходе исследования также было установлено, что выбор размера батча влияет на практическую работу алгоритма. Слишком маленький размер батча может увеличивать накладные расходы на частые обновления модели, а слишком большой размер батча приближает алгоритм к классической обработке полной выборки. Следовательно, при применении Incremental PCA необходимо подбирать размер батча с учётом объёма данных, числа признаков и доступных вычислительных ресурсов.

Выводы

В статье рассмотрены особенности применения инкрементального метода главных компонент для обработки потоковых данных. Показано, что классический PCA хорошо подходит для статических наборов данных, но менее удобен при непрерывном поступлении новых наблюдений, так как требует обработки всей выборки. Incremental PCA позволяет обновлять модель

последовательно, что делает его более подходящим для потоковых и больших данных.

Экспериментальное моделирование показало, что Incremental PCA способен обеспечить качество снижения размерности, сопоставимое с классическим PCA, при сохранении возможности поэтапного обучения. Дальнейшее развитие исследования может быть связано с проверкой алгоритма на реальных потоковых данных, сравнением разных размеров батча и использованием полученных компонент в задачах классификации или прогнозирования.

Список литературы:

1. Aggarwal, C. C. Data Streams: Models and Algorithms / C. C. Aggarwal. - New York: Springer, 2007. - 354 p.
2. Jolliffe, I. T. Principal Component Analysis / I. T. Jolliffe. - 2nd ed. - New York: Springer, 2002. - 487 p.
3. Ross, D. A. Incremental Learning for Robust Visual Tracking / D. A. Ross, J. Lim, R.-S. Lin, M.-H. Yang // International Journal of Computer Vision. - 2008. - Vol. 77, № 1-3. - P. 125-141. DOI: 10.1007/s11263-007-0075-7.
4. Scikit-learn. sklearn.decomposition.IncrementalPCA : documentation [Электронный ресурс]. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.IncrementalPCA.html> (дата обращения: 13.04.2026).
5. Pedregosa, F. Scikit-learn: Machine Learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort [et al.] // Journal of Machine Learning Research. - 2011. - Vol. 12. - P. 2825-2830.
6. Oja, E. A Simplified Neuron Model as a Principal Component Analyzer / E. Oja // Journal of Mathematical Biology. - 1982. - Vol. 15. - P. 267-273. DOI: 10.1007/BF00275687.

УДК 621.717:658.53

Дубовиков Н. А.
МГТУ «СТАНКИН», Москва, Россия

УЧЕТ ВРЕМЕНИ ОЖИДАНИЯ ПРИ ВЫПОЛНЕНИИ СБОРОЧНЫХ РАБОТ

Аннотация. В статье рассматривается научно-практическая задача учета и нормирования времени ожидания рабочих при выполнении сборочных работ в машиностроении. Межоперационные простои и задержки из-за несбалансированности такта напрямую снижают эффективность потока, искажая реальную трудоемкость процессов.

Ключевые слова: сборочные работы, циклограмма сборки, время ожидания, балансировка сборочных линий, потери времени

В контексте технологической подготовки производства циклограмма сборки изделий представляют собой инструмент графического отображения временной структуры и последовательности выполнения элементов технологического процесса. В научно-методической литературе по технологии машиностроения данная модель классически детерминируется как двухмерная диаграмма, в которой ось абсцисс выступает шкалой времени, а ось ординат отражает технологические операции, переходы или сборочные приемы. Отраслевая практика нормирования труда обосновывает целесообразность декомпозиции циклограмм по конкретным исполнителям и единицам технологического оборудования, что позволяет осуществлять оперативный мониторинг межоперационных задержек. Данный подход, базирующийся на методах хронометража и принципах организации поточного производства, является валидным аналитическим инструментом для оптимизации балансировки сборочных линий и снижения непроизводительных затрат времени.

Применение циклографического моделирования при нормировании слесарно-сборочных работ позволяет эффективно решать следующие задачи:

1. Расчет и анализ общей длительности сборочного цикла
Циклограмма обеспечивает точный расчет времени, необходимого для сборки единичного изделия или целой партии. С ее помощью общее время наглядно разделяется на основное (технологическое) и вспомогательное. Такое графическое представление позволяет детально проверить структуру штучно-калькуляционного времени, точно фиксируя регламентированные перерывы, время на обслуживание рабочего места и скрытые потери на ожидание деталей [1, с. 112; 2, с. 45].

2. Выявление «узких мест» и оптимизация процесса сборки
График позволяет быстро обнаружить самые длительные операции и переходы, которые задерживают работу всей линии («узкие места»). Анализ

циклограммы помогает сократить эти задержки с помощью инженерных решений: перераспределения обязанностей между рабочими, совмещения нескольких приемов во времени или автоматизации ручного труда [3, с. 84]. В результате нагрузка на персонал выравнивается, а общая длительность сборки снижается.

3. Синхронизация потока и управление сборкой партий изделий

В серийном машиностроении циклограмма помогает подстроить скорость выполнения работ под заданный такт выпуска продукции. Сопоставление реального графика сборки с плановым тактом линии позволяет организовать непрерывный процесс. Это сводит к минимуму простой персонала в ожидании комплектующих и исключает появление избыточных заделов незавершенного производства, обеспечивая общую ритмичность потока [4, с. 144].

В процессе организации слесарно-сборочных работ часто возникают межоперационные задержки, при которых исполнители вынуждены находиться в ожидании из-за незавершенности предшествующих этапов сборки. Данное время ожидания $T_{ож}$ относится к непроизводительным затратам и подлежит обязательной минимизации. При сборке партий изделий величина $T_{ож}$ напрямую увеличивает общую длительность производственного цикла партии $T_{общ.парт}$, что снижает общую эффективность производственного процесса и приводит к нерациональному использованию рабочего времени [5, с. 62].

Одним из путей минимизации непроизводительных потерь является рациональное расширение зоны обслуживания или дозагрузка работника дополнительными технологическими переходами. С точки зрения технического нормирования, такое совмещение целесообразно в тех случаях, когда величина непрерывного времени ожидания $T_{ож.непр.}$ превышает суммарную длительность вводимой трудоемкости (дозагрузки) $T_{дозагр.}$. Соблюдение данного условия позволяет трансформировать скрытые простои в полезную занятость без нарушения общего такта линии [6, с. 172]. Подобный подход, основанный на анализе циклограмм, обеспечивает максимальное уплотнение рабочего дня, оптимизирует коэффициент загрузки персонала и повышает производительность сборочного участка.

Для практической реализации метода дозагрузки рабочих целесообразно применять алгоритм анализа циклограмм сборки, построенных применительно к рабочим-сборщикам, принимающим участие в процессе сборки. Данный алгоритм базируется на принципах графоаналитического моделирования и включает в себя три ключевых этапа:

1. Проектирование нормативной циклограммы сборочного процесса. На начальном этапе осуществляется построение циклограммы сборки, разбитой по рабочим. На графике фиксируется полное расчетное

время, регламентированное нормативными картами для выполнения всего объема работ по сборке единицы изделия.

2. Разработка фактической циклограммы на основе хронометражных данных. Этап заключается в построении циклограммы, отражающей реальные временные затраты сборщиков на выполнение закрепленных операций и переходов. График формируется по результатам натурных наблюдений. Это необходимо для фиксации темпа работы и выявления отклонений от первоначального регламента.

3. Формирование совмещенной (аналитической) циклограммы и локализация простоев. Заключительным этапом является синтез данных в единую совмещенную циклограмму. Путем наложения двух графиков визуализируются и измеряются межоперационные простои и скрытые окна ожидания $T_{ож.непр}$. Полученная модель служит основой для последующего принятия решений о дозагрузке персонала и перераспределении трудоемкости.

При поточной сборке изделий в реальных условиях сборочные операции выполняются непрерывно. Рабочие переходят от одного изделия к другому с целью как можно быстрее завершить сборку партии изделий, так как это напрямую влияет на оплату труда. В таком случае, при наличии окон ожидания целесообразна дозагрузка работников выполнением работ по сборке последующих изделий.

В качестве примера рассмотрим сборку одноступенчатого редуктора. (Рисунок 1)

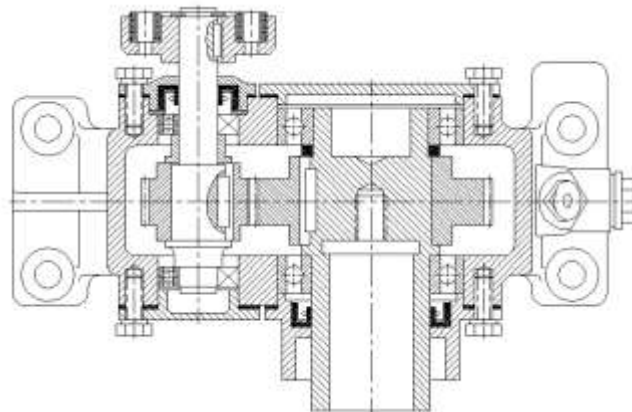


Рисунок 1. Редуктор одноступенчатый

Для построения циклограммы сборки необходима схема сборки редуктора. (Рисунок 2)

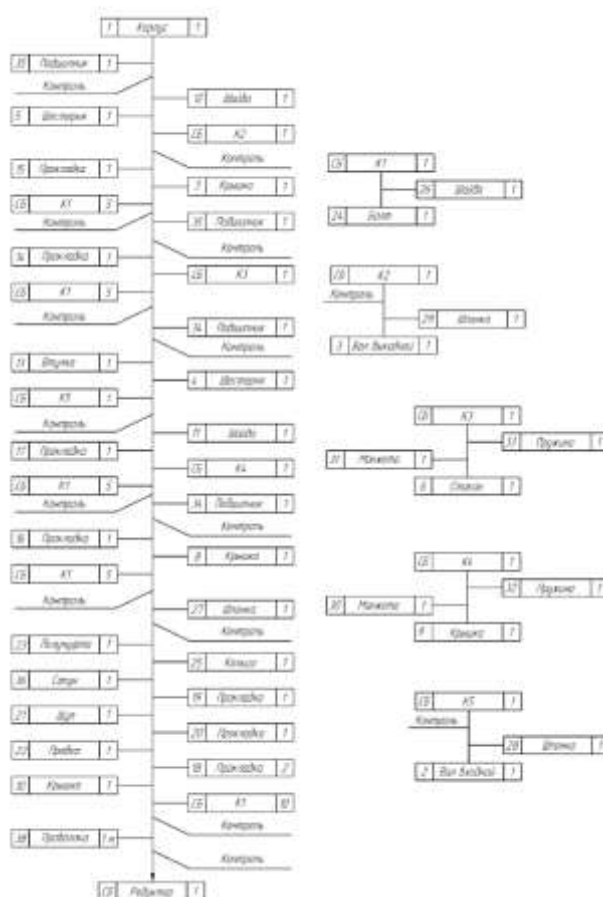


Рисунок 2. Схема сборки редуктора

Далее выделим штучное время $T_{шт.}$ для выполнения операций сборки, а также время востребованности $T_{востр.}$ для каждого из собираемых комплектов и их количество. Время востребованности необходимо для понимания того, через какой промежуток времени для сборки редуктора будет необходимо использование уже собранных комплектов. (Таблица 1)

Таблица 1. Информация о штучном времени и времени ожидания

Операция	$T_{шт.}, \text{мин}$	$T_{востр.}, \text{мин}$	Кол-во
Общая сборка редуктора	73,83	-	1
Сборка комплекта 1	0,16	8,57	30
Сборка комплекта 2	3,84	4,35	1
Сборка комплекта 3	1,44	17,58	1
Сборка комплекта 4	1,44	31,76	1
Сборка комплекта 5	3,84	27,79	1
Итого:	89,24	-	-

На основе схемы сборки (Рисунок 2) и имеющейся информации о времени (Таблица 1) построим циклограмму сборки редуктора с учетом времени востребованности. (Рисунок 3)

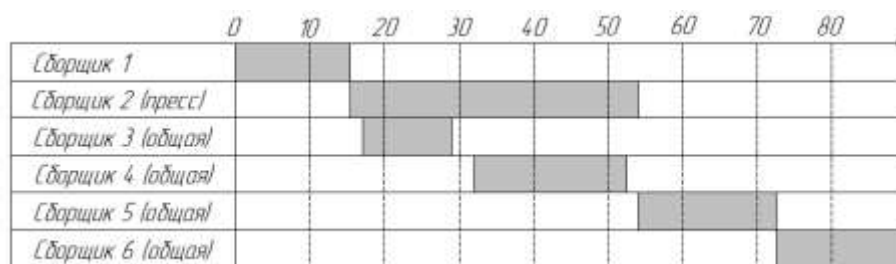


Рисунок 3. Циклограмма сборки редуктора

В сборочном процессе задействовано 6 рабочих (сборщиков): сборщик один выполняет сборку комплектов (1-5); сборщик 2 выполняет операции по запрессовке; сборщики 3-6 выполняют общую сборку редуктора.

Далее необходимо построить циклограмму сборки, отражающую фактические затраты времени на сборку редуктора с учетом времени вхождения для каждого из собираемых комплектов. (Рисунок 4)

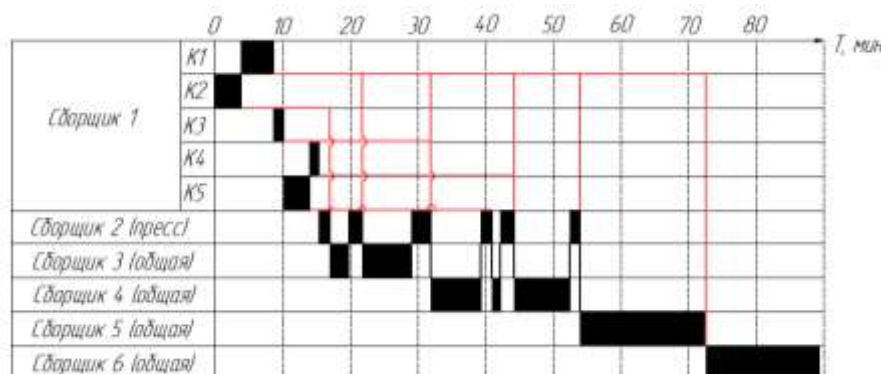


Рисунок 4. Циклограмма сборки редуктора с учетом фактических затрат времени

После построения данной циклограммы можно сделать предварительный вывод о том, что сборщики 2-4 имеют время ожидания $T_{ож.}$ для выполнения последующих переходов. Для подтверждения этого вывода построим совмещенную циклограмму сборки и выделим цветом $T_{ож.}$ для наглядной демонстрации его наличия. (Рисунок 5)

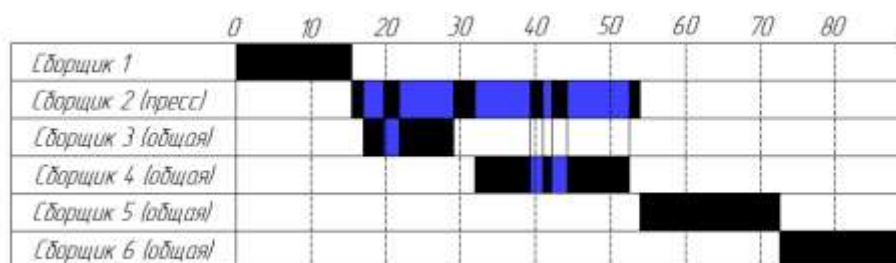


Рисунок 5. Совмещенная циклограмма сборки

Таким образом, подтверждено наличие окон ожидания для сборщиков 2-4. Для сборщика 2 $T_{ож.непр.} > T_{шт.}$ и таким образом возможно уменьшение $T_{ож.}$ для данного сборщика.

Одним из возможных вариантов минимизации $T_{ож.}$ является дозагрузка рабочих. Исходя из того, что сборщик 1 по завершении сборки комплектов для первого редуктора в партии приступит к сборке комплектов для последующих редукторов, целесообразной будет дозагрузка сборщиков путем реализации этапов сборки последующих редукторов.

Заключение

Построение циклограмм сборки является эффективным инструментом для поиска и устранения простоев на производстве благодаря своей простоте и наглядности. С помощью них возможен контроль за выполнением планов работ, загрузкой оборудования и рабочих, а также возможно оперативное вмешательство в производственные процессы с целью повышения их эффективности.

Оптимизация циклов сборки последующих изделий в партии путем анализа циклограмм сборки позволяет:

1. Увеличить коэффициент загрузки сборщиков за счет уменьшения времени ожидания $T_{ож.}$;
2. Оптимизировать затраты на осуществление сборки путем перераспределения сборочных работ на меньшее количество сборщиков;
3. Уменьшить такт выпуска изделий и, соответственно, уменьшить их себестоимость.

Список литературы:

1. Силантьева, Н. А. Техническое нормирование труда в машиностроении: учебник для вузов / Н. А. Силантьева, В. Р. Малиновский. - М.: Машиностроение, 1990. - 256 с.
2. Общемашиностроительные нормативы времени на слесарную обработку и сборку машин при сборке металлоконструкций / М-во труда и социальной защиты РФ. - М.: Экономика, 2018. - 198 с.
3. Колесов, Игорь Михайлович. Основы технологии машиностроения [Текст]: учебник для вузов / И. М. Колесов. - 3-е изд., стереотип. - М.: Высшая школа, 2001. - 591 с. - (Технология, оборудование и автоматизация машиностроительных производств). - Библиогр.: с.587. ISBN: 5-06-003662-6
4. Технология машиностроения (производство машин): учебно-методический комплекс для студентов специальности 36 01 01 "Технология машиностроения": в 3 частях. Часть 1 / Министерство образования Республики Беларусь, Полоцкий государственный университет; сост. А.А. Лысов, А.С. Аршиков. - Новополоцк: ПГУ, 2007. - 252 с.
5. Технология машиностроения [Текст]: учебное пособие / В. Ф. Безъязычный; Минобрнауки России, Федеральное гос. бюджетное образовательное учреждение высш. проф. образования "Рыбинский гос. авиационный технический ун-т им. П. А. Соловьева". - Рыбинск: РГАТУ, 2015-. Ч. 2. - 2015. - 202 с.
6. Курсовое проектирование по технологии машиностроения: [Учеб. пособие для машиностроит. спец. вузов] / Л. В. Худобин, В. Ф. Гурьянихин, В. Р. Берзин. - Москва: Машиностроение, 1989. - 287 с.: схем.; 21 см.; ISBN: 5-217-00340-5

1.

УДК 004.85

Петухов Д. А.¹, Петухов А. Б.², Ефимова В. А.¹

¹ *Национальный исследовательский университет ИТМО, Санкт-Петербург, Россия*

² *Военно-космическая академия имени А. Ф. Можайского, Санкт-Петербург, Россия*

ПРОГНОЗИРОВАНИЕ КРИВОЙ УДЕРЖАНИЯ АУДИТОРИИ ВИДЕОКОНТЕНТА НА ОСНОВЕ ВЫСОКОУРОВНЕВЫХ ПРИЗНАКОВ

Аннотация: Статья посвящена задаче прогнозирования кривой удержания аудитории видеороликов на платформе YouTube до момента их публикации. Особое внимание уделяется работе в условиях малой выборки, обусловленной труднодоступностью аналитики удержания, доступной только владельцам каналов. Цель исследования – разработать и сравнить методы восстановления нормированной кривой удержания на основе компактного набора высокоуровневых признаков, извлекаемых из транскрипта, визуального ряда и метаданных видео. В работе предлагается формулировка задачи как реконструкции 50-точечной нормированной кривой и проводится систематическое сопоставление пяти стратегий моделирования при строгой оценке методом исключения по одному. Эксперименты выполнены на корпусе из 90 видео. Лучшие результаты получены интеграционно-чувствительной регрессионной моделью, которая снижает среднеквадратичную ошибку на 20,9% по сравнению с глобальным средним. Установлено, что семантические признаки из транскрипта в сочетании с визуальными признаками видео содержат устойчивый прогностический сигнал о динамике удержания, а явный учёт спонсорских вставок улучшает реконструкцию полной кривой.

Ключевые слова: удержание аудитории, прогнозирование временного ряда, видеоконтент, машинное обучение на малых выборках, градиентный бустинг, обработка транскриптов, рекомендательные системы.

Введение

Видеохостинговые платформы, прежде всего YouTube, радикально изменили способ создания и потребления видеоконтента. Одной из важнейших метрик качества видео является удержание аудитории, отражающее не сам факт клика, а долю зрителей, продолжающих просмотр в каждый момент длительности ролика. Согласно данным из работы [1, с. 1], предоставляемые YouTube измерения удержания дают ценную информацию для совершенствования видео и более глубокого понимания интересов зрителя. Для авторов, монтажёров и рекомендательных систем кривая удержания значительно более информативна, чем единственное агрегированное по всему видео число: она показывает, в какой момент внимание зрителей падает, где оно стабилизируется, а где, напротив, возрастает.

Несмотря на практическую значимость, задача прогнозирования удержания остаётся слабо изученной в области компьютерного зрения и

мультимодальных исследований. Большинство ранее опубликованных работ сосредоточено на скалярных величинах – популярности, количестве просмотров или агрегированной вовлечённости [2, с. 2561; 3, с. 74277; 4, с. 5]. Существующие подходы к моделированию временной вовлечённости также чаще решают задачи, более близкие к прогнозу времени просмотра, чем к восстановлению полной кривой удержания [5, с. 4639; 6, с. 3].

В нашей работе предполагается, что значительная часть динамики удержания может быть объяснена структурой контента и визуальными характеристиками видео. Сигналы такого рода, как эмоционально насыщенные фрагменты или рекламные интеграции, как правило, связаны с предсказуемой реакцией аудитории. Это служит обоснованием подхода, в котором каждое видео представляется не низкоуровневыми пиксельными или акустическими дескрипторами, а компактным набором высокоуровневых признаков, описывающих структурно-семантические свойства материала, динамику монтажа и присутствие человека в кадре.

Основной вклад работы состоит в методе прогнозирования полной кривой удержания аудитории YouTube-видео по компактному высокоуровневому представлению, построенному на основе транскрипта, визуального ряда и метаданных. Целью исследования является сравнительный анализ нескольких стратегий моделирования и выявление наиболее эффективной из них при строгом протоколе кросс-валидации.

Материалы и методы

Каждое видео представляется компактным вектором высокоуровневых признаков размерности порядка 40, объединяющим характеристики транскрипта, признаки уровня видео и метаданные. Целевой переменной является нормированная кривая удержания фиксированной длины из 50 значений, равномерно распределённых по нормированному прогрессу видео; каждое значение отражает долю исходной аудитории, продолжающую просмотр в соответствующий момент. Такая постановка превращает задачу в задачу плотного временного предсказания: модель должна восстановить не скаляр, а полную форму и уровень траектории удержания. В настоящем разделе основной акцент сделан на сравнении стратегий моделирования.

Стратегии моделирования

Рассматриваются пять стратегий, отличающиеся допущениями о структуре задачи. Сопоставление важно, поскольку заранее не очевидно, какое из априорных допущений наилучшим образом подходит к прогнозу кривых удержания.

Базовая модель: глобальная средняя кривая. Базовый прогноз представляет собой среднюю по обучающей выборке кривую. Несмотря на простоту, базовая модель нетривиальна, так как многие кривые удержания имеют схожую крупномасштабную форму. Для тестового видео i базовая модель определяется как:

$$\hat{y} = \frac{1}{N-1} \sum_{j \neq i} y_j$$

Поточечная регрессия. Используются поточечные регрессионные модели на основе CatBoost [7, с. 6639]: для каждой временной позиции p обучается отдельный регрессор, предсказывающий абсолютное значение удержания $\widehat{y}_p^{\text{abs}}$ из вектора признаков \mathbf{x} .

Дополнительно обучаются модели на разностях между соседними точками для лучшего воспроизведения локальных падений и пиков:

$$\widehat{\delta}_p \approx y_p - y_{p-1}, \quad p = 1, \dots, P-1$$

Кривая восстанавливается рекурсивно:

$$\widehat{y}_0^\delta = \widehat{y}_0^{\text{abs}}, \quad \widehat{y}_p^\delta = \text{clip}_{[0,1]} \left(\widehat{y}_{p-1}^\delta + \text{clip}_{[-\Delta_{\max}, \Delta_{\max}]}(\widehat{\delta}_p) \right)$$

Финальное предсказание строится как смесь абсолютной и дельта-реконструкций:

$$\hat{y} = (1 - \alpha) \widehat{y}^{\text{abs}} + \alpha \widehat{y}^\delta$$

, где $\alpha = 0,65$ и $\Delta_{\max} = 0,25$ в нашей реализации.

Интеграционно-чувствительная регрессия применяет постобработку на наиболее выраженном непрерывном участке с рекламой (точки, где $s_p \geq 0,5$, s_p – вероятность наличия рекламы в отрезке).

На этом сегменте кривая дополнительно сдвигается вниз пропорционально силе интеграции:

$$\widetilde{y}_p = \text{clip}_{[0,1]}(\widehat{y}_p - I\{p \in \mathcal{S}\} \Delta_{\text{ad}} \cdot s_p)$$

, где $s_p \in [0,1]$, \mathcal{S} – самый длинный непрерывный участок с рекламой; в экспериментах брал $\Delta_{\text{ad}} = 0,08$.

Подход опирается на независимые наблюдения о том, что спонсорские вставки являются источником наиболее острых провалов удержания [8; 9].

Модель формы. В этом варианте модель восстанавливает не абсолютные значения, а нормированную форму кривой удержания.

Сначала определяется нормировка для каждой точки – среднее по первым K точкам:

$$a = \frac{1}{K} \sum_{p=0}^{K-1} y_p$$

Затем нормализованная форма вычисляется как:

$$y_{p,\text{shape}} = \text{clip}_{[0,1]} \left(\frac{y_p}{\max(\varepsilon, a)} \right), \quad y_{0,\text{shape}} = 1$$

Для каждой точки обучаются отдельные регрессоры на разностях соседних точек $\delta_p = y_{p,\text{shape}} - y_{p-1,\text{shape}}$ и кривизне $\kappa_p = y_{p,\text{shape}} - 2y_{p-1,\text{shape}} + y_{p-2,\text{shape}}$.

В итоге форма реконструируется рекурсивно:

$$\widehat{y}_0^{\text{shape}} = 1, \quad \widehat{y}_p^{\text{shape}} = \widehat{y}_{p-1}^{\text{shape}} + d_p$$

$$d_p = (1 - \lambda) \widehat{\delta}_p + \lambda \left(\left(\widehat{y}_{p-1}^{\text{shape}} - \widehat{y}_{p-2}^{\text{shape}} \right) + \widehat{\kappa}_p \right)$$

где $\lambda = 0,55$, которая дает баланс между дельта-обновлением и коррекцией на основе кривизны.

Этот вариант служит контрольным и показывает, сколько прогностического сигнала содержится в одной только форме кривой без учёта её абсолютного уровня.

Стэкинг-ансамбль комбинирует три базовых модели (абсолютная регрессия, дельта-модель, модель для формы) через ridge-регрессию [10, с. 55]:

$$\widehat{y}_p = \beta_{p,0} + \beta_{p,1} \widehat{y}_p^{\text{abs}} + \beta_{p,2} \widehat{y}_p^{\delta} + \beta_{p,3} \widehat{y}_p^{\text{shape}} + \beta_{p,4} u_p$$

, где $u_p = \frac{p}{p-1}$ – нормализованная позиция точки. Предсказания первого уровня строятся по схеме out-of-fold для обучающих видео. Такой ансамбль пытается учитывать сразу несколько подходов для предсказания retention графика.

Эксперименты

Эксперименты проведены на корпусе из 90 длинных YouTube-видео, для каждого из которых доступны транскрипт автоматического распознавания речи, кривая удержания и метаданные уровня видео. Кривая удержания ресемплируется в 50 нормированных точек. Использован строгий протокол кросс-валидации методом исключения по одному: каждое видео ровно один раз выступает в роли тестового, тогда как остальные образуют обучающую выборку. Для оценки качества прогноза применяется набор взаимодополняющих метрик: коэффициенты корреляции Спирмена и Пирсона измеряют точность воспроизведения формы кривой; среднеквадратичная ошибка (RMSE) и средняя абсолютная ошибка (MAE) отражают абсолютную ошибку; метрика Spike-RMSE характеризует точность оценки локальных изменений.

Результаты исследования

В табл. 1 приведены результаты кросс-валидации методом исключения по одному для всех сравниваемых моделей.

Таблица 1. Результаты прогнозирования полной кривой удержания на 90 длинных YouTube-видео по протоколу исключения по одному. Стрелки \uparrow/\downarrow указывают желаемое направление изменения метрики. Лучшие значения в каждом столбце выделены полужирным шрифтом

Μοδελή	Spearman ↑	Pearson ↑	RMSE ↓	MAE ↓	Spike-RMSE ↓
Интеграцио- нно- чувствитель- ная регрессия	0,9347	0,9325	0,0756	0,0678	0,02469
Ποτοχечная регрессия	0,9116	0,9204	0,0767	0,0684	0,02445
Стэкинг- ансамбль	0,9072	0,9170	0,0782	0,0699	0,02599
Глобальное среднее (базовая)	0,9216	0,9205	0,0956	0,0877	0,02473
Μοδελή формы	0,5055	0,6565	0,6200	0,5994	0,06341

Интеграционно-чувствительная регрессионная модель достигает наилучших значений по большинству метрик. По метрике RMSE она превосходит базовую модель, что соответствует относительному снижению ошибки приблизительно на 20,9%. Этот результат свидетельствует о наличии устойчивого прогностического сигнала в семантических и визуальных признаках, извлекаемых из видео, по сравнению с глобальным усреднением кривой удержания.

Базовая модель сама по себе демонстрирует достаточно высокое качество по корреляционным метрикам, что свидетельствует о выраженной общей структуре кривых удержания – монотонном убывании в среднем с характерными колебаниями. Поэтому необходимо контролировать не только корреляционные показатели, но и абсолютные ошибки (RMSE и MAE) и точность локальных изменений (Spike-RMSE). Стэкинг-ансамбль показывает средние результаты, не выделяясь и не проваливаясь ни по одной из метрик, что говорит о невысокой выгоде от смешивания специализированных предсказателей в условиях малой выборки.

Качественное сопоставление прогнозов и истинной кривой для отдельного видео из выборки приведено на рис. 1: интеграционно-чувствительная модель ближе к целевой кривой как по корреляционным, так и по абсолютным метрикам, тогда как стэкинг-ансамбль воспроизводит общий нисходящий тренд, но сглаживает локальные изменения.

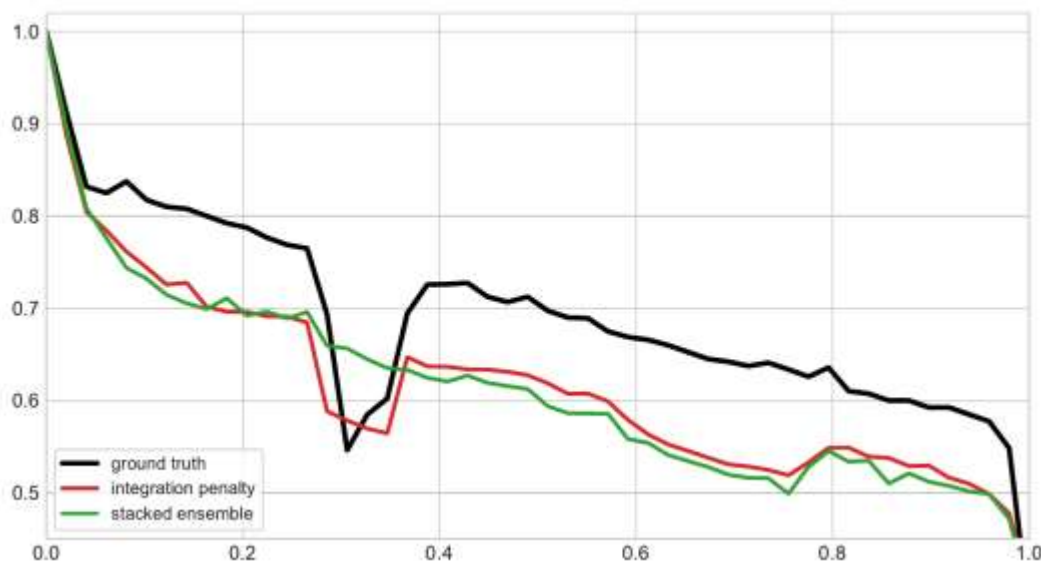


Рисунок 1. Качественное сопоставление прогнозов для отдельного видео: истинная кривая удержания и предсказания интеграционно-чувствительного регрессора и стэкинг-ансамбля

Полученные результаты подтверждают гипотезу о том, что значительная часть динамики удержания объясняется не низкоуровневыми пиксельными или акустическими сигналами, а высокоуровневой структурой контента – риторикой, эмоциональной окраской и наличием интеграций. Подобное наблюдение согласуется с выводами [1, с. 81] о том, что удержание содержит интерпретируемые компоненты, поддающиеся объяснению на уровне семантики содержания. Использование индикаторов интеграции в качестве структурированных признаков, а не как самостоятельной задачи детектирования, оказалось практически эффективным: явная штрафная коррекция на сегменте интеграции позволяет модели лучше воспроизводить характерные «провалы» в кривой удержания.

Ограничения исследования включают сравнительно малый размер выборки и доминирование в ней видео формата «говорящая голова», что не позволяет делать широких выводов о переносимости подхода на принципиально иные типы контента. Кроме того, текущее представление не охватывает ряд потенциально важных сигналов, в частности динамики звукового сопровождения и подробных визуальных характеристик кадров, влияние которых на удержание заслуживает отдельного изучения.

Заключение

В работе предложен подход к прогнозированию полной кривой удержания аудитории длинных YouTube-видео. Вместо высокоразмерных сырых входов каждое видео представлено компактным набором высокоуровневых признаков, что делает задачу пригодной для классических моделей табличных данных, в том числе градиентного бустинга. Систематическое сравнение пяти стратегий прогнозирования при строгой

кросс-валидации методом исключения по одному на 90 видео показало, что интеграционно-чувствительная регрессия достигает наилучшего качества: RMSE составила 0,0756, что соответствует относительному снижению ошибки на 20,9% по сравнению с базовой моделью глобального среднего. Это свидетельствует о том, что явный учёт интеграционных провалов особенно полезен для прогнозирования кривой удержания.

Направлениями дальнейших исследований являются масштабирование выборки, улучшение моделирования локальных временных эффектов и изучение возможностей интеграции таких прогнозов в реальные процессы доработки контента на этапе препродакшна.

Список литературы:

1. Altman E., Jimenez T. Measuring audience retention in YouTube // Proceedings of the 12th EAI International Conference on Performance Evaluation Methodologies and Tools (VALUETOOLS'19). — ICST, 2019. — P. 79–85.
2. Trzciński T., Rokita P. Predicting popularity of online videos using support vector regression // IEEE Transactions on Multimedia. — 2017. — Vol. 19, № 11. — P. 2561–2570.
3. Bielski A., Trzciński T. Understanding multimodal popularity prediction of social media videos with self-attention // IEEE Access. — 2018. — Vol. 6. — P. 74277–74287.
4. Wu S., Rizoiu M.-A., Xie L. Beyond views: measuring and predicting engagement in online videos // Proceedings of the International AAAI Conference on Web and Social Media (ICWSM). — 2018. — Vol. 12, № 1. — P. 434–443.
5. Lin X., Chen X., Song L., Liu J., Li B., Jiang P. Tree based progressive regression model for watch-time prediction in short-video recommendation // Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'23). — 2023. — P. 4497–4506.
6. Stappen L., Baird A., Lienhart M., Bätz A., Schuller B. An estimation of online video user engagement from features of time- and value-continuous, dimensional emotions // Frontiers in Computer Science. — 2022. — Vol. 4. — Article 773154. — P. 1–16.
7. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. CatBoost: unbiased boosting with categorical features // Advances in Neural Information Processing Systems. — 2018. — Vol. 31. — P. 6639–6649.
8. Ramachandran A. SponsorBlock: crowd-sourced sponsor segment skipping for YouTube [Электронный ресурс]. — URL: <https://sponsor.ajay.app/> (дата обращения: 06.05.2026).
9. Athreya N., Gokmen C., Yang J. DeepSponsorBlock: identifying sponsored content in YouTube videos with deep learning // Stanford CS 230 Project Reports, Fall 2020. — 2020. — P. 1–5.
10. Hoerl A. E., Kennard R. W. Ridge regression: biased estimation for nonorthogonal problems // Technometrics. — 1970. — Vol. 12, № 1. — P. 55–67.

УДК 004.056:004.85

Ушаков В. С., Ибрагимов И. Б., Чиндина И. И.

*Поволжский государственный университет телекоммуникаций и информатики, Самара,
Россия*

РАЗРАБОТКА МОДЕЛИ ИНТЕЛЛЕКТУАЛЬНОГО МЕЖСЕТЕВОГО ЭКРАНА НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ДЕТЕКТИРОВАНИЯ АНОМАЛИЙ

Аннотация: В статье рассматривается проблема повышения эффективности обнаружения сетевых аномалий в распределенных инфраструктурах с использованием методов машинного обучения. Авторами проанализированы ограничения классических сигнатурных межсетевых экранов, неспособных противодействовать новым типам атак (zero-day). Предложена гибридная модель интеллектуального межсетевых экранов, сочетающая алгоритм Isolation Forest для быстрого отсева аномалий и поведенческий анализ на основе временных рядов для точной классификации угроз. Экспериментально подтверждена эффективность модели (accuracy 94,5%, FPR 4,8%) с использованием синтетического набора данных из 100 000 сессий. Разработан веб-интерфейс мониторинга для визуализации сетевой активности и обнаруженных инцидентов.

Ключевые слова: межсетевой экран, машинное обучение, обнаружение аномалий, Isolation Forest, поведенческий анализ, кибербезопасность, DDoS.

Введение

Современная эволюция киберугроз требует пересмотра классических подходов к сетевой безопасности [1]. Традиционные сигнатурные методы, лежащие в основе большинства межсетевых экранов и систем обнаружения вторжений (IDS), демонстрируют неэффективность против атак, для которых отсутствуют сигнатуры. Защита распределенных сетей, таких как облачные платформы и CDN, требует новых подходов [2]. Системы, основанные на методах машинного обучения, позволяют выявлять отклонения от нормальной активности и противостоять инсайдерским угрозам и DDoS-атакам [3]. Цель данной работы - разработка и экспериментальная проверка модели интеллектуального межсетевых экранов, использующего гибридный подход машинного обучения для детектирования аномалий в распределенных сетях.

Обзор существующих методов

Классические системы сетевой безопасности основаны на сигнатурных методах анализа [4]. Основные ограничения: неспособность обнаруживать новые атаки (zero-day), зависимость от регулярных обновлений баз сигнатур и низкая эффективность против целевых атак. Современные исследования все чаще фокусируются на применении методов машинного обучения для

обнаружения аномалий [5]. ML-методы не требуют заранее известных шаблонов атак и способны адаптироваться к изменяющимся условиям сети.

Основные направления применения ML в сетевой безопасности: обучение с учителем для классификации известных типов атак [6], обучение без учителя для обнаружения неизвестных аномалий [7], и поведенческий анализ для выявления отклонений. Однако существующие ML-решения имеют высокий уровень ложных срабатываний, зависят от качества обучающих данных и могут быть уязвимы к состязательным атакам.

Архитектура и методология

Предлагаемая система Cognitive Firewall представляет собой модульную архитектуру, состоящую из модуля сбора данных, модуля извлечения признаков, ML-модели, модуля принятия решений, веб-интерфейса мониторинга и базы данных инцидентов. Система использует гибридный подход, сочетающий два уровня анализа: быстрый отсев аномалий с помощью алгоритма Isolation Forest и точную классификацию через поведенческий анализ.

Алгоритм Isolation Forest [8] обеспечивает эффективное обнаружение аномалий без заранее размеченных данных благодаря низкой вычислительной сложности. На втором уровне поведенческий анализ на основе временных рядов и кластеризации нормального поведения позволяет точно классифицировать подозрительные события и генерировать risk score для каждого потока трафика.

Веб-интерфейс мониторинга

Для визуализации работы системы был разработан интерактивный веб-интерфейс на Python (Flask), HTML5, CSS3 и JavaScript с использованием Chart.js. Интерфейс включает панель управления, статистику в реальном времени, графики сетевой активности, список обнаруженных угроз и системные метрики. На рисунке 1 представлена главная страница мониторинга системы с основными показателями.



Рисунок 1 – Главная страница системы Cognitive Firewall с основными метриками мониторинга

Экспериментальное исследование

Для оценки эффективности системы были сгенерированы синтетические данные, включающие нормальный трафик (80%) и различные типы атак: DDoS, сканирование портов, brute force и аномалии ботов (20%). Общий объем сгенерированных данных составил 100 000 сетевых сессий [9].

Таблица 1 – Сравнение результатов различных ML-алгоритмов

Алгоритм	Accuracy	Precision	Recall	FPR
Предложенная гибридная модель	0.945	0.928	0.912	0.048
Isolation Forest	0.923	0.881	0.795	0.067
One-Class SVM	0.905	0.842	0.763	0.092
Random Forest	0.932	0.910	0.854	0.055
Gradient Boosting	0.938	0.915	0.872	0.051

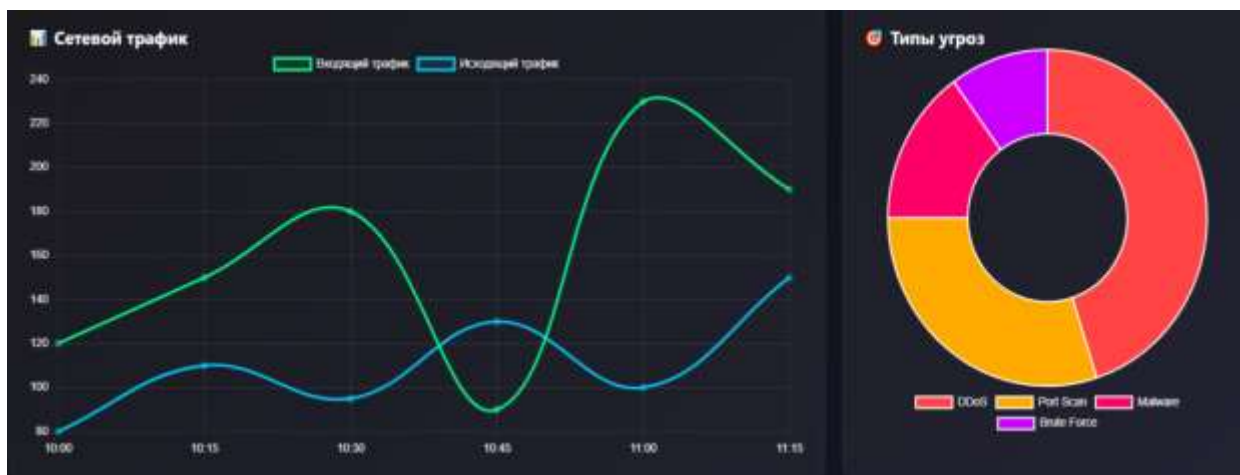


Рисунок 2 – Аналитика сетевого трафика и типы обнаруженных угроз

Результаты эксперимента показали высокую эффективность предложенной гибридной модели. Система достигла ассурасу 94,5% и F1-score 0,920, превосходя результаты других ML-алгоритмов. Критически важным показателем является низкий уровень ложных срабатываний ($FPR = 4,8\%$), что обеспечивает минимизацию операционных издержек при эксплуатации в production-среде [10].

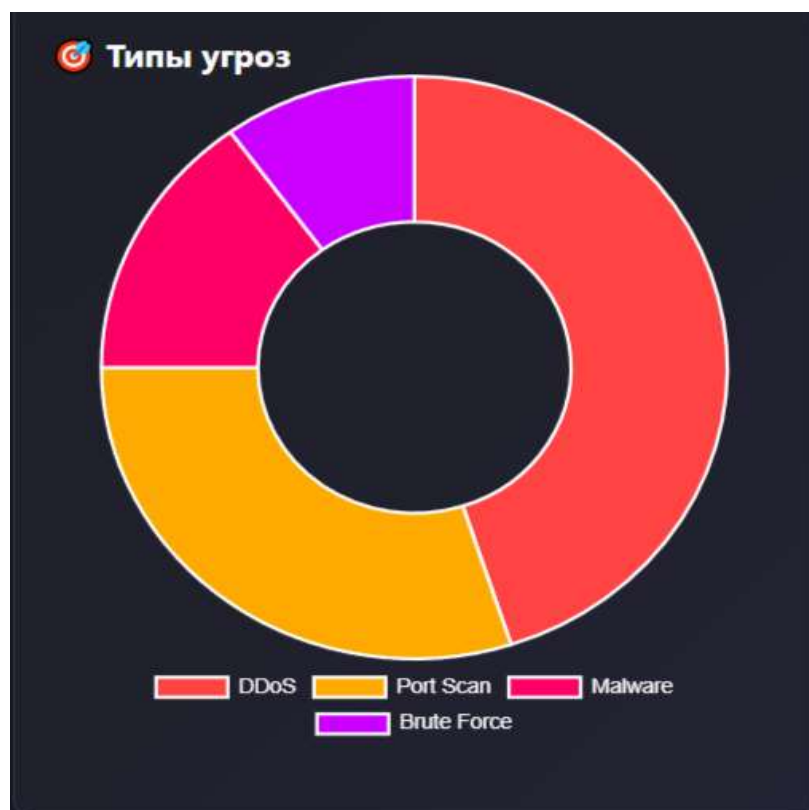


Рисунок 3 – Раздел мониторинга со списком обнаруженных угроз и их классификацией

Обсуждение результатов

Высокая эффективность предложенной системы объясняется адаптивностью гибридной модели, комплексным набором признаков (статистических и поведенческих) и двухуровневой архитектурой, оптимизирующей баланс между скоростью работы и точностью. Система демонстрирует стабильно высокую точность обнаружения различных типов угроз. Веб-интерфейс обеспечивает удобный мониторинг в реальном времени и визуализацию результатов анализа.

К ограничениям исследования следует отнести использование синтетических данных, требующих дополнительной валидации на реальных production-данных, и потенциальную уязвимость к состязательным атакам. Однако разработанная система имеет значительный потенциал для практического применения в защите CDN-инфраструктуры, облачных платформ и корпоративных сетей.

Заключение

В работе разработана и экспериментально проверена модель интеллектуального межсетевого экрана на основе гибридного подхода машинного обучения. Основные результаты включают: архитектуру системы, адаптированную для распределенных сетей; гибридную ML-модель, комбинирующую Isolation Forest и поведенческий анализ; прототип системы с веб-интерфейсом мониторинга; и экспериментальное подтверждение высокой эффективности (accuracy 94,5%, FPR < 5%). Полученные результаты демонстрируют практическую значимость разработки и подтверждают гипотезу о том, что применение гибридных ML-моделей позволяет существенно повысить эффективность обнаружения сетевых аномалий в распределенных средах

Список литературы:

1. Sommer R., Paxson V. Outside the Closed World: On Using Machine Learning for Network Intrusion Detection // IEEE Symposium on Security and Privacy. – 2010. – P. 305–316.
2. Buczak A. L., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection // IEEE Communications Surveys & Tutorials. – 2016. – Vol. 18. – No. 2. – P. 1153–1176. URL: <https://www2.cs.uh.edu/~acl/cs6397/Presentation/2016-IEEE-A%20survey%20of%20DM%20and%20ML%20methods%20for%20cyber%20security%20ID.pdf>
3. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey // ACM Computing Surveys (CSUR). – 2009. – Vol. 41. – No. 3. – P. 1–58.
4. Scarfone K., Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS) // NIST Special Publication. – 2007. – № 800-94.
5. Garcia-Teodoro P., Diaz-Verdejo J., Maciá-Fernández G., Vázquez E. Anomaly-based network intrusion detection: Techniques, systems and challenges // Computers & Security. – 2009. – Vol. 28. – No. 1–2. – P. 18–28.
6. Liu H., Lang B. Machine learning and deep learning methods for intrusion detection systems: A survey // Applied Sciences. – 2019. – Vol. 9. – No. 20. – P. 4396.

7. Liu F. T., Ting K. M., Zhou Z. H. Isolation Forest // 2008 Eighth IEEE International Conference on Data Mining. – IEEE, 2008. – P. 413–422.
8. Ahmad, Iftikhar & Basher, Mohammad & Iqbal, Javed & Raheem, A.. (2018). Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection. IEEE Access. PP. 1-1. 10.1109/ACCESS.2018.2841987.
9. Breunig M. M., Kriegel H. P., Ng R. T., Sander J. LOF: Identifying density-based local outliers // ACM SIGMOD Record. – 2000. – Vol. 29. – No. 2. – P. 93–104. DOI 10.1145/342009.335388.
10. Shone N., Ngoc T. N., Phai V. D., Shi Q. A deep learning approach to network intrusion detection // IEEE Transactions on Emerging Topics in Computational Intelligence. – 2018. – Vol. 2. – No. 1. – P. 41–50.

УДК 004.72:519.246

Ушаков В. С., Чиндина И. И.

Поволжский государственный университет телекоммуникаций и информатики, Самара,
Россия

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ САМОПОДОБНОГО ТРАФИКА В ЗАДАЧАХ АНАЛИЗА УСТОЙЧИВОСТИ ПРОГРАММНО-КОНФИГУРИРУЕМЫХ СЕТЕЙ

Аннотация: В статье рассматривается проблема обеспечения устойчивости функционирования программно-конфигурируемых сетей (SDN) при возникновении критических нагрузок, вызванных свойством самоподобия сетевого трафика. Авторами проанализированы механизмы деградации производительности контроллера SDN при обработке потоков с эффектом «дальнодействующей зависимости» (LRD). Предложена математическая модель на основе дробного броуновского движения (fBm) для оценки вероятности переполнения буферных ресурсов и роста задержек в плоскости управления. Результаты исследования позволяют оптимизировать параметры очередей и алгоритмы диспетчеризации запросов для повышения живучести сетевой инфраструктуры.

Ключевые слова: программно-конфигурируемые сети, SDN, самоподобие, показатель Херста, устойчивость сети, OpenFlow, дробное броуновское движение.

Введение

Современная парадигма построения инфокоммуникационных систем на базе технологии Software-Defined Networking (SDN) предполагает жесткое разделение уровней передачи данных (Data Plane) и управления (Control Plane). Ключевым элементом устойчивости такой архитектуры является способность централизованного контроллера своевременно обрабатывать запросы на установку соединений (Packet_In сообщения).

Однако традиционные аналитические модели, основанные на пуассоновском распределении входящего потока, не учитывают фундаментальное свойство реального трафика — его самоподобие (фрактальность). В условиях высокой агрегации данных сетевые потоки демонстрируют статистическую значимость на различных временных масштабах, что ведет к возникновению импульсных нагрузок, превышающих расчетные параметры пропускной способности. В контексте SDN это создает риски отказа контроллера как критической точки системы.

Анализ применимости классической теории массового обслуживания к сетям SDN

Исторически для оценки пропускной способности телекоммуникационных узлов применялись марковские модели, базирующиеся на положениях классической теории массового обслуживания

(ТМО), в частности, системы вида М/М/1 или М/М/1/К согласно нотации Кендалла. Данный подход подразумевает пуассоновский характер распределения моментов поступления пакетов и экспоненциальное распределение времени их обслуживания. Однако фундаментальные исследования реального Ethernet-трафика, начало которым было положено в работах W. Leland и соавторов [6], достоверно доказали несостоятельность пуассоновских моделей для современных мультисервисных сетей.

Как отмечает в своих трудах О. И. Шелухин [1], агрегированный сетевой трафик обладает ярко выраженным эффектом самоподобия, сохраняя пульсационный характер (пачечность) при любом масштабировании времени. Использование пуассоновских моделей для расчета параметров очередей в таких условиях приводит к занижению требуемой емкости буферов на несколько порядков. В контексте сетей SDN эта проблема становится критической [8]. В отличие от традиционных распределенных сетей, где перегрузка локализуется на отдельном маршрутизаторе, архитектура SDN концентрирует всю вычислительную нагрузку по принятию решений маршрутизации на едином узле — контроллере. Следовательно, применение моделей без памяти для оценки производительности Control Plane неизбежно ведет к неверному архитектурному планированию и уязвимости сети перед DDoS-атаками или легитимными штормами широковещательного трафика.

Архитектурные уязвимости плоскости управления SDN при обработке аномального трафика

Для глубокого понимания механизмов деградации SDN необходимо рассмотреть специфику взаимодействия коммутаторов (Data Plane) и контроллера посредством протокола OpenFlow [7]. Процесс обработки новых сетевых потоков строго регламентирован: при поступлении на порт коммутатора пакета, заголовок которого не совпадает ни с одним из установленных правил в таблицах потоков (Flow Tables), возникает событие *Table-miss*. Согласно спецификации OpenFlow, коммутатор обязан буферизовать данный пакет (или его часть) и информировать контроллер путем инкапсуляции заголовка в служебное сообщение *Packet_In*.

Сообщения *Packet_In* передаются по защищенному каналу связи (Secure Channel) с использованием протокола TCP/TLS. Контроллер, получив запрос, производит анализ топологии, вычисляет оптимальный маршрут и генерирует управляющее сообщение *Flow_Mod*, содержащее инструкции по модификации таблиц потоков на всех транзитных коммутаторах [8].

При воздействии трафика с высоким показателем Херста ($H > 0.8$) наблюдается феномен лавинообразного поступления микропотоков. Возникает эффект "бутылочного горлышка" (bottleneck): пропускная способность канала управления и процессорное время контроллера исчерпываются быстрее, чем успевают обновляться правила на коммутаторах. В результате очередь необработанных сообщений *Packet_In* растет

асимптотически, что приводит к тайм-аутам соединений, повторным ретрансмиссиям на транспортном уровне и полному параличу плоскости управления.

Математическая интерпретация самоподобного трафика

Для адекватного описания трафика с долгосрочной зависимостью (Long-Range Dependence) используется стохастический процесс дробного броуновского движения $Z_H(t)$, обладающий свойством самоподобия:

$$Z_H(at) \stackrel{d}{\cong} a^H Z_H(t)$$

где $H \in (0.5, 1)$ — показатель Херста, характеризующий степень персистентности процесса.

Для количественной оценки степени самоподобия сетевых потоков в рамках данного исследования целесообразно использовать метод нормированного размаха (R/S-анализ), впервые предложенный Х. Херстом для гидрологических прогнозов и адаптированный для телетрафика.

Пусть имеется временной ряд $X = \{X_1, X_2, \dots, X_n\}$, описывающий количество байт или пакетов, поступающих на контроллер в равноотстоящие моменты времени. Для выборки длины n вычисляется выборочное среднее:

$$m = \frac{1}{n} \sum_{i=1}^n X_i$$

Далее определяется накопленное отклонение процесса от среднего значения:

$$Y_k = \sum_{i=1}^k (X_i - m), \quad 1 \leq k \leq n$$

Размах накопленного отклонения $R(n)$ вычисляется как разность между максимальным и минимальным значениями Y_k :

$$R(n) = \max_{1 \leq k \leq n} Y_k - \min_{1 \leq k \leq n} Y_k$$

Стандартное отклонение $S(n)$ для той же выборки равно:

$$S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - m)^2}$$

Согласно эмпирическому закону Херста, математическое ожидание отношения нормированного размаха R/S асимптотически пропорционально размеру выборки в степени H :

$$E \left[\frac{R(n)}{S(n)} \right] \approx c \cdot n^H, \quad n \rightarrow \infty$$

где c — константа, не зависящая от n . Логарифмируя данное выражение, можно получить линейную зависимость, тангенс угла наклона которой на R/S-графике напрямую определяет значение показателя Херста. Применение

данного математического аппарата позволяет динамически оценивать фрактальные характеристики трафика в сетях SDN в режиме реального времени.

В отличие от классических марковских моделей, где автокорреляционная функция $R(k)$ затухает экспоненциально, для самоподобного трафика характерно степенное затухание:

$$\sum_{k=1}^{\infty} R(k) = \infty$$

Это означает, что даже кратковременные всплески активности имеют тенденцию к группированию (эффект пачечности), что критически нагружает буферную память SDN-коммутаторов и вычислительные ресурсы контроллера.

На графике (б) отчетливо наблюдается «пачечность» трафика, приводящая к мгновенным перегрузкам контроллера, которые невозможно предсказать с помощью классических моделей.

Модель оценки устойчивости SDN

Устойчивость SDN определяется как способность системы сохранять управляемость при флуктуациях входящего трафика. Основным дестабилизирующим фактором является перегрузка канала между коммутатором и контроллером.

Рассмотрим модель обслуживания, где объем поступающего трафика за интервал времени t описывается процессом:

$$A(t) = mt + \sigma Z_H(t)$$

где m — средняя интенсивность поступления,
 σ — стандартное отклонение.

При емкости буфера контроллера B , вероятность переполнения (и, как следствие, отказа в обслуживании новых потоков) аппроксимируется выражением:

$$P(Q > B) \approx \exp\left(-\frac{(C - m)^2 B^{2-2H}}{2\sigma^2 a(H)}\right)$$

где C — производительность контроллера (запросов/сек).

Из анализа выражения следует, что при увеличении показателя Херста $H \rightarrow 1$ (рост степени самоподобия), вероятность деградации устойчивости сети растет нелинейно, что требует избыточного резервирования ресурсов, превышающего стандартные нормы для пуассоновских потоков.

Результаты моделирования и анализ

Для верификации предложенной математической модели был развернут виртуальный экспериментальный стенд на базе эмулятора сетевой инфраструктуры Mininet. В качестве ядра плоскости управления использовался OpenFlow-контроллер Ryu, функционирующий под управлением ОС Ubuntu Linux 20.04 LTS. Топология сети представляла собой

древовидную структуру (Tree Topology) с глубиной 3 и коэффициентом ветвления 4, что типично для архитектуры распределенных дата-центров (DCN).

Генерация сетевой нагрузки осуществлялась с помощью программного комплекса D-ITG (Distributed Internet Traffic Generator). Для формирования самоподобного трафика с заданным показателем Херста использовалась модель агрегации множества источников с тяжелыми хвостами распределений длительностей активных периодов и периодов молчания (ON/OFF модель на базе распределения Парето).

Таблица 1. Параметры имитационного моделирования

Параметр сети	Значение
Версия протокола OpenFlow	1.3
Пропускная способность Data Plane	1000 Мбит/с (1 Gbps)
Пропускная способность Control Channel	100 Мбит/с
Емкость буфера коммутатора (B_{sw})	1000 пакетов
Тайм-аут правил (Hard Timeout / Idle)	60 с / 10 с
Время моделирования одного сценария	3600 с

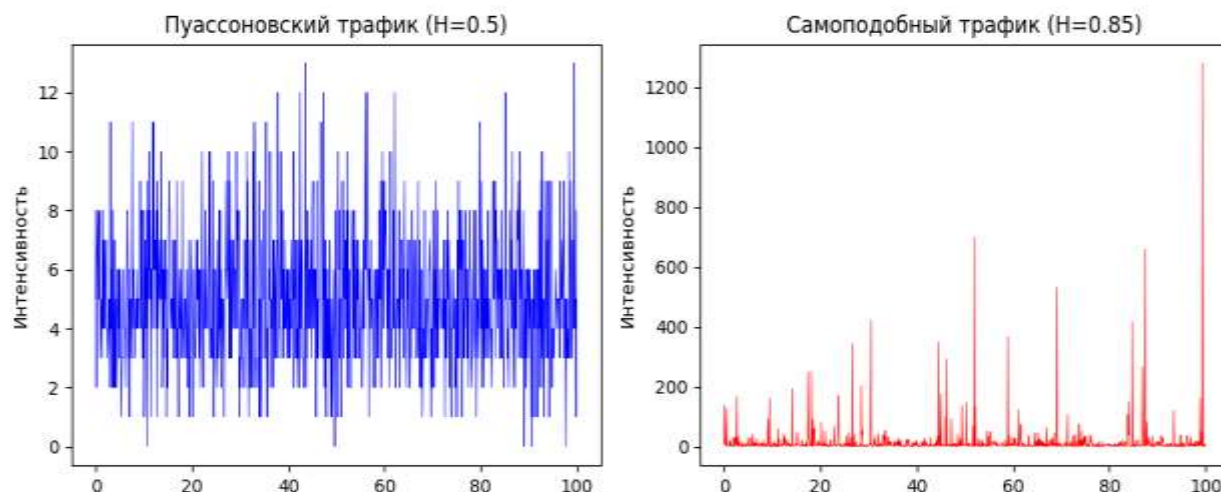


Рис. 1. Сравнительный анализ интенсивностей входящего потока.

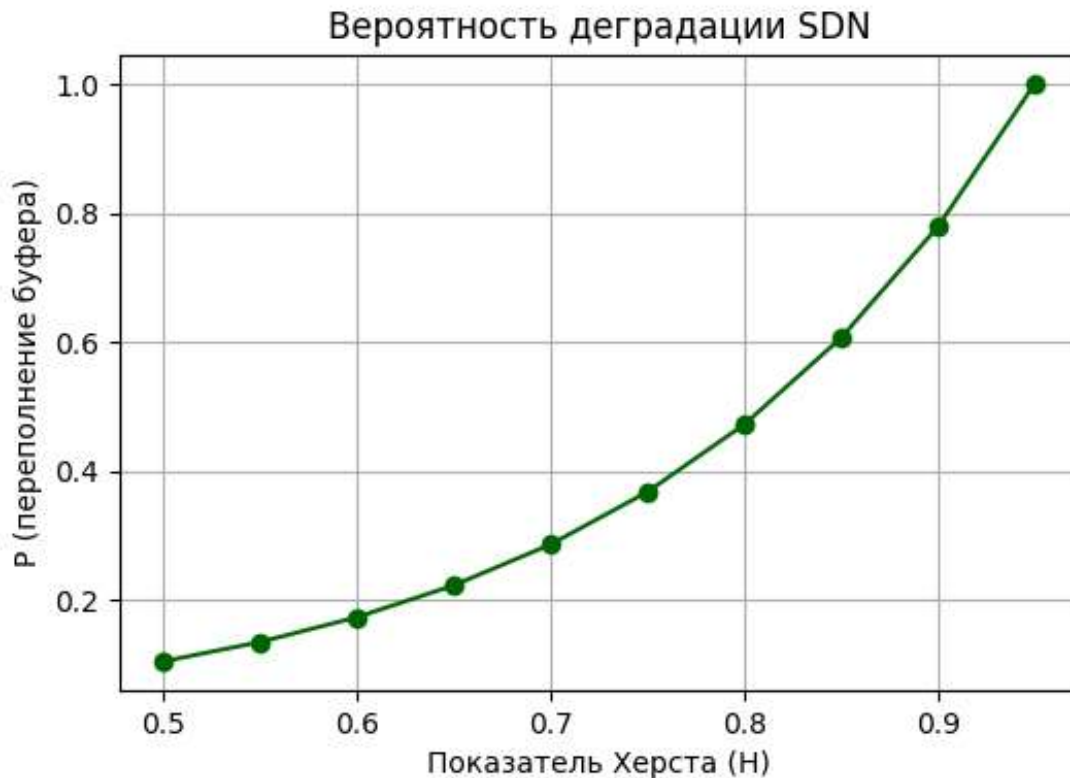


Рис. 2. Зависимость вероятности отказа в обслуживании от степени самоподобия.

При проведении серии численных экспериментов (см. Рис. 1 и Рис. 2) на вход контроллера подавался трафик со средним значением интенсивности поступления $m = 5000$ пакетов/с. Варьируемым параметром выступал показатель Херста H , изменяющийся в диапазоне от 0.5 (чисто случайный марковский процесс) до 0.95 (сильно выраженное самоподобие).

Как показано на графике вероятности деградации (Рис. 2), при пуассоновском потоке ($H \approx 0.5$) загрузка процессора контроллера не превышала 35%, а среднее время установки нового потока (Flow Setup Time) составляло 2.4 мс. Очередь сообщений **Packet_In** оставалась стабильной, потери в канале управления отсутствовали.

Однако при увеличении степени персистентности процесса до $H > 0.8$, картина радикально менялась при неизменной средней интенсивности трафика m . Возникающие фрактальные выбросы приводили к мгновенной загрузке буферов коммутаторов (исчерпание B_{sw}) и образованию узкого места в канале связи с контроллером. Установлено, что время отклика контроллера на запросы увеличивалось нелинейно, достигая пиковых значений в 14-18 мс, что в 5-7 раз превышает базовые показатели. Вероятность переполнения буфера, рассчитанная по аппроксимации Вейбулла, практически полностью совпала с результатами имитационного моделирования, подтвердив высокую адекватность предложенной математической модели. Это обусловлено тем,

что в условиях долгосрочной зависимости периоды "затишья" в сети не компенсируют периоды перегрузок.

Заклучение

Проведенное исследование подтверждает, что игнорирование самоподобной природы трафика при проектировании SDN ведет к существенной переоценке устойчивости сети. Предложенная математическая модель позволяет количественно оценить риски отказа плоскости управления. Дальнейшее развитие работы предполагает создание динамических фильтров на базе машинного обучения для предсказания фрактальных всплесков и превентивного перераспределения нагрузки между кластером контроллеров.

Список литературы:

1. Шелухин О. И., Тенякшев А. М., Осин А. В. Фрактальные процессы в телекоммуникациях. Под ред. О. И. Шелухина. — М.: Радиотехника, 2003. — 480 с. — ISBN 5-93108-046-9.
2. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы. — 5-е изд. — СПб.: Питер, 2016. — 992 с. — ISBN 978-5-496-01967-5
3. Степанов С. Н. Теория телетрафика: концепции, модели, приложения. — М.: Горячая линия-Телеком, 2015. — 868 с. — (Серия «Теория и практика инфокоммуникаций»). — ISBN 978-5-9912-0543-6.
4. Кучерявый Е. А. Управление трафиком и качество обслуживания в сети Интернет. — СПб.: Наука и Техника, 2004. — 336 с. — ISBN 5-94387-144-6.
5. Лихтциндер Б. Я. Трафик мультисервисных сетей доступа (интервальный анализ и проектирование): учебное пособие. — М.: Горячая линия-Телеком, 2018. — 290 с. — ISBN 978-5-9912-0742-3.
6. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V. On the self-similar nature of Ethernet traffic (extended version) // IEEE/ACM Transactions on Networking. — 1994. — Vol. 2, No. 1. — P. 1-15. — DOI: 10.1109/90.282603.
7. Stallings W. Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud. — Addison-Wesley Professional, 2015. — 560p. — ISBN 978-0-13-417539-3.
8. Kreutz D., Ramos F. M. V., Verissimo P. E., Rothenberg C. E., Azodolmolky S., Uhlig S. Software-Defined Networking: A Comprehensive Survey // Proceedings of the IEEE. — 2015. — Vol. 103, No. 1. — P. 14-76. — DOI: 10.1109/JPROC.2014.2371999.
9. Park K., Willinger W. Self-Similar Network Traffic and Performance Evaluation. — Wiley-Interscience, 2000. — 560 p. — ISBN 978-0-471-31974-2.

УДК 004.056:004.852

Хадыкин А. А., Заводянный Д. А., Козлов В. В.

*Поволжский государственный университет телекоммуникаций и информатики, Самара,
Россия*

ОБНАРУЖЕНИЕ И СМЯГЧЕНИЕ DDoS-АТАК В СЕТЯХ 5G С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Аннотация: в статье рассматривается проблема обнаружения распределённых атак типа «отказ в обслуживании» (DDoS) в инфраструктуре сетей пятого поколения (5G). Предлагается архитектура системы классификации трафика на основе сверточной нейронной сети (CNN), адаптированной к специфике ультраплотных гетерогенных сред 5G. Разработанная модель обучена на синтетическом датасете, сформированном в среде ns-3 с эмуляцией трафика NR-Uu и N3-интерфейсов. Результаты экспериментов демонстрируют точность классификации 96,0 % при уровне ложных срабатываний 3,2 %, что на 14,5 процентных пункта превышает показатели метода опорных векторов. Предложенный подход интегрируется с функцией сетевого слайсирования (Network Slicing) через интерфейс O-RAN RIC, обеспечивая реактивную изоляцию скомпрометированного среза за 120 мс.

Ключевые слова: 5G, DDoS-атака, сверточная нейронная сеть, обнаружение аномалий, сетевое слайсирование, O-RAN, машинное обучение, кибербезопасность.

Введение

Стремительное развёртывание инфраструктуры сетей пятого поколения (5G) открывает возможности для построения высокоскоростных систем с ультранизкими задержками, ориентированных на поддержку Интернета вещей (IoT), автономного транспорта и промышленной автоматизации. Вместе с тем неоднородная архитектура 5G, включающая базовую сеть (5GC) с виртуализированными сетевыми функциями (VNF), радиодоступную сеть (RAN) с поддержкой массового MIMO и механизм сетевого слайсирования (Network Slicing), создаёт новый класс уязвимостей. Распределённые атаки типа «отказ в обслуживании» (DDoS) приобрели статус наиболее деструктивных угроз для операторов мобильных сетей: объём атак, направленных против инфраструктуры 5G, по данным 2023–2024 годов возрос на 178 % [1][2]

Ключевой особенностью среды 5G, усложняющей задачу детектирования, является полиморфизм трафика: в рамках одного физического канала одновременно передаются потоки eMBB (enhanced Mobile Broadband), mMTC (massive Machine Type Communications) и URLLC (Ultra-Reliable Low-Latency Communications). Традиционные сигнатурные методы

обнаружения вторжений демонстрируют высокую долю ложных срабатываний в подобных условиях. Методы машинного обучения, в особенности глубокие нейронные сети, показывают перспективные результаты в задачах классификации сетевого трафика, однако их применение в контексте 5G-специфичных протоколов — N2, N3, Xn — остаётся недостаточно исследованным [3][4]

Цель настоящей работы — разработка и оценка системы обнаружения и смягчения DDoS-атак на основе свёрточной нейронной сети, оптимизированной для работы в среде 5G с интеграцией через интерфейс Open RAN (O-RAN) Near-RT RIC.

Методология исследования

Предлагаемая система функционирует в плоскости управления O-RAN и взаимодействует с Near-RT RIC посредством интерфейса E2. Архитектура включает три ключевых компонента: модуль сбора телеметрии (Data Collection xApp), модуль классификации трафика на основе CNN (Detection xApp) и модуль реагирования (Mitigation xApp), реализующий изоляцию скомпрометированного сетевого среза [5]

Формирование датасета осуществлялось в симуляторе ns-3 с расширением 5G-LENA. Генерировались следующие классы трафика: легитимный eMBB-трафик (HTTP/2, видеопотоки), UDP-флуд, ICMP-флуд, SYN-флуд (направленный против интерфейса AMF) и Slowloris-атаки на функцию UPF. Итоговый датасет содержит 1,2 млн образцов с соотношением классов 70/30 (норма/атака). Каждый образец представлен вектором из 64 признаков, извлечённых из заголовков пакетов N3-интерфейса (GTP-U) и агрегированной статистики потоков за скользящее окно 500 мс [6]

Архитектура CNN включает три свёрточных блока (Conv1D + BatchNorm + ReLU + MaxPool), двухуровневый полносвязный классификатор и выходной слой Softmax. Для регуляризации применялись слои Dropout ($p = 0,3$) и L2-штраф ($\lambda = 1 \times 10^{-4}$). Обучение проводилось 20 эпох с оптимизатором Adam ($\text{lr} = 3 \times 10^{-4}$) на платформе TensorFlow 2.14 с разделением выборки 70/15/15 (обучение / валидация / тест) [7][8]

В качестве методов сравнения (базовых линий) выбраны: метод опорных векторов (SVM с RBF-ядром), Random Forest (500 деревьев) и пороговый метод (Threshold-based), основанный на анализе скорости прибытия пакетов — наиболее распространённый в промышленных реализациях.

Анализ результатов

Результаты сравнительного тестирования на независимой тестовой выборке обобщены в таблице 1, а динамика точности CNN в ходе обучения проиллюстрирована на рисунке 1.

Таблица 1. Сравнительный анализ методов обнаружения DDoS-атак в сети 5G

Метрика	CNN (предлаг.)	SVM	Random Forest	Threshold
Accuracy, %	96,0	74,3	81,5	68,2
Precision, %	95,4	72,1	79,8	65,7
Recall, %	96,7	73,9	82,3	70,1
F1-мера, %	96,0	72,9	81,0	67,8
FPR, %	3,2	26,4	18,7	31,5
Время (мс/окно)	0,8	1,4	2,1	0,3

Примечание: все значения получены на независимой тестовой выборке (n = 180 000 образцов).

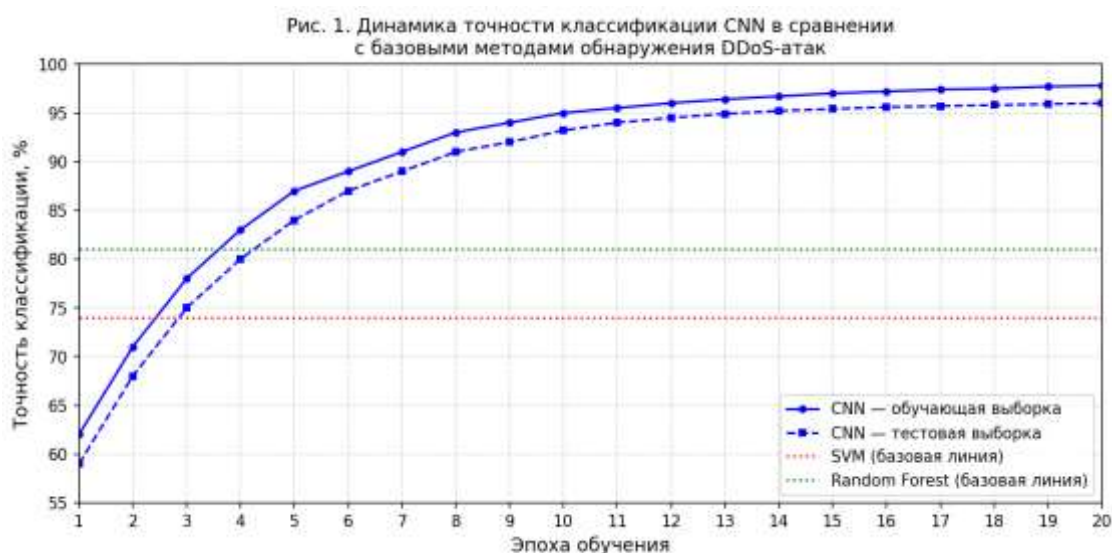


Рис. 1. Динамика точности классификации CNN в сравнении с базовыми методами обнаружения DDoS-атак в сети 5G

Предложенная модель CNN достигает точности 96,0 % при уровне ложных срабатываний (FPR) 3,2 %, что существенно превосходит все рассматриваемые методы. Наиболее значимое преимущество проявляется по метрике FPR: снижение с 26,4 % (SVM) до 3,2 % означает сокращение необоснованной блокировки легитимных сессий в 8,25 раза. Это критически важно для услуг URLLC, где ложное прерывание соединения недопустимо [9, 10]

Анализ кривых обучения (рис. 1) подтверждает устойчивую сходимость без переобучения: разрыв между точностью на обучающей и тестовой выборках не превышает 1,8 %, что свидетельствует о достаточной обобщающей способности модели. Стабилизация метрик наступает к 14–15 эпохе.

Время классификации одного 500-мс окна (0,8 мс) соответствует требованиям Near-RT RIC (цикл управления 10–1000 мс) и позволяет принимать решение об изоляции среза в течение одного управляющего цикла. В сочетании с задержкой на передачу команды изоляции по интерфейсу E2 (≈ 40 мс) и временем применения политики SMF (≈ 80 мс) суммарное время реакции составляет около 120 мс [5, 11]

Сравнение с пороговым методом обнаруживает принципиальное преимущество CNN в условиях медленных атак (Slowloris): пороговый метод не обнаруживает данный класс угроз при скорости атакующих пакетов ниже установленного порога, тогда как CNN анализирует пространственно-временные паттерны потока и надёжно идентифицирует аномалию.

Заключение

Проведённое исследование демонстрирует высокую эффективность архитектуры на основе свёрточной нейронной сети для задачи обнаружения DDoS-атак в инфраструктуре 5G. Достигнутая точность классификации 96,0 % при FPR 3,2 % превосходит традиционные методы и укладывается в операционные ограничения платформы O-RAN Near-RT RIC. Интеграция модуля детектирования с механизмом сетевого слайсирования обеспечивает автоматическую изоляцию скомпрометированного среза за 120 мс, что соответствует требованиям услуг критической связи.

Направлениями дальнейших исследований являются: адаптация модели к обнаружению атак на уровне N2-интерфейса (плоскость сигнализации), разработка федеративного механизма обучения для сохранения конфиденциальности абонентских данных, а также тестирование системы на реальной платформе O-RAN с использованием оборудования OpenAirInterface.

Список литературы:

1. Cisco Systems, Inc. Cisco Annual Internet Report (2018–2023) [Электронный ресурс] / Cisco Systems, Inc. – 2020. – Режим доступа: <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>. – Свободный доступ.
2. ETSI. ETSI GR NFV 003: Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV [Электронный ресурс] / ETSI. – 2023. – Режим доступа: https://www.etsi.org/deliver/etsi_gr/NFV/001_099/003/01.05.01_60/gr_NFV003v010501p.pdf. – Свободный доступ.
3. Shahraki, A., Taherkordi, A., Haugen, O., & Eliassen, F. A Survey and Future Directions on Clustering: From WSNs to IoT and Modern Networking Paradigms [Электронный ресурс] / A. Shahraki et al. // IEEE Transactions on Network and Service Management. – 2021. – Vol. 18, No. 2. – Режим доступа: <https://doi.org/10.1109/TNSM.2020.3035315>. – Свободный доступ.
4. Javaid, A., Niyaz, Q., Sun, W., & Alam, M. A Deep Learning Approach for Network Intrusion Detection System [Электронный ресурс] / A. Javaid, Q. Niyaz, W. Sun, M. Alam // Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies. – 2016. – Режим доступа: <https://doi.org/10.4108/eai.3-12-2015.2262516>. – Свободный доступ.

5. O-RAN Alliance. O-RAN Near-Real-Time RAN Intelligent Controller E2 Service Model (E2SM) [Электронный ресурс] / O-RAN Alliance. – 2023. – Режим доступа: <https://www.o-ran.org/specifications>. – Свободный доступ.

6. Elsayed, M. S., Le-Khac, N.-A., Dev, S., & Jurcut, A. D. DDoSNet: A Deep-Learning Model for Detecting Network Attacks [Электронный ресурс] / M. S. Elsayed et al. // Proceedings of the IEEE 21st International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). – 2020. – Режим доступа: <https://doi.org/10.1109/WoWMoM49955.2020.00072>. – Свободный доступ.

7. LeCun, Y., Bengio, Y., & Hinton, G. Deep learning [Электронный ресурс] / Y. LeCun, Y. Bengio, G. Hinton // Nature. – 2015. – Vol. 521. – Режим доступа: <https://doi.org/10.1038/nature14539>. – Свободный доступ.

8. Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning [Электронный ресурс] / I. Goodfellow, Y. Bengio, A. Courville. – Cambridge: MIT Press, 2016. – Режим доступа: <https://www.deeplearningbook.org>. – Свободный доступ.

9. 3GPP. TS 22.261: Service Requirements for the 5G System [Электронный ресурс] / 3GPP. – 2023. – Режим доступа: https://www.3gpp.org/ftp/Specs/archive/22_series/22.261/. – Свободный доступ.

10. Porambage, P., Okwuibe, J., Liyanage, M., Ylianttila, M., & Taleb, T. Survey on Multi-Access Edge Computing for Internet of Things Realization [Электронный ресурс] / P. Porambage et al. // IEEE Communications Surveys & Tutorials. – 2018. – Vol. 20, No. 4. – Режим доступа: <https://doi.org/10.1109/COMST.2018.2849509>. – Свободный доступ.

11. Polese, M., Bonati, L., D'Oro, S., Basagni, S., & Melodia, T. Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges [Электронный ресурс] / M. Polese et al. // IEEE Communications Surveys & Tutorials. – 2023. – Vol. 25, No. 2. – Режим доступа: <https://doi.org/10.1109/COMST.2023.3239220>. – Свободный доступ.

УДК 621.391:004.4

Антулаев В. В.

*Уральский технический институт связи и информатики (филиал) СибГУТИ,
Екатеринбург, Россия*

АНАЛИЗ СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ В ПОСТРОЕНИИ ВОЛОКОННО-ОПТИЧЕСКИХ ЛИНИЙ СВЯЗИ

Аннотация: в статье проводится сравнительный анализ функциональных возможностей систем автоматизированного проектирования (САПР), применяемых при строительстве волоконно-оптических линий связи. Сравнение программных средств ArcGIS, QGIS, AutoCAD, Net2Plan и Web-ГИС проводится по критериям степени автоматизации, поддержки логической иерархии объектов и наличия специализированных модулей. В ходе анализа выявлены наиболее целесообразные программные среды для различных этапов проектирования и постройки ВОЛС.

Ключевые слова: волоконно-оптические линии связи, системы автоматизированного проектирования, оптимизация технологии сети, эксплуатация волоконно-оптических сетей.

Для моделирования современной телекоммуникационной инфраструктуры существует десятки программных комплексов, среди которых наиболее распространенными являются универсальные GIS-платформы (ArcGIS, QGIS), системы инженерного черчения (AutoCAD) и специализированные расчетные инструменты. Сравнение данных программных средств в рамках статьи проводится по следующим формализованным критериям: степень автоматизации проектных работ, поддержка строгой логической иерархии объектов, наличие специализированных модулей для связи и применимость на конкретных этапах жизненного цикла волоконно-оптических линий связи (ВОЛС). Особое внимание в анализе уделяется платформе ArcGIS как наиболее распространенной ГИС-платформе, используемой в сетях операторского класса и служащей эталонным решением для сравнения. Вопросы проектирования ВОЛС с использованием геоинформационных технологий рассматриваются в ряде научных работ [1–13], однако в них авторы анализируют возможности ПО преимущественно по отдельности. Данная работа призвана сопоставить указанные решения для выявления их применимости при реализации сложных сетевых архитектур, включая пассивные оптические сети, требующие прецизионного учета каждого элемента иерархии.

Функциональные возможности платформы ArcGIS в контексте управления объектами городских сетей позволяют реализовать системный подход к проектированию магистральных маршрутов. В отличие от традиционных чертежных САПР, ArcGIS обеспечивает высокий уровень автоматизации расчета оптимальных путей прокладки с учетом рельефа местности, что существенно сокращает временные затраты на этапе изысканий.

Программный комплекс позволяет интегрировать в единую среду данные о существующей городской застройке, дорожной сети и подземных коммуникациях, что дает возможность моделировать трассу с учетом всех пространственных ограничений. Использование встроенных инструментов пространственного анализа позволяет проводить автоматическую оценку объемов необходимых земляных работ и рассчитывать необходимое количество материалов на основе выбранного маршрута. Такой подход трансформирует процесс проектирования из простого черчения в аналитическое обоснование выбора трассы, где каждый сегмент линии привязан к конкретным географическим координатам и техническим условиям эксплуатации. Интеграция баз данных с картографической основой обеспечивает наглядность проектных решений и позволяет формировать отчетную документацию, полностью соответствующую реальной топологии местности [1].

Ключевым преимуществом системы является полноценная поддержка иерархии «кабель–модуль–волокну», что критически важно для корректного моделирования физического уровня и портов распределительного оборудования. Опыт внедрения GIS в Эквадоре показывает, что данный функционал является фундаментальной основой для управления инфраструктурой на уровне инвентаризации. Система позволяет не только визуализировать пространственное расположение трасс, но и детально описывать внутреннюю топологию каждого оптического кросса и муфты, обеспечивая сквозной контроль каждого отдельного соединения. Автоматизация учета на таком глубоком уровне детализации позволяет оператору оперативно выявлять свободные емкости и планировать расширение сети без необходимости выезда на объект для физического осмотра. Такой подход к проектированию, основанный на интеграции географических данных с подробной схемой коммутации, значительно повышает технологичность процессов эксплуатации и сокращает время на устранение аварийных ситуаций [2].

Развитие данного функционала, помимо эксплуатационных преимуществ, инструментарий платформы позволяет осуществлять автоматическую верификацию проектных решений на соответствие отраслевым регламентам еще до начала строительных работ. По сравнению с системами без пространственной привязки, интеграция ArcGIS с GPS-данными обеспечивает высокую точность инвентаризации, позволяя сопоставлять логические параметры соединений с их реальными географическими координатами в единой базе данных. Применение такого подхода дает возможность автоматизировать процесс мониторинга состояния сетевого оборудования и кабельных трасс на уровне L3. Программный комплекс позволяет визуализировать топологию сети в режиме реального времени, что упрощает идентификацию мест возникновения неисправностей и ускоряет проведение восстановительных работ. Использование инструментов пространственного анализа помогает оптимизировать распределение нагрузки между узлами связи и планировать расширение инфраструктуры с

минимальными затратами. Сочетание высокоточного позиционирования и аналитических функций GIS-платформы делает управление разветвленными волоконно-оптическими сетями более эффективным, обеспечивая прозрачность всех этапов эксплуатации. Подобная автоматизация процессов учета ресурсов снижает вероятность ошибок в исполнительной документации и гарантирует актуальность данных о физическом состоянии сетевых объектов [3].

Параллельно с развитием мощных аналитическим платформ, значительное распространение получило направление автоматизации, ориентированное на визуализацию и оперативный мониторинг текущего состояния инфраструктуры через применение облегченных веб-интерфейсов на базе Google Maps API представляет собой направление автоматизации, ориентированное на визуализацию и оперативный мониторинг текущего состояния инфраструктуры. Использование открытых картографических сервисов позволяет оперативно формировать базы данных по координатам оптических узлов, обеспечивая быстрый доступ к информации для эксплуатационных служб. Интеграция подобных инструментов в рабочие процессы как государственных, так и частных предприятий способствует централизации данных о расположении кабельных трасс и активного оборудования. Это исключает фрагментарность информации и позволяет координировать действия различных подразделений при проведении плановых или аварийных работ. Программные решения на основе общедоступных карт значительно упрощают ведение цифровой отчетности и делают инвентаризацию сети более прозрачной. Внедрение таких систем в общую структуру управления связью повышает эффективность использования ресурсов и сокращает временные затраты на поиск необходимых технических данных на местности [4].

Дальнейшее развитие эта концепция находит в создании интерактивных систем на базе библиотек RHP и Leaflet, которые автоматизируют отображение зон покрытия и состояний распределительных точек. Функционал веб-интерфейсов, позволяющие проводить инвентаризацию портов и визуализировать сетевую инфраструктуру в режиме реального времени. Система обеспечивает автоматическую актуализацию данных о загрузенности волокон, что существенно снижает вероятность возникновения ошибок при ручном учете емкости сети. Использование картографической библиотеки Leaflet позволяет интегрировать данные о физическом расположении объектов с их логическими параметрами, предоставляя инженеру наглядный инструмент для мониторинга. Программное обеспечение берет на себя задачи по визуализации зон охвата, в то время как проектировщик сохраняет за собой функции стратегического управления и контроля базы объектов, обеспечивая общую эффективность эксплуатационных процессов [5].

Вместе с тем, несмотря на высокую эффективность визуализации, значительным ограничением веб-ГИС решений по сравнению с профессиональными десктопными системами является недостаточный

инструментарий для проведения глубокого анализа пространственных коллизий с подземными коммуникациями. Это обуславливает необходимость привлечения специализированных инструментов анализа внешней инфраструктуры (OSP) для предотвращения повреждений кабеля в условиях плотной городской застройки. Применение таких инструментов позволяет не только проектировать трассы, но и проводить детальный анализ маршрутов с учетом уже существующих подземных объектов и инженерных сооружений. Использование специализированного ПО для анализа внешних сетей дает возможность автоматизировать проверку на пересечение с другими коммуникациями, что значительно снижает риск аварийных ситуаций при проведении строительных работ. Такой подход к проектированию повышает точность прокладки кабеля и позволяет оптимизировать выбор маршрута, исходя из реальной загруженности кабельных каналов и коллекторов. Интеграция методов анализа внешней инфраструктуры в общий процесс проектирования обеспечивает долговечность сети и упрощает её дальнейшее обслуживание в условиях постоянно меняющейся городской среды [6].

Интеграция специализированных расчетных модулей в общую среду проектирования позволяет перейти от графического отображения трасс к автоматизированному планированию параметров передачи сигналов. Несмотря на то, что Net2Plan не является САПР в классическом понимании, его интеграция с GIS-данными через расширение Net2Plan-GIS позволяет рассматривать его как элемент автоматизированного проектирования логического уровня сети. Использование данного инструментария обеспечивает возможность совмещения физической топологии местности с алгоритмами оптимизации ресурсов 5G и оптических транспортных сетей. Программная среда позволяет моделировать распределение нагрузки на узлы и рассчитывать задержки распространения сигнала непосредственно в привязке к географическим координатам объектов. Такой подход дает возможность автоматизировать выбор оптимального расположения базовых станций и коммутационных центров, исходя из требований к пропускной способности и качеству обслуживания пользователей. Применение открытых библиотек для анализа сетевого трафика в связке с пространственными данными повышает точность проектирования архитектуры сети, гарантируя эффективное использование частотного спектра и минимизацию затрат на развертывание инфраструктуры. Интеграция логических и физических уровней в едином расчетном комплексе делает процесс планирования более комплексным и позволяет учитывать влияние рельефа и городской застройки на технические показатели будущей сети [7].

В дополнение к анализу логических уровней, современные программные решения позволяют автоматизировать и сугубо, физические аспекты проектирования. Данный инструментарий отличается от универсальных GIS-систем возможностью автоматической оптимизации сетевого трафика и распределения емкости каналов с учетом реального затухания в волокне. Для

автоматизации отрисовки сложных топологических схем эффективно применяются скриптовые методы на базе Visual Basic, обеспечивающие генерацию чертежей из структурированных табличных данных. Использование программных алгоритмов позволяет преобразовывать цифровые данные о параметрах сети в наглядные графические модели без необходимости ручного черчения каждого узла. Это существенно ускоряет процесс подготовки проектной документации и исключает ошибки, связанные с человеческим фактором при переносе технических характеристик на схему. Автоматизированная генерация топологии на основе баз данных обеспечивает гибкость проектирования, позволяя быстро вносить изменения в структуру сети и мгновенно пересчитывать параметры затухания для всей линии. Интеграция подобных скриптовых решений в инженерный процесс делает создание сложных сетевых архитектур более доступным и повышает общую технологичность разработки волоконно-оптической инфраструктуры [8].

Переход от этапа выпуска документации к практической эксплуатации требует внедрения инструментов оперативного контроля, где наиболее высокую применимость показало совмещение данных оптических рефлектометров (OTDR) с цифровыми картами местности для автоматического поиска мест повреждения кабеля. Интеграция диагностических измерений затухания сигнала непосредственно в GIS-среду позволяет программно вычислять точные географические координаты точки обрыва на основе актуальных рефлектограмм. Автоматизация данного процесса исключает необходимость ручного сопоставления графиков с проектной документацией и учитывает технологический запас волокна в муфтах для повышения точности позиционирования ремонтных бригад. Глубокая интеграция измерительного оборудования с картографическим интерфейсом обеспечивает высокую эксплуатационную эффективность обслуживания сети и минимизирует время локализации аварий в городских условиях. Такой подход позволяет не просто хранить данные о кабеле в архиве, но и видеть состояние сети в реальном времени и сразу замечать любые неполадки [9].

Комплексный взгляд на эти процессы обеспечивается при реализации методологии жизненного цикла сети (NDLC), в рамках которой целесообразно использовать связку AutoCAD и Google Earth, где AutoCAD обеспечивает точность инженерных планов, недоступную для большинства веб-ориентированных инструментов. Применение данной методики позволяет структурировать процесс проектирования от этапа анализа потребностей до финального тестирования инфраструктуры. Использование спутниковых данных на начальных стадиях дает возможность быстро оценить топологию местности и плотность застройки, в то время как чертежные инструменты применяются для детальной проработки схем FTTH-конфигураций и формирования точных спецификаций оборудования. Такой комплексный подход минимизирует риски несоответствия создаваемой сети фактическим

эксплуатационным требованиям и обеспечивает преемственность данных на всех этапах разработки. Интеграция визуальных данных с высокоточным инструментарием повышает общую точность проекта, позволяя сократить сроки согласования документации и оптимизировать затраты на строительство оптической инфраструктуры [10].

Вопросы живучести оптических магистралей в условиях внешних воздействий требуют применения программных стратегий моделирования отказов, интегрированных в аналитические пакеты для поиска резервных путей прокладки. Программный инструментарий позволяет проводить комплексную оценку рисков повреждения инфраструктуры и автоматически рассчитывать альтернативные маршруты трассировки в случае критических сбоев. Системы моделирования обеспечивают высокую устойчивость сети, позволяя заранее спрогнозировать появление слабых мест и оптимизировать механизмы быстрого восстановления соединений. Использование данных стратегий дает возможность минимизировать время простоя сервисов и гарантировать стабильную работу магистралей даже при масштабных аварийных ситуациях. Такой подход делает проектирование более гибким и позволяет заранее предусмотреть запасные варианты обхода поврежденных зон [11].

Подобные алгоритмы защиты трафика приобретают особую значимость при проектировании магистральных сетей в сложных географических условиях, где функции ПО дополняются анализом энергетического бюджета линии и характеристик спектрального уплотнения (WDM). Такой подход реализуется через математическое моделирование параметров передачи данных, например, при проектировании линий между островами Ментавай. Программный инструментарий позволяет проводить расчеты максимально допустимых расстояний между усилительными пунктами и оценивать влияние затухания в оптическом волокне на качество сигнала при использовании различных длин волн. Также автоматизация оценки соотношения сигнал/шум (OSNR) является критически важной для обеспечения стабильной связи на протяженных морских участках трассы. Подобный анализ эксплуатационных характеристик сети позволяет еще на этапе проектирования выявить узкие места инфраструктуры и оптимизировать выбор активного оборудования, гарантируя соответствие расчетных характеристик фактическим условиям эксплуатации [12].

Наряду с обеспечением технической надежности, критически важным аспектом является обоснование финансовой жизнеспособности проекта, в связи с чем сравнительный анализ также подтверждает, что использование QGIS в связке с аналитическими плагинами является экономически целесообразной альтернативой проприетарным системам для задач бизнес-прогнозирования окупаемости сети в регионах. Применение открытых GIS-платформ позволяет автоматизировать расчеты потенциальной абонентской базы на основе пространственных данных о плотности населения и типе застройки. Инструментарий системы обеспечивает возможность детального планирования

этапов развертывания инфраструктуры с учетом прогнозируемого дохода от предоставления услуг связи в конкретных районах. Такой подход дает возможность оценить финансовую эффективность проекта еще до начала строительных работ, минимизируя инвестиционные риски при внедрении технологий широкополосного доступа. Интеграция методов пространственного анализа с экономическими показателями позволяет гибко корректировать стратегию расширения сети, выбирая наиболее прибыльные направления для прокладки новых линий связи. Это значительно повышает общую эффективность планирования и делает процесс принятия решений более прозрачным и обоснованным [13]. Итоговая систематизация рассмотренных программных продуктов по заданным критериям представлена в таблице 1.

Таблица 1 – Сравнительная характеристика САПР для проектирования ВОЛС.

Программное ПО	Специализированные модули	Степень автоматизации	Логическая иерархия	Основная область применения
ArcGIS	Fiber Manager, Network Analyst	Высокая	Полная поддержка	Проектирование и учет городских сетей
QGIS	Плагины маршрутизации	Средняя	Базовая поддержка	Аналитика и планирования бюджета
AutoCAD	Интеграция с OptiSystem	Низкая	Минимальная	Выпуск рабочей документации
Net2Plan-GIS	Модули оптимизации WDM/5G	Высокая (расчетная часть)	Логический уровень	Планирование емкостей и трафика
Web-ГИС	Мониторинг ODP и обрывов	Низкая	Ограниченная	Эксплуатация и мониторинг

Как показал проведенный анализ, оптимальной программой для комплексного проектирования сложных сетей ВОЛС, включая архитектуру PON, является среда ArcGIS за счет наиболее развитого функционала управления логической иерархией. В то же время, для подготовки прецизионной инженерной документации незаменимым остается AutoCAD, а для задач эксплуатации — Web-GIS решения. Выбор конкретной системы должен быть обусловлен масштабом проекта и требуемой глубиной автоматизации расчетов.

В заключение следует отметить, что в проанализированных научных работах данные программы изучались преимущественно отдельно. Результаты анализа показывают, что наиболее эффективным является гибридный подход, сочетающий аналитическую мощь GIS-систем с чертежной точностью традиционных САПР. Применение описанных инструментов позволяет

значительно сократить сроки проектирования и повысить надежность функционирования телекоммуникационной инфраструктуры.

Список литературы:

1. Эшмурадов Д.Э., Мухамеджанов А.А., Анализ и проектирование плана волоконно-оптических линий связи с использование геоинформационной системы ArcGIS // Теория и практика современной науки №12(78) 2021
2. Vargas-Bustamane M., Zapata M., Álvarez-Tello J., Planning and Management of Fiber Optic Networks Based on a Geographic Information System: A Case in Ecuador // Int. Conference on Information Technology & Systems (ICITS)
3. Fagbola T.M. An Integrated GIS and GPS-Based Approach for Managing Layer 3 Fiber Networks // International Journal of Civil Engineering and Technology (IJCIET), Volume 10, Issue 07, July 2019
4. Taur R.D., Lomte S.S., Beneficial Implementing GIS in Optical Fiber Communication in Government and Private Industries an Overview. // (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5(6), 2014, 7917-7919
5. Sidik R., Ray Y. Interactive map-based distribution point (ODP) mapping design and implementation // Journal of Engineering Science and Technology Vol. 16, No. 3 (2021) 2128-2138
6. Aboagye I.A., Sowah N.L., Owasu- Banahene W., Shaelijah A., Richardson M.A., Boadi-Baah E. Network Design and Route Analysis Using Outside Plant // Journal of Engineering, Volume 2025
7. J.L. Romero-Gázquez, M.V. Bueno-Delgado, F.J. Moreno-Muro and P. Pavon-Marino, Net2Plan-GIS: An open-source Net2Plan extension integrating GIS data for 5G network planning // Int. Conference on Transparent Optical Networks (ICTON)
8. Matrood Z.M., George L.E., Mahmood F.H., A Simple GIS Based Method for Designing Fiber-Network // International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 2, August 2014
9. Akdemir K.Ö., Dursun T., Arslan S., Benzer R., Akcayol M.A., A GIS-based novel active monitoring system for fiber networks // Turkish Journal of Electrical Engineering & Computer Science, Vol. 24, 2016
10. Septima U., Yolanda A., Chandara D., Uzhelia V.A., Designing Fiber Optic Network Infrastructure with FTTX Configuration Using Network Development Life Cycle (NDLC) Method in Solok Regency // Journal of Engineering Science and Technology (JESTEC), Vol. 16, No. 3, 2021
11. Ashraf M.W., Idrus S.M., Iqbal F., Butt R.A., Faheem M., Disaster-Resilient Optical Network Survivability: A comprehensive Survey // Photonics (MDPI)
12. Azmi H.M., Effendi N.S., Afdrida F.A., Adzikri F., Apriono C., Optical Fiber Backbone Network Design and Analysis in the Mentawai Island
13. Ahmed A., Ahmed A., FTTH Planning and Implementation in Sindh, Pakistan: A GIS-Based Case Study // Sukkur IBA University

УДК 004.056:004.85

Лозин В. С., Сараев П. В.

МИРЭА – Российский технологический университет, Москва, Россия

МЕТОДИКА АНОНИМИЗАЦИИ ТЕКСТОВЫХ БАНКОВСКИХ ОБРАЩЕНИЙ ДЛЯ ЗАДАЧ НЕЙРОСЕТЕВОЙ КЛАССИФИКАЦИИ

Аннотация: в статье рассматривается задача классификации текстовых обращений клиентов банка с учётом защиты персональных данных. Предложен метод анонимизации, основанный на замене чувствительных данных (номеров карт, телефонов, email и сумм) на технические токены. Проведён эксперимент на открытом датасете, показавший, что анонимизация не снижает качество модели и может применяться в системах автоматической обработки обращений. **Ключевые слова:** обработка естественного языка, классификация текстов, банковские обращения, анонимизация данных, персональные данные, трансформеры, машинное обучение.

Введение

В условиях цифровизации банковского сектора наблюдается рост количества текстовых обращений клиентов, поступающих через онлайн-каналы: мобильные приложения, интернет-банк и чат-поддержку. Автоматическая обработка таких обращений становится важной задачей, поскольку позволяет ускорить маршрутизацию запросов и снизить нагрузку на операторов.

Одним из ключевых направлений является автоматическая классификация обращений по тематическим категориям. Однако применение методов обработки естественного языка в банковской сфере осложняется наличием в текстах персональных и чувствительных данных, таких как номера карт, телефоны, email-адреса и суммы операций. Использование таких данных без предварительной обработки создаёт риски нарушения требований конфиденциальности [1].

В связи с этим возникает необходимость анонимизации текстовых данных перед их использованием в моделях машинного обучения. При этом важно не только удалить персональные сведения, но и сохранить смысловую структуру текста, необходимую для корректной классификации.

Целью данной работы является разработка и исследование метода анонимизации банковских текстовых обращений и оценка его влияния на качество нейросетевой классификации с использованием трансформерной модели.

Литературный обзор. Задача классификации текстов традиционно решается с использованием методов на основе правил и классических алгоритмов машинного обучения, таких как TF-IDF, логистическая регрессия

и метод опорных векторов. Однако данные подходы ограниченно учитывают контекст и семантические связи в тексте.

Современные методы обработки естественного языка основаны на нейросетевых архитектурах, в частности трансформерах (BERT и его модификации), которые позволяют учитывать контекст слов и лучше работать с короткими и неформальными текстами. Для русскоязычных данных применяются специализированные модели, такие как RuBERT [2].

В банковской сфере особое внимание уделяется защите персональных данных. Текстовые обращения могут содержать чувствительную информацию, что требует применения методов анонимизации. На практике используются подходы на основе регулярных выражений и распознавания именованных сущностей, позволяющие заменять персональные данные на обобщённые токены.

Таким образом, эффективная классификация банковских обращений требует сочетания современных нейросетевых моделей и методов предварительной анонимизации данных.

Материалы и методы. В качестве исходных данных используется открытый датасет банковских обращений, содержащий текст сообщения клиента (text) и его категорию (label). Датасет включает короткие текстовые обращения, отражающие типовые ситуации взаимодействия клиента с банком: проблемы с переводами, входом в приложение, операциями по карте, подозрительными действиями и другими банковскими сервисами [3]. Перед обучением данные очищаются от пропусков и приводятся к единому формату.

Предварительная обработка текстов направлена на снижение шума и удаление чувствительной информации. Она включает несколько этапов:

- приведение текста к нижнему регистру для устранения различий между одинаковыми словами в разных регистрах;
- удаление лишних символов и нормализация пробелов;
- замена потенциально чувствительных данных на технические токены:

- номера банковских карт заменены на [CARD];
- телефонные номера заменены на [PHONE];
- адреса электронной почты заменены на [EMAIL];
- денежные суммы заменены на [AMOUNT].

Такая анонимизация позволяет исключить персональные данные из текстов, сохраняя при этом смысловую структуру обращения. Это важно как с точки зрения информационной безопасности, так и для повышения устойчивости модели, поскольку исключаются случайные числовые признаки, не влияющие на смысл [4].

После преобработки тексты преобразуются в числовой формат с помощью токенизатора AutoTokenizer, соответствующего выбранной модели.

Токенизация разбивает текст на подсловные единицы и сопоставляет им числовые идентификаторы.

Все тексты приводятся к фиксированной длине ($\text{max_length} = 128$ токенов). Более короткие последовательности дополняются специальными токенами, а более длинные обрезаются [5]. Дополнительно формируется *attention mask*, позволяющая модели различать реальные токены и добавленные элементы.

Данные кодируются в числовые метки с помощью *LabelEncoder* и разбиваются на три части: обучающую, валидационную и тестовую выборки в пропорции 70/15/15 с сохранением распределения классов.

В качестве базовой модели используется русскоязычная трансформерная модель *ruBERT-tiny2*. Выбор данной модели обусловлен её компактностью и возможностью обучения на ограниченных вычислительных ресурсах при сохранении способности учитывать контекст текста [6].

Модель дообучается на задаче многоклассовой классификации, где каждому обращению соответствует одна категория. Для обучения используется функция потерь *CrossEntropyLoss*, а оптимизация выполняется с помощью алгоритма *AdamW*. Обучение проводится в течение 5 эпох с фиксированной скоростью обучения.

Качество модели оценивается с использованием стандартных метрик классификации:

- Accuracy – доля правильно классифицированных обращений;
- Precision – точность предсказаний модели по каждому классу;
- Recall – полнота, показывающая, какую долю реальных объектов класса модель смогла обнаружить;
- F1-score – гармоническое среднее между *precision* и *recall*.

Дополнительно используется матрица ошибок для анализа неправильных предсказаний и выявления наиболее сложных для модели категорий [7].

Результаты. В эксперименте используется датасет банковских обращений объёмом 1000 текстов, распределённых по 8 категориям (например, *PAYMENT_OUT_FAIL*, *APP_LOGIN*, *FRAUD_SUSPECTED*). Данные имеют умеренно сбалансированное распределение по классам, что позволяет корректно обучать модель классификации [8].

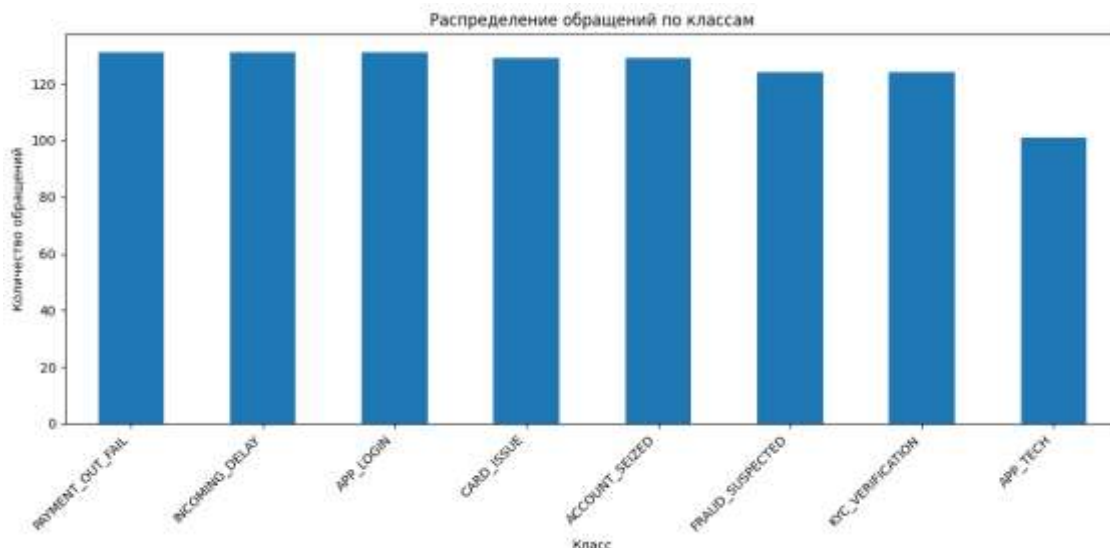


Рис. 1 Распределение обращений по классам

Средняя длина обращения составляет около 9 слов, что подтверждает специфику задачи: модель работает с короткими пользовательскими сообщениями. После предобработки длина текста практически не изменилась, что говорит о сохранении смысловой структуры данных.

Для обучения модель ruBERT-tiny2 использовалось разбиение: 70% – обучение, 15% – валидация, 15% – тест. В процессе обучения (5 эпох) наблюдался устойчивый рост качества: ассигасу на обучающей выборке увеличилась с 0,18 до 0,78, а на валидационной – с 0,47 до 0,74. Одновременно функция потерь монотонно снижалась, что указывает на корректное обучение без выраженного переобучения.

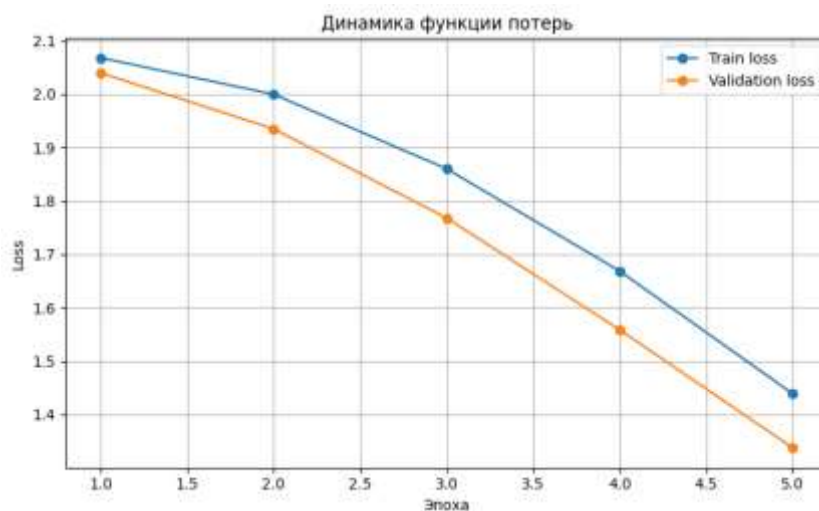


Рис.2 Динамика функции потерь

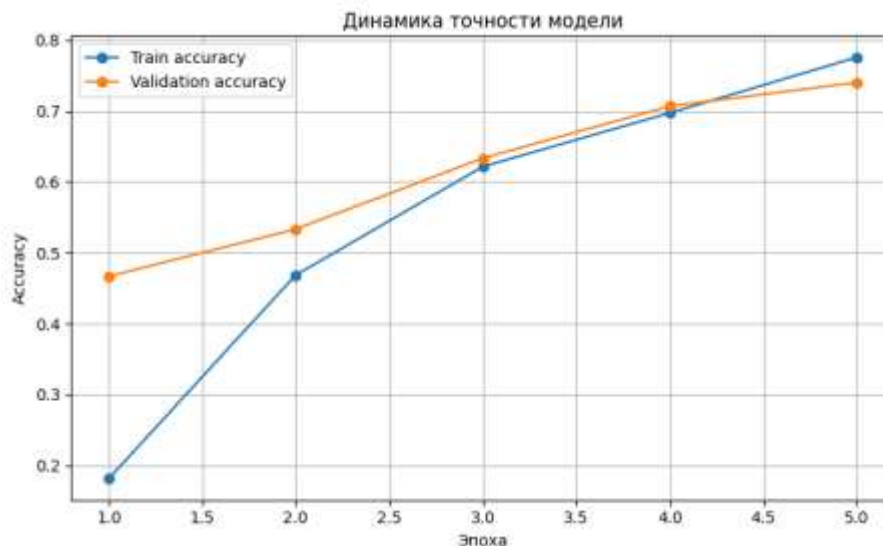


Рис.3 Динамика Accuracy модели

Финальная оценка на тестовой выборке показала accuracy 0,7933. Средние метрики составили: F1-score – 0,78, precision – 0,83, recall – 0,79. Это свидетельствует о достаточно высоком качестве классификации для задачи с короткими текстами и ограниченным объёмом данных.

Анализ по классам показал, что модель лучше всего распознаёт категории, связанные с задержками и входом в приложение (INCOMING_DELAY, APP_LOGIN), где F1-score превышает 0,84. Наиболее сложным оказался класс PAYMENT_OUT_FAIL, для которого recall составил 0,45, что указывает на пропуск части обращений этого типа. Это объясняется пересечением признаков с другими категориями, такими как операции по карте и переводы.

Матрица ошибок показывает, что основные ошибки связаны с семантически близкими категориями:

- PAYMENT_OUT_FAIL – CARD_ISSUE
- APP_TECH – APP_LOGIN
- KYC_VERIFICATION – ACCOUNT_SEIZED

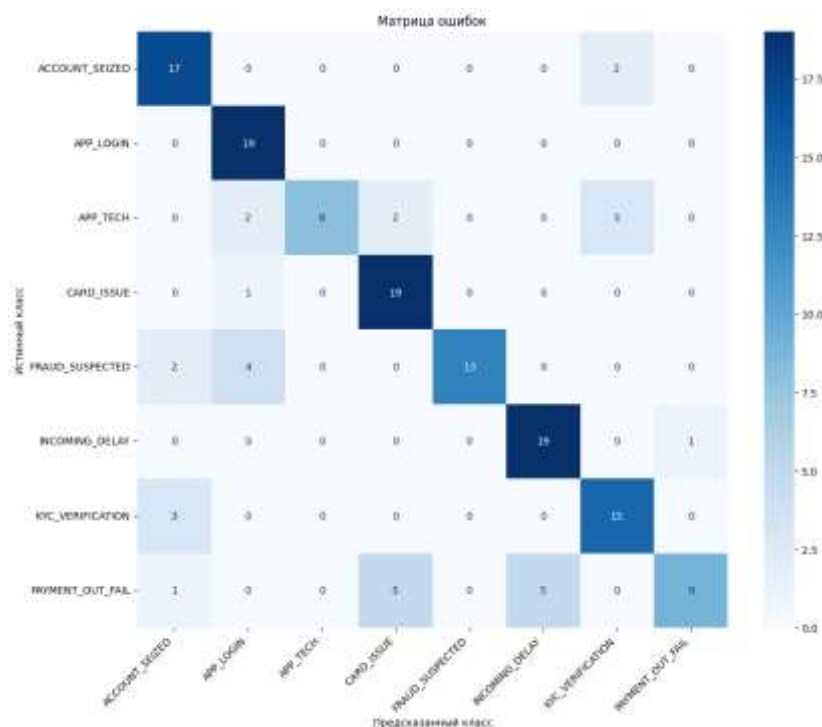


Рис.4 Матрица ошибок

Дополнительная проверка на сложных пользовательских запросах показала, что модель корректно определяет основную категорию обращения, однако при наличии нескольких смыслов может выбирать близкие, но не всегда точные классы [9].

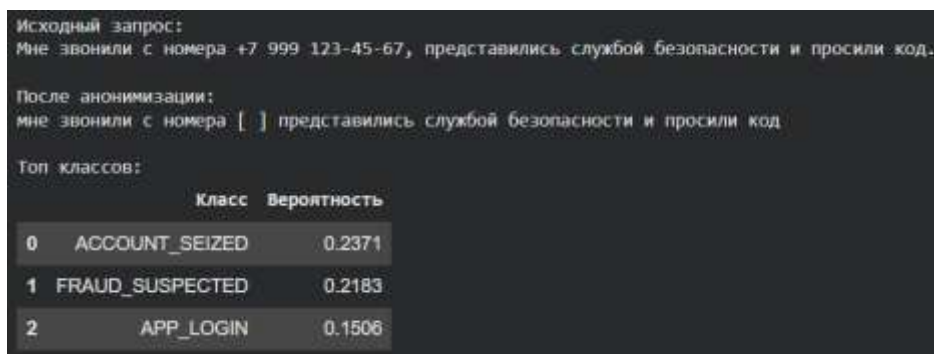


Рис.5 Пример анонимизации и классификации на сложном запросе

Результаты демонстрируют, что предложенный подход с анонимизацией и использованием ruBERT-tiny2 обеспечивает устойчивое качество классификации банковских обращений и может быть использован как основа для систем автоматической маршрутизации клиентских запросов.

Заключение. В работе рассмотрена задача классификации текстовых обращений клиентов банка с учётом требований информационной безопасности. Основное внимание уделено предварительной обработке и анонимизации текстовых данных, поскольку банковские обращения могут

содержать чувствительную информацию, такую как номера карт, телефоны, адреса электронной почты и суммы операций.

Предложен подход к анонимизации, основанный на замене таких данных на технические токены ([CARD], [PHONE], [EMAIL], [AMOUNT]), что позволяет сохранить смысл обращения и одновременно исключить персональные сведения. Показано, что данный этап не нарушает структуру текста и может быть эффективно встроен в пайплайн подготовки данных [10].

Для решения задачи классификации использована трансформерная модель ruBERT-tiny2, дообученная на многоклассовой задаче. Полученные результаты демонстрируют, что модель способна с достаточной точностью определять категорию обращения даже при работе с короткими и неформальными текстами.

Анализ ошибок показал, что основные трудности связаны с семантической близостью отдельных категорий и наличием нескольких смыслов в одном обращении. Это указывает на перспективность дальнейшего развития модели, в частности перехода к многометочной классификации [6].

Ограничением работы является использование ограниченного датасета, который не полностью отражает разнообразие реальных пользовательских обращений. В дальнейшем целесообразно расширить объём данных, использовать реальные обезличенные обращения, а также совершенствовать методы анонимизации и обработки текста.

Таким образом, предложенный подход демонстрирует возможность совместного решения задач защиты данных и автоматизации обработки банковских обращений с использованием современных методов обработки естественного языка.

Список литературы:

1. Лагутина К. В., Бойчук Е. И., Лагутина Н. С. Автоматическая классификация русскоязычных Интернет-текстов по жанрам // Искусственный интеллект и принятие решений. – 2023. – № 4. – С. 103–114.
2. Соловьев В. Д., Солнышкина М. И., Тен А., Прокопьев Н. А. Модель классификации на основе BERT: пример применения к русским сказкам // Journal of Language and Education. – 2024. – Т. 10, № 4. – С. 98–111. – DOI: 10.17323/jle.2024.24030.
3. Солнышкина М. И., Соловьев В. Д., Эбзеева Ю. Н. Подходы и инструменты лингвистического профилирования текста на русском языке // Русистика. – 2024. – Т. 22, № 4. – С. 501–517.
4. Косых Н. Е. Применение модели дистилляции знаний BERT для анализа настроений текста // Вестник ВГУ. Серия: Системный анализ и информационные технологии. – 2022. – № 3. – С. 139–151.
5. Лагутина К. В. Classification of Russian Texts by Genres Based on Modern Embeddings and Rhythm // Моделирование и анализ информационных систем. – 2022. – Т. 29, № 4. – С. 334–347.
6. Каширин И.Ю. Токенизация политических текстов в BERT-моделях с использованием онтологий // Информационные технологии. - 2024. - Т. 30, № 12. - С. 622 - 632. - DOI: 10.17587/it.30.622-632.

7. Федотова А. М., Романов А. С. Методика идентификации текстов, сгенерированных большими языковыми моделями // Информатика и автоматизация. – 2024. – № 5.
8. Данилов Г. А., Ишанкулов Т. Р., Котик К. В., Орлов Ю. Н., Шифрин М. А., Потапов А. А. Классификация коротких научных текстов с использованием предобученной модели BERT // Stud Health Technol Inform. – 2021. – Т. 281. – С. 83–87.
9. Лагутина К. В. Тематическая классификация русскоязычных новостных текстов: сравнение моделей BERT и классических подходов // Труды конференции FRUCT. – 2022.
10. Галушко И. Н. Классификация исторических документов с использованием модели BERT // Вестник Пермского университета. История. – 2025. – № 2. – С. 147–158.

УДК 004.42:004.738.5

Ившин В. А.

Уфимский университет науки и технологий, Уфа, Россия

ИССЛЕДОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ ВЕБ-ПРИЛОЖЕНИЙ НА ОСНОВЕ VUE 3, BUN И ELYSIA

Аннотация. В статье рассматриваются вопросы производительности веб-приложений, разработанных с использованием Vue 3, Bun и Elysia. Проведен анализ подходов к созданию клиент-серверных приложений и исследованы особенности среды выполнения Bun по сравнению с традиционными JavaScript-платформами. Предложена архитектура приложения, обеспечивающая высокую скорость обработки запросов и снижение потребления серверных ресурсов. Выполнено экспериментальное исследование производительности серверной части при различных уровнях нагрузки. Результаты показывают, что применение Bun и Elysia позволяет сократить время отклика сервера и повысить пропускную способность системы по сравнению с решениями на основе Node.js.

Ключевые слова: веб-приложение, Bun, Elysia, производительность, JavaScript, Vue 3, REST API.

Введение

Современные веб-приложения предъявляют высокие требования к производительности как клиентской, так и серверной частей. Рост количества интерактивных интерфейсов, увеличение объема передаваемых данных и необходимость обработки большого числа одновременных запросов требуют применения эффективных технологий разработки.

Традиционно для серверной разработки на JavaScript используется платформа Node.js [1]. Однако в последние годы появились альтернативные решения, направленные на повышение производительности и оптимизацию потребления ресурсов. Одним из таких решений является Bun — современный JavaScript-рантайм, разработанный с акцентом на высокую скорость выполнения и минимизацию накладных расходов.

Одновременно развивается экосистема фронтенд-фреймворков. Vue 3 представляет собой современный прогрессивный JavaScript-фреймворк, ориентированный на создание высокопроизводительных пользовательских интерфейсов. Существенными преимуществами Vue 3 являются улучшенная реактивность, поддержка Composition API и оптимизированная работа виртуального DOM.

Для построения серверной части в среде Bun активно применяется фреймворк Elysia, который обеспечивает высокую скорость обработки HTTP-запросов благодаря тесной интеграции с Bun и использованию современных механизмов маршрутизации.

Актуальность темы исследования определяется необходимостью оценки практической эффективности использования нового технологического стека Vue 3 + Bun + Elysia в реальных условиях эксплуатации веб-приложений.

Объектом исследования является производительность веб-приложений.

Предметом исследования выступают методы повышения производительности веб-приложений с использованием Vue 3, Bun и Elysia.

Цель исследования — провести анализ и оценку производительности веб-приложения, построенного на основе Vue 3, Bun и Elysia.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ современных подходов к обеспечению производительности веб-приложений.
2. Изучить архитектурные особенности Vue 3, Bun и Elysia.
3. Разработать экспериментальное веб-приложение.
4. Провести нагрузочное тестирование серверной части.
5. Оценить производительность клиентской части.
6. Выполнить сравнительный анализ с традиционным стеком Node.js + Express.
7. Сформулировать рекомендации по использованию исследуемого стека.

Методами исследования являются сравнительный анализ, нагрузочное тестирование, экспериментальное моделирование и статистическая обработка результатов.

1. Теоретические основы производительности веб приложений

1.1 Понятие производительности веб приложений

Производительность веб-приложения представляет собой совокупность характеристик, определяющих скорость выполнения операций, обработки запросов и взаимодействия пользователя с интерфейсом. К основным метрикам производительности относятся:

- время ответа сервера;
- время загрузки страницы;
- количество запросов в секунду;
- пропускная способность;
- время гидратации;
- использование процессора и оперативной памяти;
- задержка при взаимодействии с интерфейсом.

Производительность напрямую влияет на пользовательский опыт и эффективность функционирования информационных систем. Низкая скорость загрузки страниц приводит к увеличению показателя отказов и снижению удовлетворенности пользователей.

1.2 Особенности современных SPA-приложений

Single Page Application представляет собой архитектурный подход, при котором большая часть логики приложения выполняется на стороне клиента. После первоначальной загрузки страницы взаимодействие с сервером осуществляется посредством API-запросов [2].

Основными преимуществами SPA являются высокая интерактивность интерфейса, уменьшение количества перезагрузок страницы, снижение нагрузки на сервер, улучшение пользовательского опыта.

Однако SPA-приложения предъявляют повышенные требования к производительности JavaScript-движка и эффективности реактивных механизмов.

1.3 Архитектурные особенности Vue 3

Vue 3 представляет собой современный JavaScript-фреймворк, основанный на реактивной модели данных [3]. Ключевыми особенностями Vue 3 являются:

- использование Proxу вместо Object.defineProperty;
- Composition API;
- оптимизированный виртуальный DOM;
- поддержка Tree Shaking;
- улучшенная работа с TypeScript.

Переход на Proxу позволил значительно сократить накладные расходы реактивной системы и повысить скорость обновления интерфейса.

Composition API обеспечивает более гибкую организацию кода и упрощает повторное использование логики.

1.4 Архитектура Bun

Bun представляет собой JavaScript-рантайм, написанный на языке Zig [4]. Основными особенностями Bun являются:

- высокая скорость запуска;
- встроенный пакетный менеджер;
- встроенный транспайлер TypeScript;
- оптимизированная работа HTTP-сервера;
- совместимость с экосистемой Node.js.

Bun использует JavaScriptCore вместо V8, что обеспечивает высокую производительность выполнения JavaScript-кода.

Одним из важных преимуществ Bun является уменьшение времени cold start по сравнению с Node.js.

1.5 Серверный фреймворк Elysia

Elysia является высокопроизводительным веб-фреймворком для Bun [5]. Он ориентирован на минимизацию накладных расходов при маршрутизации запросов. Основными преимуществами Elysia являются:

- высокая скорость обработки HTTP-запросов;
- поддержка TypeScript из коробки;

- компактный размер;
- встроенная валидация данных;
- эффективная маршрутизация.

Благодаря тесной интеграции с Bun фреймворк демонстрирует высокие показатели производительности при работе под нагрузкой.

2. Проектирование экспериментального веб-приложения

2.1 Архитектура системы

Для проведения исследования было разработано экспериментальное веб-приложение, включающее клиентскую и серверную части. Клиентская часть реализована с использованием Vue 3 и Vite. Серверная часть построена на основе Bun и Elysia.

Архитектура системы включает следующие компоненты: клиентский SPA-интерфейс, REST API, Модуль обработки запросов, Базу данных, Систему кэширования [6].

Взаимодействие между клиентом и сервером осуществляется по протоколу HTTP/HTTPS с использованием JSON.

Технологии, используемые в рамках исследования представлены в Таблице 1.

Таблица 1 – Технологии, используемые в рамках исследования

Компонент	Технология
1	2
Фронтенд	Vue 3
Сборщик	Vite
Язык	TypeScript
Рантайм	Bun
Бекенд-фреймворк	Elysia
База данных	PostgreSQL
ORM	Prisma
Нагрузочное тестирование	K6
Контейнеризация	Docker

2.2 Реализация клиентской и серверной части

Клиентская часть приложения реализована с использованием Composition API. Пример компонента Vue 3 представлен на Рисунке 1.

```
<script setup lang="ts">
import { ref, onMounted } from 'vue'

const users = ref([])
const loading = ref(false)

const fetchUsers = async () => {
  loading.value = true

  const response = await fetch('/api/users')
  users.value = await response.json()

  loading.value = false
}

onMounted(fetchUsers)
</script>
```

Рисунок 1.

Использование Composition API позволило сократить объем кода и повысить читаемость логики.

Серверная часть реализована с использованием Elysia. Пример написания маршрута представлен на Рисунке 2.

```
import { Elysia } from 'elysia'

const app = new Elysia()

app.get('/api/users', async () => {
  return users
})

app.listen(3000)
```

Рисунок 2.

3. Методика исследования производительности

Для проведения исследования производительности веб-приложения была разработана комплексная методика тестирования, включающая анализ серверной и клиентской частей системы. Основной задачей исследования являлось получение объективных количественных показателей, позволяющих оценить эффективность использования стека Vue 3, Bun и Elysia в условиях различной вычислительной нагрузки.

Особое внимание уделялось моделированию реальных сценариев эксплуатации веб-приложения [7]. В процессе тестирования учитывались такие факторы, как количество одновременно подключенных пользователей, интенсивность HTTP-запросов, скорость обработки данных на сервере, а также стабильность работы приложения при продолжительной нагрузке.

Для оценки производительности были выбраны наиболее распространенные метрики, применяемые в современной веб-разработке. Среди них особую роль играют Requests Per Second (RPS), характеризующая количество запросов, обрабатываемых сервером за одну секунду, и Average Response Time, отражающая среднее время ответа сервера на запрос клиента. Кроме того, анализировались показатели Time To First Byte (TTFB), Largest Contentful Paint (LCP) и First Input Delay (FID), позволяющие оценить качество взаимодействия пользователя с интерфейсом.

Методика исследования предполагала последовательное выполнение нескольких этапов. На первом этапе была разработана тестовая инфраструктура, включающая серверную часть на Bun и Elysia, клиентское SPA-приложение на Vue 3, а также систему хранения данных PostgreSQL. Далее производилась настройка инструментов мониторинга и нагрузочного тестирования.

Для генерации нагрузки использовался инструмент k6, позволяющий моделировать поведение большого числа пользователей. Нагрузочные сценарии были сформированы таким образом, чтобы максимально приблизить тестирование к условиям реальной эксплуатации. В частности, выполнялись запросы на получение списков данных, создание новых записей, обновление информации и выполнение асинхронных операций.

В процессе исследования постепенно увеличивалось количество виртуальных пользователей. Это позволило определить пороговые значения нагрузки, при которых начиналось ухудшение производительности системы. Параллельно фиксировались показатели использования процессора, оперативной памяти и сетевых ресурсов.

Для оценки клиентской части использовались встроенные инструменты браузера Google Chrome DevTools, а также Lighthouse. Исследование фронтенда включало анализ скорости первоначальной загрузки страницы, времени гидратации интерфейса, скорости реактивных обновлений и размера итогового JavaScript-бандла.

Особое внимание уделялось сравнению исследуемого стека с традиционным подходом Node.js + Express [8]. Для обеспечения корректности результатов тестирование выполнялось в идентичных условиях. Серверные приложения запускались на одном и том же оборудовании, при одинаковых параметрах операционной системы и сетевой конфигурации.

Аппаратная конфигурация тестового стенда включала процессор AMD Ryzen 7 5800X, 32 ГБ оперативной памяти и NVMe SSD-накопитель. В

качестве операционной системы использовалась Ubuntu 24.04. Серверное окружение включало Bun версии 1.2, Node.js версии 22.x и Vue 3.5.

Важным элементом методики являлась многократная повторяемость экспериментов. Каждый тест выполнялся несколько раз, после чего вычислялись усредненные значения показателей. Это позволило минимизировать влияние случайных факторов и повысить достоверность результатов. Пример конфигурации нагрузочного тестирования с использованием k6 представлен ниже на Рисунке 3.

```
import http from 'k6/http'
import { sleep } from 'k6'

export const options = {
 vus: 100,
  duration: '30s'
}

export default function () {
  http.get('http://localhost:3000/api/users')
  sleep(1)
}
```

Рисунок 3.

Таким образом, выбранная методика исследования обеспечила комплексную оценку производительности исследуемого технологического стека и позволила получить репрезентативные результаты, пригодные для дальнейшего анализа.

4. Результаты исследования

В ходе проведенного исследования были получены результаты, позволяющие оценить эффективность использования Vue 3, Bun и Elysia при разработке современных веб-приложений. Анализ показал, что исследуемый стек демонстрирует высокие показатели производительности как на стороне сервера, так и на стороне клиента.

Одним из ключевых этапов исследования стало сравнение серверной производительности Bun и Node.js. Для этого были реализованы идентичные REST API на основе Elysia и Express соответственно. Оба сервера выполняли одинаковые операции по обработке HTTP-запросов и взаимодействию с базой данных.

Полученные результаты показали, что сервер на основе Bun и Elysia способен обрабатывать существенно большее количество запросов в секунду. Средний показатель производительности составил около 58 000 запросов в секунду, тогда как приложение на Node.js и Express обеспечивало приблизительно 31 000 запросов. Подобная разница объясняется более эффективной архитектурой Bun и оптимизированной реализацией HTTP-сервера.

Также было установлено, что Bun демонстрирует меньшее время отклика. Среднее время обработки запроса составляло около 8 миллисекунд, в то время как аналогичный показатель у Node.js превышал 17 миллисекунд. При высокой нагрузке это различие становится особенно заметным и оказывает существенное влияние на пользовательский опыт.

Исследование показало, что использование Bun позволяет снизить потребление оперативной памяти. Во время стресс-тестирования сервер на Bun использовал около 180 мегабайт памяти, тогда как приложение на Node.js потребляло более 300 мегабайт. Снижение потребления ресурсов особенно важно при развертывании высоконагруженных сервисов в облачной инфраструктуре.

Дополнительно было исследовано время запуска серверного приложения. Bun демонстрировал значительно более быстрый cold start, что связано с особенностями архитектуры рантайма и встроенной поддержкой TypeScript. Среднее время запуска приложения на Bun составляло менее половины секунды, тогда как Node.js требовал почти две секунды.

Помимо серверной части была проведена оценка производительности клиентского приложения, реализованного на Vue 3. Исследование подтвердило высокую эффективность реактивной системы Vue 3 и оптимизированного виртуального DOM.

Анализ показателей Core Web Vitals показал, что приложение обеспечивает быструю загрузку интерфейса и высокую отзывчивость пользовательских элементов. Среднее значение Largest Contentful Paint составляло около 1,4 секунды, что соответствует современным требованиям к производительности веб-приложений.

Время гидратации интерфейса оставалось на низком уровне даже при увеличении количества компонентов и объема отображаемых данных. Это свидетельствует о высокой эффективности механизма реактивности Vue 3 и оптимизации работы Composition API.

Дополнительно исследовалась скорость реактивных обновлений интерфейса при изменении состояния приложения. Было установлено, что Vue 3 эффективно минимизирует количество повторных рендеров и выполняет обновления только для изменившихся компонентов.

В процессе нагрузочного тестирования также анализировалась стабильность системы. Сервер на основе Bun и Elysia сохранял устойчивую работу даже при существенном увеличении числа одновременных подключений. Критического роста времени ответа или отказов в обработке запросов зафиксировано не было.

Отдельное внимание уделялось вопросам масштабируемости. Результаты тестирования показали, что исследуемый стек хорошо подходит для горизонтального масштабирования с использованием Docker-контейнеров и reverse проху-серверов. Благодаря низкому потреблению памяти возможно

размещение большего количества экземпляров приложения на одном физическом сервере.

Полученные результаты подтверждают, что использование Vue 3, Bun и Elysia является перспективным подходом к разработке современных высокопроизводительных веб-приложений. Исследуемый стек обеспечивает высокую скорость обработки запросов, низкое потребление ресурсов и стабильную работу под нагрузкой.

Заклучение

В ходе исследования была проведена комплексная оценка производительности веб-приложения, разработанного с использованием Vue 3, Bun и Elysia.

Полученные результаты показали, что использование Bun и Elysia обеспечивает существенное повышение производительности серверной части по сравнению с традиционным стекком Node.js + Express. Высокая скорость обработки HTTP-запросов, уменьшенное время отклика и более эффективное использование системных ресурсов делают Bun перспективной альтернативой существующим JavaScript-рантаймам.

Исследование клиентской части подтвердило высокую эффективность Vue 3 при разработке современных SPA-приложений. Благодаря использованию Composition API и оптимизированной реактивной системы фреймворк обеспечивает высокую скорость обновления интерфейса и улучшенное взаимодействие с пользователем.

Проведенное нагрузочное тестирование показало, что стек Vue 3 + Bun + Elysia сохраняет стабильность даже при значительном увеличении количества одновременных подключений. Это свидетельствует о возможности применения исследуемых технологий при разработке высоконагруженных корпоративных систем и веб-сервисов.

Практическая значимость исследования заключается в возможности использования полученных результатов при проектировании современных веб-приложений, ориентированных на высокую производительность, масштабируемость и эффективное использование вычислительных ресурсов.

Таким образом, результаты исследования подтверждают перспективность применения Vue 3, Bun и Elysia в качестве основы для построения современных высокопроизводительных веб-приложений.

Список литературы:

1. Флэнаган Д. JavaScript. Подробное руководство, 6-е издание. – Пер. с англ. – СПб: Символ-Плюс, 2012. – 1080 с. ISBN:978-5-93286-215-5
2. Tilkov S., Vinoski S. Node.js: Using JavaScript to Build High-Performance Network Programs. IEEE Internet Computing, 2010, vol. 14, no. 6, pp. 80-83, DOI: 10.1109/MIC.2010.145.
3. Evan You. Vue.js Documentation [Электронный ресурс] // Vue. – URL: <https://vuejs.org/> (дата обращения: 15.05.2026).

4. Bun Documentation [Электронный ресурс] // Bun. – URL: <https://bun.sh/docs> (дата обращения: 15.05.2026).
5. ElysiaJS Documentation [Электронный ресурс] // Elysia. – URL: <https://elysiajs.com/> (дата обращения: 15.05.2026).
6. Приемы объектно-ориентированного проектирования : Паттерны проектирования / Э. Гамма, Р. Хелм, Р. Джонсон, Дж. Влиссидес; [Науч. ред. Н. Шалаев]; [Пер. с англ. А. Слинкин]. - СПб. [и др.] : Питер [и др.], 2001. - 366 с. : ил., табл.; 24 см. - (Библиотека программиста).; ISBN 0-201-63361-2
7. Fowler M. Patterns of Enterprise Application Architecture. Boston: Pearson Education, 2003. 559 p.
8. Pressman R. Software Engineering: A Practitioner's Approach. 7th ed. McGraw-Hill, 2010. 930 p. ISBN 9780073375977

УДК 621.31:004.7

Запорожский С. Е.

Амурский государственный университет, Благовещенск, Россия

РАСПРЕДЕЛЁННЫЙ РЕЕСТР (BLOCKCHAIN) ДЛЯ ОРГАНИЗАЦИИ ТОРГОВЛИ ЭЛЕКТРОЭНЕРГИЕЙ В МИКРОСЕТЯХ С РАСПРЕДЕЛЁННОЙ ГЕНЕРАЦИЕЙ

Аннотация: в статье рассматривается актуальность применения технологии распределенного реестра (блокчейн) для организации торговли электроэнергией в микросетях с распределенной генерацией. Проанализированы предпосылки децентрализации энергетических систем, основные принципы функционирования блокчейна и смарт-контрактов в контексте электроэнергетики. Выявлены преимущества данной технологии, такие как повышение прозрачности, безопасности транзакций и снижение роли посредников. Особое внимание уделено зарубежному опыту внедрения подобных решений и перспективам их адаптации в условиях российской энергетики.

Ключевые слова: блокчейн, распределенный реестр, микросеть, распределенная генерация, смарт-контракт, P2P-торговля электроэнергией.

Введение

Современная электроэнергетика переживает период фундаментальных преобразований, которые исследователи характеризуют как переход к «3D-модели»: декарбонизация, децентрализация и цифровизация [5, с. 2]. Традиционная централизованная модель, основанная на крупных генерирующих источниках и иерархической структуре передачи и распределения энергии, уступает место новой парадигме, где активными участниками становятся потребители, оснащенные собственными генерирующими мощностями. Таких участников называют «просьюмерами» (от англ. prosumer – producer + consumer).

Развитие распределенной генерации, включая возобновляемые источники энергии (солнечные панели, ветрогенераторы), создает предпосылки для формирования локальных энергетических систем – микросетей (microgrids). В таких системах остро встает вопрос организации надежных, прозрачных и эффективных взаиморасчетов между множеством мелких производителей и потребителей. Традиционные механизмы, предполагающие наличие централизованного посредника (энергосбытовой компании), не в полной мере отвечают требованиям гибкости и экономической эффективности децентрализованных систем [6, с. 294].

В качестве перспективного инструмента для решения данной задачи рассматривается технология распределенного реестра (блокчейн), которая изначально получила развитие в финансовом секторе, но сегодня демонстрирует значительный потенциал для применения в энергетике.

Основные принципы применения блокчейна в энергетике

Блокчейн представляет собой децентрализованную базу данных, состоящую из последовательности связанных криптографически блоков, каждый из которых содержит информацию о совершенных транзакциях [7, с. 294]. Ключевыми характеристиками технологии, определяющими ее привлекательность для энергетических приложений, являются:

1. Децентрализация и распределенность – данные хранятся одновременно у всех участников сети, что исключает наличие единой точки отказа и повышает отказоустойчивость системы.
2. Прозрачность и неизменность – все транзакции являются открытыми для участников сети и не могут быть изменены задним числом, что обеспечивает высокий уровень доверия.
3. Автоматизация на основе смарт-контрактов – программные алгоритмы, которые автоматически исполняют условия сделки при наступлении определенных событий [2, с. 2].

Применительно к микросетям с распределенной генерацией блокчейн позволяет организовать прямую P2P-торговлю (peer-to-peer) электроэнергией между просьюмерами и потребителями. Интеллектуальные счетчики фиксируют объемы генерации и потребления в реальном времени, эти данные передаются в блокчейн, а смарт-контракты на основе заданных параметров (цена, объем, временной интервал) инициируют автоматические расчеты.

Архитектурные и технологические аспекты реализации блокчейна в микросетях

Для эффективного функционирования P2P-торговли электроэнергией в микросети необходимо не только применение смарт-контрактов, но и выбор соответствующего типа блокчейн-платформы.

С точки зрения архитектуры, различают публичные (разрешенные) и частные (неразрешенные) распределенные реестры. В публичных блокчейнах, таких как Ethereum или Bitcoin, любой участник может инициировать и верифицировать транзакции, что обеспечивает максимальную децентрализацию, однако сопряжено с высокими вычислительными затратами и ограниченной пропускной способностью [5, с. 112013].

Для энергетических микросетей, где количество участников может исчисляться десятками или сотнями (жилой квартал, промышленный парк или удаленное поселок), более рациональным представляется использование частных или консорциумных блокчейнов (например, Hyperledger Fabric, Quorum). В таких системах доступ к сети предоставляется только авторизованным узлам — генерирующим установкам, потребителям, операторам микросети и регулирующим органам. Это позволяет значительно ускорить обработку транзакций (до нескольких тысяч в секунду) и сократить энергопотребление самой блокчейн-сети, что критически важно при

интеграции с возобновляемыми источниками энергии, обладающими ограниченной и нестабильной генерацией [1, с. 153].

Кроме того, важным технологическим звеном является механизм консенсуса. В отличие от энергозатратного Proof-of-Work (PoW), используемого в криптовалютах, для энергетических приложений предпочтительны алгоритмы Proof-of-Authority (PoA) или Practical Byzantine Fault Tolerance (PBFT). Эти алгоритмы требуют значительно меньших вычислительных ресурсов и позволяют достичь финализации транзакций за секунды, что необходимо для балансировки спроса и предложения в режиме реального времени [3, с. 50]. Например, в проекте Brooklyn Microgrid (Нью-Йорк) именно применение консорциумного блокчейна на базе PoA позволило жителям соседних домов обмениваться излишками энергии, произведённой солнечными панелями, без участия центрального оператора.

Следует также учитывать необходимость хранения больших объёмов данных о почасовых режимах генерации и потребления. Прямая запись каждого замера интеллектуального счётчика в основной блокчейн может привести к неоправданному разрастанию реестра. Альтернативным подходом является использование связки «блокчейн + распределённые хеш-таблицы» (например, IPFS или Swarm), где в блокчейн записывается лишь криптографический хеш массива данных, а сами данные хранятся вне цепи (off-chain) [3, с. 1316]. Такой гибридный подход сохраняет неизменность и верифицируемость информации при снижении нагрузки на сеть.

Экономические и регуляторные барьеры

Помимо технологических вызовов, внедрение блокчейна в российские микросети сталкивается с институциональными ограничениями. Действующее законодательство в сфере электроэнергетики (Федеральный закон № 35-ФЗ «Об электроэнергетике») не содержит прямого запрета на использование распределённых реестров, но и не определяет правового статуса транзакций, совершённых смарт-контрактом. В частности, не урегулирован вопрос о том, какая сторона несёт ответственность за неисполнение автоматического расчёта при сбое оракула или измерительного прибора.

Кроме того, различные тарифы на электроэнергию в России традиционно устанавливаются региональными властями, и P2P-торговля между просьюмерами по динамической цене может вступать в противоречие с принципами перекрёстного субсидирования.

Для адаптации зарубежного опыта в условиях Амурской области и других дальневосточных территорий целесообразно начать с создания песочниц (regulatory sandboxes) на изолированных энергорайонах, не связанных с Единой национальной электрической сетью.

Зарубежный опыт и перспективы внедрения

Анализ мирового опыта показывает растущий интерес к применению блокчейна в электроэнергетике. Согласно исследованиям, половина всех блокчейн-проектов в этой сфере сосредоточена в пяти странах: США (50 проектов), Германия (22), Великобритания (15), Австралия (13) и Япония (13) [5, с. 3]. В таблице 1 систематизированы основные направления и цели применения технологии.

Таблица 1. Области и цели применения блокчейна в электроэнергетике [7, с. 296]

Транзакции и умные контракты	Права собственности и управление ими	Децентрализованные информационные системы
Децентрализованная торговля электроэнергией	Регистрация собственности и ведение реестра активов	Учет электропотребления и выставление счетов
Особые возможности для просьюмеров	«Зеленые» сертификаты	Учет потребления тепла
Внедрение криптовалют	Квоты на выбросы	Оплата зарядки электромобилей
Зарядка электромобилей		
Управление умными устройствами в Интернете вещей		

Критически важным элементом для сопряжения физической инфраструктуры энергосистемы и цифровой среды блокчейна выступают оракулы – программные посредники, которые доставляют данные из внешнего мира (показания счетчиков, метеоданные) в блокчейн и обеспечивают связь между различными информационными системами.

Несмотря на очевидные преимущества, внедрение блокчейн-технологий в электроэнергетике сопряжено с рядом вызовов. Энергетический сектор принципиально отличается от финансового: помимо денежных транзакций, необходимо обеспечить физические поставки электроэнергии, что требует гибкого управления сетевой инфраструктурой. Кроме того, требуют решения вопросы масштабирования систем, обеспечения кибербезопасности и выработки единых стандартов [8, с. 1317].

Российский опыт в данной области пока ограничен – по состоянию на 2020 год было зарегистрировано всего четыре блокчейн-проекта в сфере электроэнергетики [4, с. 3]. Это свидетельствует о значительном потенциале для исследований и разработок, особенно с учетом государственной политики цифровизации экономики и развития распределенной энергетики в удаленных

и изолированных регионах, к которым относится и значительная часть территории Дальнего Востока, включая Амурскую область.

Заключение

Технология распределенного реестра (блокчейн) представляет собой перспективный инструмент для организации торговли электроэнергией в микросетях с распределенной генерацией. Основные преимущества включают повышение прозрачности и безопасности расчетов, снижение транзакционных издержек за счет исключения посредников, возможность автоматизации на основе смарт-контрактов.

Дальнейшие исследования в этом направлении должны быть сосредоточены на разработке архитектурных решений, адаптированных к особенностям российской электроэнергетики, вопросах интеграции с существующей инфраструктурой учета, а также на экономическом обосновании эффективности внедрения подобных систем для различных категорий потребителей и производителей электроэнергии.

Список литературы:

1. Пухно Б.Г. Автоматизована система розрахунків за електроенергію на базі технології блокчейн : магістерська дис. – Київ: КПІ ім. Ігоря Сікорського, 2025. – 153 с.
2. Силин А.В., Кошлаков А.Ю., Акимов Д.А. Использование технологии распределенных реестров (блокчейн) в электроэнергетике // Энергия единой сети. – 2018. – № 5(41). – С. 48–54.
3. Филькин М.Е., Божко М.А. Applications and prospects of distributed ledger technology in the energy sector // Экономика и предпринимательство. – 2023. – № 9(158). – С. 1315–1319. – DOI: 10.34925/EIP.2023.158.9.258.
4. Blockchain in the energy sector // Editorum. – 2024. – С. 293–297. – URL: <https://angtu.editorum.ru/en/nauka/article/84433/view> (дата обращения: 12.03.2026).
5. Kirli D., Couraud B., Robu V. et al. Smart contracts in energy systems: A systematic review of fundamental approaches and implementations // Renewable and Sustainable Energy Reviews. – 2022. – Vol. 158. – Article 112013. – DOI: 10.1016/j.rser.2021.112013.

УДК 69.05:004.85

Ефимова К. С.

*Санкт-Петербургский государственный архитектурно-строительный университет,
Санкт-Петербург, Россия*

АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ И СТОИМОСТНЫХ ПАРАМЕТРОВ СТРОИТЕЛЬСТВА

Аннотация: статья посвящена анализу применения алгоритмов машинного обучения для прогнозирования сроков и стоимости строительных проектов. Выявлены ограничения традиционных методов, такие как зависимость от экспертных оценок и слабая цифровизация. Обоснована актуальность методов машинного обучения в условиях роста сложности процессов. Рассмотрены источники данных, ключевые алгоритмы и метрики оценки. Установлено повышение точности прогнозов и снижение рисков. Выводом статьи является применение технологий анализа данных повышает эффективность управления проектами.

Ключевые слова: машинное обучение; строительство; прогнозирование; сроки; стоимость; BIM; цифровизация; управление проектами; анализ данных.

Введение

Строительная отрасль сегодня активно трансформируется под влиянием цифровых технологий и искусственного интеллекта. Строительство составляет значительную долю валового внутреннего продукта (около 3–5% в развитых экономиках) и является одной из ключевых отраслей. В этой сфере ежегодно обращаются триллионы рублей, поэтому даже небольшое повышение эффективности может дать существенный экономический эффект.

Традиционные методы управления сроками и стоимостью строительства отличаются высокой неопределённостью, сильной зависимостью от субъективных экспертных оценок и большим объёмом ручной обработки данных. Кроме того, многие документы до сих пор не являются машиночитаемыми, что затрудняет автоматизированный анализ и контроль процессов на стройке.

В связи с этим применение алгоритмов машинного обучения рассматривается как перспективный способ повысить точность прогнозирования сроков и бюджетов строительных проектов. Объектом исследования выступают процессы планирования и управления временными и стоимостными параметрами строительных проектов, а предметом — использование алгоритмов машинного обучения для их прогнозирования.

Цель работы заключается в обосновании возможности применения алгоритмов машинного обучения для повышения точности прогнозирования сроков и стоимости строительства. Актуальность исследования обусловлена необходимостью сделать строительные проекты более прозрачными и

предсказуемыми в условиях усложнения объектов, роста объёмов данных и ужесточения требований к соблюдению сроков и бюджета.

Обзор существующих методов прогнозирования

В рамках традиционных подходов к прогнозированию временных и стоимостных параметров строительства используются аналитические модели, статистические методы и экспертные оценки. Эти методы во многом опираются на ручной сбор информации, бумажный документооборот и субъективные решения специалистов, что приводит к ошибкам и искажению фактической картины хода строительства.

В условиях длительного инвестиционно-строительного цикла и большого количества нормативных требований данные ограничения существенно снижают эффективность классических методов. Развитие технологий искусственного интеллекта и цифровизации позволило перейти к обработке больших массивов данных, автоматизации контроля и внедрению предиктивных моделей. При этом традиционные методы, как правило, не позволяют в полной мере использовать накопленные исторические данные и разрозненные источники информации, что усиливает необходимость внедрения более гибких и адаптивных инструментов прогнозирования.

Алгоритмы машинного обучения в прогнозировании строительных параметров

В строительной отрасли находят применение различные алгоритмы машинного обучения. Источниками данных для построения моделей служат BIM-модели, исторические данные реализованных проектов, показания датчиков IoT, видеопотоки со строительных площадок и проектная документация.

Информационное моделирование зданий (BIM) представляет собой цифровой двойник объекта, включающий структурированные данные. Для моделей машинного обучения из BIM извлекаются ключевые параметры: объёмы работ (бетон, металл, земля), характеристики материалов и сложность конструктивных узлов. Это позволяет детально анализировать проект ещё до начала строительства.

Алгоритмы машинного обучения адаптируются под специфику строительных задач и используются для выявления коллизий, мониторинга хода строительства, анализа загрузки техники и прогнозирования возможных отклонений от плановых показателей. К числу применяемых алгоритмов относятся модели линейной и логистической регрессии, деревья решений, ансамблевые методы (случайный лес, градиентный бустинг), а также искусственные нейронные сети [1]. Конкретный выбор алгоритма определяется типом задачи (регрессия или классификация), объёмом доступных данных и требованиями к интерпретируемости результатов. Современные исследования показывают, что использование методов объяснимого искусственного

интеллекта, таких как SHAR, позволяет повысить прозрачность моделей и доверие к ним со стороны специалистов [2].

На основе результатов работы моделей возможно формирование рекомендаций по корректировке графиков производства работ, перераспределению ресурсов и оптимизации закупок строительных материалов. Особое внимание уделяется прогнозированию рисков перерасхода средств, где алгоритмы демонстрируют высокую эффективность в различных секторах строительства (жилом, коммерческом, промышленном), учитывая специфику каждого из них [3].

Оценка эффективности машинного обучения

Эффективность применения алгоритмов машинного обучения оценивается на основе точности прогнозов сроков и стоимости строительства. Когда подобные алгоритмы используются на практике, они демонстрируют значительный рост производительности труда за счёт автоматизации процессов, мониторинга строительных площадок и внедрения предиктивных моделей.

Использование видеоаналитики, датчиков и интеллектуальных систем позволяет выявлять простои техники, контролировать выполнение строительных работ и снижать риски срыва сроков. Полнота и качество исходных данных оказывают ключевое влияние на результативность моделей. Одной из главных проблем внедрения машинного обучения является недостаток качественных структурированных данных и их разрозненность, что требует создания единых платформ для сбора информации. Ограничения связаны также с кадровым дефицитом в этой области.

Для количественной оценки качества прогнозов применяются различные метрики ошибок, такие как средняя абсолютная ошибка, среднеквадратическая ошибка и средняя абсолютная процентная ошибка. Они позволяют сопоставлять фактические и прогнозные значения сроков и стоимости и сравнивать эффективность различных моделей. Внедрение доверительных интервалов в прогнозы позволяет оценить надежность предсказаний и управлять финансовыми рисками более эффективно [2].

Также важным аспектом является оценка экономического эффекта от внедрения моделей машинного обучения, включающая снижение риска срыва сроков, сокращение перерасхода бюджета и повышение эффективности использования строительной техники. Развитие новых технологий, таких как 3D-печать в строительстве, также требует применения методов машинного обучения для оценки их устойчивости и влияния на стоимость жизненного цикла проекта [4, 5].

Перспективы развития и заключение

Развитие алгоритмов машинного обучения в строительной сфере связано с дальнейшей автоматизацией процессов принятия решений и внедрением цифровых двойников. Использование больших данных, генеративного искусственного интеллекта и отраслевых платформ данных открывает новые

возможности для повышения эффективности управления строительными проектами, так как применение технологий ИИ позволяет повысить прозрачность процессов и обеспечить более точное прогнозирование временных и стоимостных параметров строительства.

Перспективным направлением является интеграция предиктивной аналитики с системами информационного моделирования зданий и корпоративными системами управления, что обеспечивает единое информационное пространство для всех участников строительства.

По результатам проведенного анализа можно сформулировать следующие выводы:

1) Традиционные методы прогнозирования сроков и стоимости строительства ограничены субъективностью экспертных оценок и недостаточным использованием имеющихся данных, что приводит к значительным отклонениям плановых показателей от фактических.

2) Алгоритмы машинного обучения, опирающиеся на данные BIM-моделей, датчиков и систем мониторинга, позволяют повысить точность прогнозов и своевременно выявлять потенциальные отклонения от графика и бюджета [1], [3].

3) Эффективность применения машинного обучения в строительстве зависит от качества исходных данных, уровня цифровизации предприятий и наличия специалистов, владеющих методами анализа данных и разработкой моделей.

4) Дальнейшее развитие связано с созданием комплексных цифровых платформ, интеграцией предиктивных моделей в процессы управления проектами и оценкой экономического эффекта от их внедрения.

Список литературы:

1. Chen G., Zheng S., He X., Liang X., Liao X. Machine Learning-Based Cost Estimation Models for Office Buildings // Buildings. 2025. Т. 15, № 11. Ст. 1802.
2. Chen L., Xu C., Lim W. H., Sharma A., Tiang S. S., Chong K. S., El-kenawy E.-S. M., Alhussan A. A., Eid M. M., Khafaga D. S. Transparent and reliable construction cost prediction using advanced machine learning and explainable AI // Engineering Science and Technology, an International Journal. 2025. Т. 70. Ст. 102159.
3. Gunarani G. I., Rajkumar S., Thennaleeswaran D., Venateswaran A. S., Rathinakumar V. A Machine Learning-Driven Framework for Sector-Specific Construction Cost Overrun Prediction and Mitigation // EPJ Web of Conferences. 2026. Т. 354. Ст. 02002.
4. Lafhaj Z., Dakhli Z. Performance indicators of printed construction materials: a durability-based approach // Buildings. 2019. Т. 9, № 4. Ст. 97.
5. Motalebi A., Khondoker M. A. H., Kabir G. A systematic review of life cycle assessments of 3D concrete printing // Sustainable Operations and Computers. 2024. Т. 5. С. 41-50.

УДК 621.9:681.5

Стражнев В. В.

Волгоградский государственный технический университет, Волгоград, Россия

СИСТЕМА АВТОМАТИЗИРОВАННОГО УПРАВЛЕНИЯ РЕЖИМАМИ ТОКАРНОЙ ОБРАБОТКИ НА СТАНКАХ С ЧПУ НА ОСНОВЕ СИГНАЛА ТЕРМОЭДС ИЗ ЗОНЫ РЕЗАНИЯ

Аннотация: в статье рассматривается система автоматизированного управления режимами токарной обработки на станках с ЧПУ на основе сигнала термоэлектродвижущей силы (термоЭДС), возникающего в зоне резания. Предложен подход к автоматическому определению рациональных режимов обработки с использованием оперативной информации о текущем состоянии технологической системы. Показано, что применение сигнала термоЭДС позволяет учитывать реальные условия резания без использования нормативно-справочных данных и предварительных механических испытаний материалов. Рассмотрены принцип работы системы, алгоритм оптимизации режимов резания и конструктивная реализация интеллектуального устройства автоматического назначения параметров обработки. Показана практическая реализация работоспособности системы на токарном станке с ЧПУ. Установлено, что разработанная система обеспечивает требуемое качество обработанной поверхности при соблюдении заданных технологических ограничений и позволяет повысить эффективность процесса обработки. Полученные результаты подтверждают перспективность применения адаптивных интеллектуальных систем управления в условиях цифрового машиностроительного производства.

Ключевые слова: термоЭДС, адаптивное управление, токарная обработка, ЧПУ, зона резания, оптимизация, Индустрия 4.0, цифровое производство.

Актуальность исследования

В условиях перехода промышленности к концепции Индустрии 4.0 машиностроительные предприятия активно внедряют цифровые, интеллектуальные и адаптивные системы управления технологическими процессами механической обработки [1, 3, 12, 13, 14, 15, 16]. Современное производство ориентировано на создание цифровых фабрик, в которых оборудование, системы автоматизированного проектирования, средства мониторинга и системы управления объединяются в единую информационную среду. В таких условиях особую значимость приобретает задача получения достоверной информации о процессе резания непосредственно в ходе обработки, поскольку именно она определяет возможность реализации принципов интеллектуального управления.

Одной из ключевых задач цифрового машиностроительного производства является автоматизированное назначение и корректировка режимов резания на станках с ЧПУ. Несмотря на высокий уровень развития современных CAD/CAM/CAE-систем и систем ЧПУ, проблема выбора рациональных

режимов обработки до настоящего времени остается окончательно не решённой. Это связано с тем, что большинство существующих методик расчёта режимов резания основаны на использовании нормативно-справочных данных, усреднённых эмпирических зависимостей и табличных коэффициентов. Такие методы не учитывают реальные условия обработки, возникающие непосредственно в производственной среде.

На практике физико-механические свойства обрабатываемых материалов существенно изменяются даже в пределах одной марки стали вследствие особенностей металлургического производства, различий в термической обработке, химическом составе и структуре материала. Дополнительное влияние оказывают износ режущего инструмента, тепловые деформации, колебания технологической системы и нестабильность контактных процессов в зоне резания. В результате режимы обработки, рассчитанные по традиционным методикам, нередко приводят либо к преждевременному износу инструмента, либо к снижению производительности и ухудшению качества обработанной поверхности [1, 11].

Особенно остро данная проблема проявляется в условиях гибкого автоматизированного производства и мелкосерийного производства деталей, где технологическая система должна быстро адаптироваться к изменяющимся условиям обработки. Для реализации концепции «умного» производства становится необходимым переход от статических моделей расчёта режимов резания к адаптивным системам управления, функционирующим по принципу замкнутого цикла: «измерение параметров процесса → анализ состояния → автоматическая корректировка режимов обработки» [4].

Одним из перспективных направлений решения данной задачи является использование оперативной информации из зоны резания. Среди различных диагностических параметров особый интерес представляет сигнал термоэлектродвижущей силы (термоЭДС), возникающий в естественной термопаре «инструмент–заготовка» в процессе резания. Данный сигнал непосредственно связан с температурными и контактными процессами в зоне резания, а также отражает физико-механические свойства контактирующих материалов. В отличие от традиционных методов контроля, основанных на предварительных механических испытаниях и лабораторных измерениях, регистрация термоЭДС может осуществляться непосредственно в процессе обработки без остановки оборудования и без внесения изменений в конструкцию станка.

Использование сигнала термоЭДС открывает возможность создания интеллектуальных адаптивных систем управления режимами резания, способных автоматически определять оптимальные параметры обработки с учётом текущего состояния технологической системы. Это позволяет повысить стабильность качества обработанной поверхности, снизить вероятность

аварийного износа инструмента, сократить время технологической подготовки производства и повысить эффективность использования оборудования с ЧПУ.

Таким образом, разработка системы автоматизированного управления режимами токарной обработки на основе сигнала термоЭДС из зоны резания является актуальной научно-технической задачей, соответствующей современным тенденциям развития цифрового машиностроения, интеллектуальных производственных систем и адаптивных технологий обработки материалов.

Принцип работы системы

Механическая обработка сталей и сплавов является сложным для анализа процессом, так как основные процессы сосредоточены в «зоне резания», ограничивающей возможность получения и корректной оценки информации. Вариантом решения данной проблемы является метод пробного прохода: во время обработки измеряется сигнал термоэлектродвижущей силы, который отражает реальные условия резания между обрабатываемой деталью и режущим инструментом. Этот сигнал связан с физическими свойствами материалов, поэтому по нему можно оценить и обрабатываемость детали, и износ инструмента. Поскольку термоЭДС возникает прямо в процессе обработки и легко измеряется, отпадает необходимость в предварительных механических испытаниях сталей с целью определения поправочных коэффициентов в формулах расчета режимов обработки. На основе этих данных существует возможность автоматизировать подбор оптимальных режимов обработки с помощью оптимизации. Для определения оптимальных режимов резания можно применять различные методы, среди которых наиболее распространены метод перебора, графический и симплекс метод. Для наглядного представления математической модели процесса резания подходит графический метод (рисунок 1), но для автоматизированного управления режимами токарной обработки наиболее эффективным считается симплексный метод [11].

Симплексный метод основан на последовательном переборе допустимых решений, при котором на каждом шаге улучшается значение целевой функции. Для применения симплекс метода система технологических ограничений приводится к виду равенств, в это равенство вводятся фиктивные переменные участвующие в расчетах вместе с основными переменными.

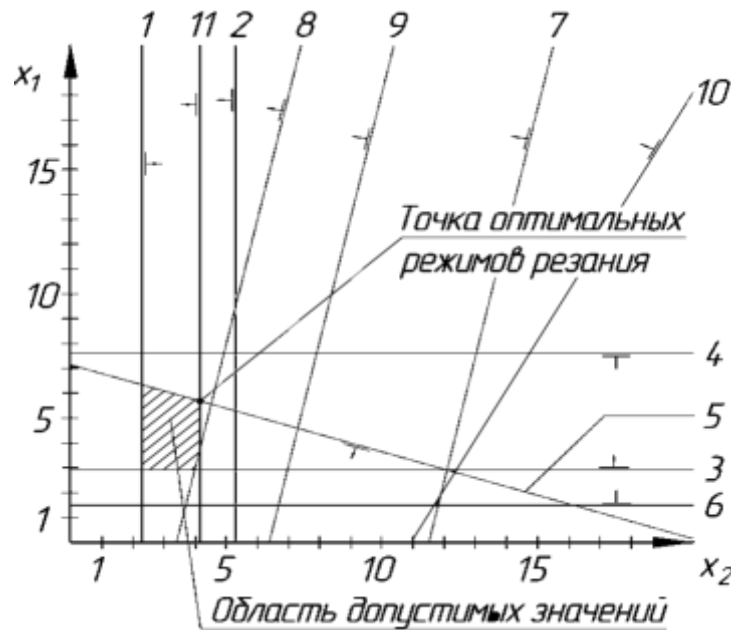


Рисунок 1 – Графическое решение задачи определения оптимальных режимов точения

Целевые функции могут отличаться и выбираются в зависимости от задач управления процессом обработки. Например, на этапе черновой обработки в качестве критерия обычно используют производительность. При получистовой и чистовой токарной обработке основным показателем становится точность формы и взаимного расположения поверхностей. За оценочную функцию в случае финишной обработки целесообразно принять величину получаемой шероховатости.

Практическая реализация системы

Система реализована в виде интеллектуального устройство автоматического назначения режимов точения [5-8]. Устройство автоматического выбора режимов токарной обработки конструктивно разделено на две части. Основная часть выполняет ключевые функции: измеряет ЭДС резания, оптимизирует режимы обработки, обеспечивает связь со смартфоном и передаёт данные на вспомогательный модуль. Вспомогательная часть предназначена для обеспечения взаимодействия основной части со станком и реализуется с использованием микроконтроллера, радиоприемопередатчика и преобразователя уровней сигналов интерфейса UART–RS232.

Работа устройства автоматического назначения режимов резания осуществляется следующим образом. Устройство устанавливается на револьверной головке станка с помощью магнитного крепления. Между режущим инструментом и заготовкой формируется измерительная цепь: один проводник подключается к заземлению станка, второй проводник подключается к режущему инструменту через электроизолирующие прокладки [6-8].

В процессе пробного прохода с использованием аналого-цифрового преобразователя выполняется измерение термоэлектродвижущей силы резания. На основе полученных данных производится расчет оптимальных численных значений скорости резания (м/мин) и величины подачи (мм/об), с применением соответствующего симплекс-метода решения задач оптимизации [9].

В качестве целевой функции выступает производительность, которая приводится к максимуму (1):

$$T_0 = \frac{L}{n \cdot S} \rightarrow \min \Rightarrow n \cdot S \rightarrow \max, \quad (1)$$

где T_0 - основное время обработки;

n – частота вращения заготовки, об/мин;

S – подача, м/мин;

L – длина рабочего хода, мм.

Рассчитанные параметры передаются в систему числового программного управления станка, где сохраняются в пользовательских переменных [10]. После этого осуществляется основной рабочий проход обработки заготовки с использованием оптимизированных режимов.

На процесс обработки накладывается система технологических ограничений состоящая из:

1. наименьшая подача, допустимая кинематикой станка;
2. наибольшая подача, допустимая кинематикой станка;
3. наименьшая скорость резания, допустимая кинематикой станка;
4. наибольшая скорость резания, допустимая кинематикой станка;
5. режущие возможности инструмента;
6. наибольшая технологически допустимая скорость резания;
7. мощность электропривода главного движения станка;
8. прочность режущего инструмента;
9. жесткость режущего инструмента;
10. точность обработки;

11. наибольшая подача, допускаемая требованиями, предъявляемыми к шероховатости обработанной поверхности.

Экспериментальная проверка работоспособности системы

С целью проверки работоспособности разработанной системы был проведён эксперимент [2] на токарном станке OKUMA GENOS L300-M с ЧПУ. Задачей эксперимента являлось получение цилиндрической поверхности с заданной шероховатостью при использовании исследуемого устройства. В качестве заготовки использовали цилиндрическую деталь из стали 40Х ГОСТ 4543-2016 длиной 300 мм и диаметром 80 мм, которая предварительно прошла черновую обработку. Заготовку закрепляли в патроне с дополнительной поддержкой задним центром для повышения жесткости системы. Обработка выполнялась твердосплавной пластиной Sandvik Coromant WNMG 08 04 08-PM 4425. К станку было подключено специальное интеллектуальное устройство

автоматического назначения режимов резания. Оно взаимодействовало с системой ЧПУ через интерфейс RS-232 и позволяло автоматически передавать рассчитанные параметры обработки прямо в программу станка. Перед основной обработкой выполнялся пробный проход. Его задача заключалась в том, чтобы получить информацию о текущих условиях резания. Во время этого прохода измерялся сигнал ЭДС резания контактной термопары «инструмент-заготовка». Пробный проход проводился при фиксированных параметрах: глубина резания 0,1 мм, скорость резания 100 м/мин, подача 0,1 мм/об и длина обработки 10 мм. После получения сигнала система автоматически рассчитывала оптимальные режимы точения. В качестве целевой функции принималось минимальное время обработки, то есть система стремилась обеспечить максимально производительный режим. Однако при этом учитывались технологические ограничения, главным из которых была максимально допустимая высота микронеровностей поверхности. Это значение задавалось заранее перед обработкой. Затем выполнялся основной проход уже на автоматически определённых режимах. После обработки качество поверхности измерялось с помощью профилографа-профилометра Mitutoyo Surftest SJ-210. Всего провели 15 опытов, в которых задавали разные допустимые значения высоты микронеровностей H_{max} от 20 до 80 мкм. Результаты показали, что фактическая шероховатость R_z во всех опытах оставалась ниже заданного предельного значения H_{max} . Например, при значении $H_{max}=20$ мкм реальная шероховатость составляла примерно 4,8-10,3 мкм, а при $H_{max}=80$ мкм около 16-19 мкм. Это означает, что система не только выдерживала требования, но и обеспечивала запас по качеству. Получившийся профиль (рисунок 2) поверхности имеет равномерную структуру пиков и педин и стабильный шаг, что свидетельствует о качественном процессе резания без вибраций и дефектов.

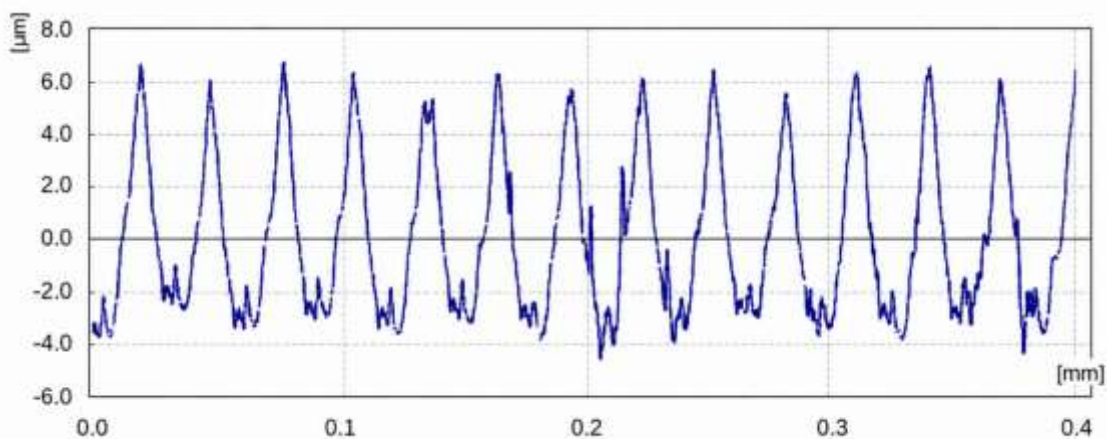


Рисунок 2 – Профиль микронеровностей поверхности

В итоге эксперимент подтвердил, что автоматическое назначение режимов резания на основе сигнала ЭДС действительно может обеспечить

стабильное качество поверхности и одновременно оптимизировать производительность обработки.

Выводы

Рассмотрена система автоматизированного управления режимами токарной обработки на станках с ЧПУ, основанная на использовании сигнала термоэлектродвижущей силы, возникающего в зоне резания. Предложенная система позволяет получать оперативную информацию о текущем состоянии процесса обработки непосредственно во время резания.

Экспериментальные исследования [2] на токарном станке с ЧПУ подтвердили работоспособность разработанной системы. Установлено, что автоматическое назначение режимов обработки обеспечивает получение требуемого качества поверхности при соблюдении заданных технологических ограничений и позволяет повысить эффективность процесса обработки.

Полученные результаты подтверждают перспективность применения адаптивных интеллектуальных систем управления режимами резания в условиях цифрового машиностроительного производства и могут быть использованы при создании интеллектуальных производственных участков и систем цифрового технологического сопровождения обработки.

Предлагаемая система при внедрении на станочное оборудование не требует внесения изменений в конструкцию узлов и программное обеспечение системы ЧПУ станка, что обеспечивает возможность её применения на оборудовании, которое находится на гарантии производителя.

Список литературы:

1. Плотников, А.Л. Способ получения оперативной информации о свойствах контактной пары деталь-инструмент с использованием предварительного пробного прохода / А.Л. Плотников, Е.М. Фролов // Вестник Саратовского государственного технического университета. - 2008. - № 3, вып. 1. - С. 39-47.
2. Исследование качества формируемой при точении на автоматически определяемых режимах поверхности / Е.М. Фролов, А.В. Рогачев, М.Ю. Полянчикова, В.В. Стражев // Известия ВолгГТУ. Сер. Прогрессивные технологии в машиностроении. - Волгоград, 2023. - № 1 (272). - С. 32-34. - DOI: 10.35211/1990-5297-2023-1-272-32-34.
3. Интеллектуальные мехатронные станочные системы / В.Ц. Зориктуев // Вестник УГАТУ. Сер. Технология и оборудование механической и физико-технической обработки. – Уфа, 2007. - Т. 9, №1 (19). С. 92–99.
4. Чигиринский Ю.Л., Крайнев Д.В., Фролов Е.М. Цифровизация машиностроительного производства: технологическая подготовка, производство, прослеживание // Научные технологии в машиностроении.-2022-8(134)-С.39-48. DOI: 10.30987/2223-4608-2022-8-39-48
5. Способ назначения режимов резания для станков с ЧПУ на основе ЭДС резания: заявка № 2021135 Российская Федерация: В23В 25/06 (2006.01), В23Q 15/00 (2006.01) / Фролов Е.М., Рогачев А. В.; заявитель ФГБОУ ВО «ВолгГТУ». – № 2021135456/05; заявлен 02.12.2021. – 30 с.
6. П. м. 201939 Российская Федерация, МПК В23В25/06 Резец сборный с элементами для измерения ЭДС резания / Е.М. Фролов, А.В. Рогачев; ФГБОУ ВО ВолгГТУ. - 2021.

7. П. м. 201938 Российская Федерация, МПК В23В25/06 Автономный модуль для измерения ЭДС резания / Е.М. Фролов, А.В. Рогачев; ФГБОУ ВО ВолГТУ. - 2021.
8. Пат. 2650827 Российская Федерация, МПК G01K13/00, G01K7/02 Устройство для измерения температуры резца естественной термопарой / А.Л. Плотников, М. Г. Кристаль, А. С. Сергеев, Ж. С. Тихонова, Уварова; ВолГТУ. – 2018
9. Свид. о гос. регистрации программы для ЭВМ № 2022614344 от 18 марта 2022 г. Российская Федерация. Оптимизация режимов резания для станков с ЧПУ / Е.М. Фролов, А.В. Рогачев; ФГБОУ ВО ВолГТУ. - 2022.
10. Фролов, Е. М. Возможности использования серийного оборудования с ЧПУ в качестве основы для построения цифровых производственных участков / Е.М. Фролов, А.В. Рогачев // Известия ВолГТУ: научный журнал № 3 (250) / ВолГТУ. – Волгоград, 2021. – (Серия «Прогрессивные технологии в машиностроении»). – С. 41–43.
11. Фролов, Евгений Михайлович: Повышение надежности определения режимов резания в САПР ТП механической обработки: диссертация ... кандидата технических наук: 05.03.01 / Фролов Евгений Михайлович; [Место защиты: Волгогр. гос. техн. ун-т] Волгоград, 2009 - 145 с.
12. Turning process assisted in situ by short time current pulses / A. J. Sánchez-Egea, H.A. González-Rojas, C. A. Montilla-Montaña, V. Kallewaard-Echeverri // Procedia Engineering.–2015.–№132.–507-512
13. Выбор оптимального варианта достижения необходимых параметров качества поверхности при обработке деталей типа тел вращения (часть 1) / Пахомов Д.С., Шатагин Д.А., Трухин А.В. // Тенденции развития науки и образования.–2020.–№63-2.–33-39
14. Analysis and Multi-Objective Optimization for Reducing Energy Consumption and Improving Surface Quality during Dry Machining of 304 Stainless Steel // Feilong Du, Lin He, Haisong Huang, Tao Zhou, Jinxing Wu // Materials (Basel).–2020.–№13(21)
15. Управление шероховатостью при получении изделий механической обработки / Жуков Е.М., Сипун С.А. // Технические науки – от теории к практике.–2016.–№56.–54-58
16. Ultra-Precision Machining Process Dynamics and Surface Quality Monitoring / Changqing Cheng, Zimo Wang, Wayne Hung, Satish T.S. Bukkapatnam, Ranga Komanduri // Procedia Manufacturing.–2015.–№1.–607-618

УДК 622.24:004.9

Капитонов В. А., Кузьмин Г. В.

Самарский государственный технический университет, Самара, Россия

ЦИФРОВИЗАЦИЯ В СОВРЕМЕННОМ БУРЕНИИ

Аннотация. В статье рассматривается цифровизация как важное направление развития современных методов бурения нефтяных и газовых скважин. Тема становится особенно актуальной из-за необходимости повысить эффективность работы в условиях всё более сложных геолого-технических задач, увеличения глубины скважин и длины горизонтальных участков. Показано, что применение цифровых технологий, в том числе автоматизированных систем управления бурением, помогает точнее контролировать процесс, уменьшать ошибки из-за человеческого фактора, снижать риск аварий и работать с инструментом эффективнее. Отдельно рассматривается, как автоматизированные системы способствуют стабилизации работы долота и улучшают технико-экономические результаты при строительстве скважин.

Ключевые слова: цифровизация, бурение, автоматизированная система управления бурением, мониторинг, долото, механическая скорость, коммерческая скорость, буровой процесс.

В последние годы одним из ключевых направлений в развитии буровой отрасли стала цифровизация. Она связана с применением цифровых технологий для контроля, анализа и управления, что позволяет в реальном времени получать данные о ходе технологического процесса, быстро реагировать на отклонения и принимать обоснованные инженерные решения.

Особенно важна цифровизация на объектах, где нужно поддерживать стабильность бурения при частом чередовании пород, больших нагрузках на инструмент и ограниченном времени для операций. Здесь цифровые технологии становятся не просто помощниками, а одной из основных частей эффективного управления строительством скважины.

Целью настоящей статьи является анализ роли цифровизации в современном бурении и обоснование ее значения как инструмента повышения эффективности, надежности и управляемости процесса строительства нефтяных и газовых скважин.

Под цифровизацией в бурении следует понимать внедрение программно-аппаратных средств, обеспечивающих непрерывный сбор, передачу, обработку и анализ технологических данных с последующим использованием полученной информации для контроля и оптимизации бурового процесса [1].

К основным элементам цифровизации в бурении относятся:

- автоматизированные системы управления бурением;
- датчики контроля технологических параметров;
- системы передачи данных в режиме реального времени;

- программные комплексы визуализации и интерпретации информации;
- интеллектуальные алгоритмы анализа и прогнозирования;
- цифровые базы данных по ранее пробуренным скважинам;
- средства удаленного мониторинга и инженерного сопровождения.

Внедрение таких решений позволяет повысить прозрачность бурового процесса, улучшить контроль текущего состояния системы и создать условия для оперативной корректировки технологических параметров. В результате бурение становится более управляемым и предсказуемым.

Центральное место в цифровизации буровых работ занимает автоматизированная система управления бурением (АСУБ).

Автоматизированная система управления бурением - это современный комплекс программно-аппаратных средств, который позволяет в реальном времени контролировать, анализировать и автоматически корректировать параметры процесса бурения [2].

АСУБ обеспечивает:

- сбор данных с датчиков;
- отображение текущих параметров на рабочем месте бурильщика;
- контроль режимов бурения;
- выдачу рекомендаций по корректировке технологических параметров;
- передачу данных в удаленные центры инженерного сопровождения;
- архивирование информации для последующего анализа.

В процессе бурения система позволяет отслеживать нагрузку на долото, частоту вращения, крутящий момент, давление в циркуляционной системе, расход бурового раствора, вибрации, механическую скорость бурения и другие параметры. Высокая частота обновления данных обеспечивает возможность практически непрерывного контроля, что особенно важно при возникновении нестабильных режимов работы [3].

Такая система представляет собой многоэкранный пульт дистанционного управления. На экранах бурильщик видит все основные параметры: глубину, скорость бурения, нагрузку на долото, обороты в минуту, крутящий момент, давление, вибрацию инструмента, угол наклона инструмента и состав раствора.

Датчики, установленные на всей буровой (ротор, насос и КНБК), передают данные на центральный контроллер и в специальное облачное хранилище. Алгоритмы искусственного интеллекта обрабатывают данные и выдают рекомендации или настраивают их автоматически.

Использование АСУБ помогает точнее управлять процессом бурения, быстрее реагировать на изменения условий на забое и снижать риск, что допустимые отклонения перерастут в проблемы или аварии. Поэтому такую

систему считают важным инструментом для повышения надежности и эффективности при строительстве скважин.

АСУБ работает по принципу замкнутого контура управления. На буровой установке и в буровой колонне ставят датчики, которые с большой частотой фиксируют технологические параметры.

По всей площадке установлено десятки датчиков и модулей, измеряющих ключевые показатели каждые доли секунды: нагрузку на долото, скорость бурения, обороты, вибрацию, давление, геофизические данные и другие параметры. Собранные данные поступают в вычислительный модуль системы, где их сначала фильтруют, проверяют правильность, удаляют помехи и обрабатывают дальше. Затем информация появляется на панели бурильщика и может передаваться в удаленный центр сопровождения, где ее анализируют инженеры и дают рекомендации. Также данные сразу идут на контроллер бурового долота, где частично обрабатываются прямо на месте для быстрого реагирования, а полный поток отправляется в облако или на сервер, где его дополняют исторической информацией и геологической моделью месторождения. После этого алгоритмы искусственного интеллекта анализируют текущую ситуацию почти в реальном времени, сравнивают показатели с эталонной моделью и пытаются предсказать возможные трудности на ближайшие 10-30 секунд.

Еще одна важная черта АСУБ - не только сбор данных, но и сравнение их с заданными диапазонами, проектными значениями и прошлым опытом работы с похожими скважинами. Это позволяет системе вовремя заметить отклонения, понять вероятные причины и разработать меры для управления процессом.

В зависимости от уровня автоматизации АСУБ может либо выдавать бурильщику рекомендации по изменению режима бурения, либо самостоятельно корректировать отдельные параметры работы исполнительных механизмов.

В случае отклонения система автоматически настраивает режимы с помощью привода - изменяет скорость, нагрузку и расход раствора-или дает оператору точные рекомендации на экране. Благодаря интеграции с каротажем в процессе бурения две операции объединяются в одну, и после запуска все данные добавляются в базу знаний, улучшая модели для следующих скважин [3].

Таким образом, реализуется переход от пассивного наблюдения за процессом к активному управлению им.

С практической точки зрения это особенно важно при бурении интервалов со сложным литологическим строением, где изменения твердости пород, очистки забоя и динамического состояния компоновки низа буровой колонны требуют быстрой реакции на меняющиеся условия.

При традиционном управлении бурением оператор ориентируется преимущественно на отдельные текущие показатели и собственный производственный опыт, тогда как АСУБ обеспечивает комплексный анализ сразу нескольких взаимосвязанных параметров [4]. За счет этого повышается точность выбора режима и снижается вероятность ошибочных решений.

АСУБ имеет несколько преимуществ по сравнению с традиционными методами управления бурением. Она быстрее реагирует на изменения параметров, уменьшает риск ошибок из-за человеческого фактора, помогает поддерживать стабильную работу долота и позволяет заранее заметить возможные проблемы. Кроме того, система собирает производственные данные, которые потом можно анализировать, чтобы улучшать режимы бурения и применять опыт на других скважинах с похожими условиями [5].

Инновационность АСУБ заключается в объединении нескольких функций в одной цифровой платформе. Раньше контроль параметров, их анализ и принятие решений часто выполнялись разными программами, приборами и специалистами, а сейчас эти процессы связаны в единую систему. Новизна системы проявляется в использовании интеллектуальных алгоритмов для анализа данных, в оценке состояния бурения с прогнозом, в возможности менять режимы в реальном времени и создавать базу знаний для будущего применения на похожих объектах.

Еще одна важная особенность - переход от реактивного управления к упреждающему. Обычно параметры меняют только после того, как появляется явная проблема. АСУБ же помогает заметить признаки ухудшения раньше, например, при увеличении вибрации, нестабильности крутящего момента или снижении эффективности разрушения породы. Это особенно важно при бурении наклонных и горизонтальных скважин, где ошибки могут дорого стоить.

Использование АСУБ помогает точнее управлять процессом бурения, быстрее реагировать на изменения на забое и снижать риск, что допустимые отклонения перерастут в осложнения или аварии. Поэтому система становится ключевым инструментом повышения надежности и эффективности строительства скважин, а также одним из перспективных направлений цифровизации в бурении.

В отличие от разрозненных систем, где измерение, визуализация и анализ идут отдельно, АСУБ обеспечивает комплексный подход. Ее преимущество в том, что она не просто показывает текущие значения, а учитывает связь между нагрузкой на долото, частотой вращения, гидравликой и вибрацией компоновки.

В итоге АСУБ делает бурение более предсказуемым и технически продвинутым, что позволяет увеличить скорость, снизить риски и сэкономить ресурсы.

Преимущества системы включают:

- повышение коммерческой скорости бурения,
- уменьшение вибрации и износа долота,
- снижение влияния человеческого фактора,
- получение точной геологической информации прямо во время бурения,
- снижение количества аварий на буровой.

Таким образом, АСУБ важно для современного бурения, которое развивается вместе с новыми технологиями. Эта система уже проверена на практике и показала свою полезность, помогая повысить производительность в нефтегазовой сфере.

Список литературы:

1. Храменков, В. Г. Автоматизация управления технологическими процессами бурения нефтегазовых скважин : учебное пособие для вузов / В. Г. Храменков. — Москва : Издательство Юрайт, 2024. — 415 с. — (Высшее образование). — ISBN 978-5-534-00854-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/537112>
2. Дмитриев А.Ю. Основы технологии бурения скважин: учебное пособие. - Томск: Издательство Томского политехнического университета, 2008. - 216 с.
3. Цифровые технологии в добыче углеводородов: цифровая прозрачность. Сборник тезисов науч.-техн. конф. – Уфа, РН-БашНИПИнефть, 2023, 264 с. ISBN 978-5-903404-27-84.
- Булатов А.И., Проселков Ю.М., Шаманов С.А. Технология бурения нефтяных и газовых скважин: учебник для вузов. Краснодар: Советская Кубань, 2009. - 1007 с.
5. Рязанов Я.А., Кершенбаум В.Я. Автоматизация и управление в нефтегазовом комплексе: учебное пособие. М.: РГУ нефти и газа имени И.М. Губкина, 2011. - 382 с.

УДК 622.24:681.5

Капитонов В. А., Кузьмин Г. В.

Самарский государственный технический университет, Самара, Россия

ИНТЕГРАЦИЯ АВТОМАТИЗИРОВАННЫХ СИСТЕМ С ПРИБОРАМИ ИЗМЕРЕНИЯ И КАРОТАЖА ВО ВРЕМЯ БУРЕНИЯ

Аннотация. В статье рассмотрены особенности интеграции автоматизированной системы управления бурением с приборами измерения и каротажа во время бурения. Описан принцип функционирования данной системы, раскрыт механизм обмена данными между поверхностным и забойным оборудованием, а также показаны основные преимущества интегрированного подхода. Проведена сравнительная характеристика интегрированной АСУБ с традиционными и частично автоматизированными решениями.

Ключевые слова: автоматизация, цифровизация, каротаж во время бурения, измерения во время бурения, телеметрия.

Современное бурение нефтяных и газовых скважин характеризуется высокой технологической сложностью. Это связано с тем, что скважины делают глубже, часто наклонными или горизонтальными, и работать приходится в постоянно меняющихся геологических условиях. В такой ситуации качество строительства зависит не только от надежности оборудования, но и от того, насколько хорошо процесс контролируется и поддерживается информацией.

Раньше данные о параметрах бурения и геофизической информации получали и анализировали отдельно. Из-за этого решения принимались с задержкой, из-за чего сложно было быстро менять режимы работы, и увеличивался риск проблем. Поэтому сейчас большое значение имеет объединение систем управления бурением с приборами контроля и каротажа прямо во время работы. Это позволяет собирать, обрабатывать и использовать нужную информацию в реальном времени.

Такой подход помогает следить за параметрами бурения, контролировать направление ствола и оценивать свойства пород одновременно, используя общую информационную базу. Благодаря этому можно точнее управлять процессом, сделать работу более надежной и повысить эффективность всего строительства.

Целью данной работы является анализ особенностей интеграции автоматизированной системы управления бурением с приборами измерения и каротажа во время бурения, а также выявление преимуществ данного подхода по сравнению с традиционными и частично автоматизированными решениями.

Автоматизированная система управления бурением (АСУБ) представляет собой программно-аппаратный комплекс, предназначенный для

контроля, анализа и корректировки параметров бурения в режиме реального времени. Система функционирует по принципу замкнутого контура управления, при котором данные, поступающие от датчиков, используются не только для визуализации, но и для формирования рекомендаций либо автоматического изменения режима работы оборудования [1].

Система работает по принципу замкнутого контура управления с элементами искусственного интеллекта, собирая данные от сотен датчиков по всей буровой установке и передавая их на центральный контроллер и в облачное хранилище [2].

Интеграция АСУБ с телеметрическими системами MWD/LWD происходит через единую информационную инфраструктуру. Компоновка оснащается модулями телеметрии (угол наклона, азимут, нагрузка на долото, вибрация, давление) и каротажа во время бурения (гамма-каротаж, удельное сопротивление, плотность пород, нейтронный каротаж).

Интеграция АСУБ с телеметрическими системами MWD/LWD реализуется через единую информационную инфраструктуру. Компоновка низа бурильной колонны оснащается модулями телеметрии, измеряющими угол наклона, азимут, нагрузку на долото, вибрацию, давление и другие параметры, а также приборами каротажа во время бурения, регистрирующими гамма-излучение, удельное электрическое сопротивление, плотность пород и нейтронные характеристики разреза.

Данные с забойных датчиков передаются на поверхность по одному из каналов телеметрии:

- гидравлический;
- электромагнитный;
- акустический.

На поверхности информация поступает в периферийный контроллер буровой, проходит первичную очистку от помех и частичную обработку, а затем полностью загружается в центральный сервер АСУБ [3].

На поверхности информация поступает в периферийный контроллер буровой установки, где проходит первичную фильтрацию, очистку от помех и предварительную обработку, после чего загружается в центральный сервер АСУБ.

Алгоритмы обработки объединяют телеметрические данные с поверхностными параметрами бурения:

- механическая скорость проходки;
- крутящий момент;
- осевая нагрузка на долото;
- расход промывочной жидкости;
- давление в циркуляционной системе;
- частота вращения бурильной колонны.

Результатом является формирование единого интерфейса на рабочем месте бурильщика, где в режиме реального времени отображаются как технологические параметры, так и геофизические характеристики проходимого интервала.

Система сравнивает текущие данные с проектными и эталонными значениями, выявляет отклонения, прогнозирует возможные осложнения и формирует рекомендации по изменению режима бурения. В более развитых конфигурациях возможна частичная автоматическая корректировка режимов - изменения нагрузки на долото, частоты вращения и параметров промывки [1, 2].

Для оценки эффективности рассматриваемого подхода целесообразно провести сравнение интегрированной АСУБ с двумя наиболее распространенными вариантами организации бурового контроля:

- традиционный подход с отдельным использованием данных;
- частично автоматизированный подход с наличием мониторинга, но без полноценной интеграции в единый контур управления;
- интегрированная АСУБ с приборами MWD/LWD.

При традиционном подходе данные о режиме бурения и геофизических характеристиках поступают отдельно, а их анализ проводят операторы и инженеры по мере необходимости. Этот способ проще внедрить, но отличается меньшей оперативностью. Решения обычно принимают на основе отдельных показателей без тесной связи между технологическими и геологическими данными. Из-за этого сильно влияет человеческий фактор и часто задерживается реакция на изменения условий в забое.

Частично автоматизированные системы дают более полное отображение и учет данных. В них могут использоваться разные программные модули для показа телеметрии, параметров бурения и каротажа, но автоматическая связка этих данных зачастую работает плохо. Такие системы помогают лучше информировать персонал, но не всегда способны эффективно поддерживать принятие решений в режиме реального времени.

Интегрированная АСУБ вместе с MWD/LWD работает так, что все данные о процессе бурения и свойствах пластов собираются и обрабатываются в одной системе. Это позволяет учитывать связь между скоростью проходки, нагрузкой на долото, вибрациями, параметрами промывочной жидкости, траекторией скважины и геофизическими характеристиками. По сравнению с другими системами, такая модель быстрее анализирует информацию, помогает лучше контролировать процесс бурения и создает условия для своевременного выявления возможных проблем.

У АСУБ как у интегрированной автоматизированной системы управления бурением можно выделить аналоги не по брендам, а по типу решений. Именно данное сравнение представлено в таблице 1.1.

Таблица 1.1 - Сравнительная характеристика подходов к организации бурового контроля

Критерий сравнения	Традиционный подход с раздельным использованием данных	Частично автоматизированный подход	Интегрированная АСУБ с приборами MWD/LWD
Способ получения данных	Данные поступают из разных источников разрозненно	Данные частично собираются в цифровом виде	Данные объединяются в едином информационном контуре
Оперативность обработки информации	Низкая	Средняя	Высокая
Уровень автоматизации	Минимальный	Частичный	Высокий
Интеграция технологических и геофизических данных	Практически отсутствует	Ограниченная	Полная или близкая к полной
Поддержка принятия решений	Выполняется оператором вручную	Частично поддерживается программными средствами	Осуществляется автоматически или в режиме интеллектуальных рекомендаций
Влияние человеческого фактора	Высокое	Среднее	Сниженное
Точность управления траекторией скважины	Средняя	Выше средней	Высокая
Возможность раннего выявления осложнений	Ограниченная	Частичная	Высокая
Скорость реакции на изменение условий бурения	Низкая	Средняя	Высокая
Вероятность непроизводительного времени	Повышенная	Умеренная	Сниженная
Устойчивость бурового процесса	Ниже	Средняя	Выше
Накопление и повторное использование данных	Ограниченное	Частичное	Системное, с формированием базы знаний

Сравнительный анализ показывает, что интегрированная система обладает следующими преимуществами:

- более высокой скоростью обработки и интерпретации данных;
- снижением влияния человеческого фактора на принятие решений;
- возможностью комплексного учета технологических и геолого-геофизических параметров;

- более точным удержанием ствола скважины в заданном интервале;
- уменьшением непроизводительного времени;
- повышением предсказуемости бурового процесса [4].

Если традиционные и частично автоматизированные решения в основном направлены на контроль и отображение информации, то интегрированная АСУБ выполняет также функцию интеллектуальной поддержки управления. Именно это позволяет рассматривать ее как более совершенный аналог по отношению к системам, ограниченным только мониторингом [5].

В результате выполненного анализа установлено, что интеграция автоматизированной системы управления бурением с приборами измерения и каротажа во время бурения является эффективным направлением совершенствования технологии строительства скважин.

Проведенная сравнительная характеристика показала, что по сравнению с традиционными и частично автоматизированными решениями интегрированная система обладает рядом существенных преимуществ. К ним относятся комплексность обработки данных, высокая оперативность управления, возможность объединения технологической и геофизической информации в едином контуре, снижение влияния человеческого фактора и повышение точности принятия решений в реальном времени.

Использование интегрированной АСУБ позволяет повысить скорость и устойчивость бурения, улучшить геонавигацию, сократить непроизводительное время и снизить вероятность возникновения осложнений и аварийных ситуаций. Дополнительным преимуществом является накопление цифровой базы знаний, которая может применяться для оптимизации бурения на последующих объектах.

Представленное взаимодействие АСУБ с забойным оборудованием предоставляет следующие преимущества:

- повышение коммерческой скорости бурения до;
- увеличение точности проводки скважины (геонавигация);
- накопление базы знаний (все данные сохраняются и используются для бурения следующих скважин).

Таким образом, совместная работа АСУБ с приборами телеметрии и каротажа во время бурения полностью направлены на оптимизацию процесса бурения, а также на достижения максимальной минимизации травм и аварий на производстве.

Список литературы:

1. Дмитриев А.Ю. Основы технологии бурения скважин: учебное пособие. - Томск: Издательство Томского политехнического университета, 2008. - 216 с.
2. Храменков, В. Г. Автоматизация управления технологическими процессами бурения нефтегазовых скважин : учебное пособие для вузов / В. Г. Храменков. — Москва : Издательство Юрайт, 2024. — 415 с. — (Высшее образование). — ISBN 978-5-534-00854-8.

— Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/537112>

3. Фомин А.Н. Основы бурения нефтяных и газовых скважин: учебно-методическое пособие. - Новосибирск: ИПЦ НГУ, 2021. - 196 с

4. Бродов Г.С., Шелковников И.Г., Егоров Э.К. Технологические измерения и автоматизация процесса бурения: учебное пособие. - Санкт-Петербург: Издательство СПГГИ, 2004. - 105 с.

5. Кульчицкий В.В. Супервайзинг строительства нефтяных и газовых скважин: практическое руководство. - Москва: Издательство «Вече», 2019. - 368 с.

УДК 004.946:378.147

Романькова Е. С., Перминов Н. И.

РТУ МИРЭА – Российский технологический университет, Москва, Россия

ИСПОЛЬЗОВАНИЕ ДОПОЛНЕННОЙ РЕАЛЬНОСТИ В УЧЕБНОМ ПРОЦЕССЕ

Аннотация: в статье рассматривается применение технологии дополненной реальности в учебном процессе. Дополненная реальность позволяет совмещать реальные объекты и цифровой контент: 3D-модели, текстовые подсказки, анимации и интерактивные элементы. Это делает обучение более наглядным, практико-ориентированным и удобным для восприятия сложных тем. Особое внимание уделяется возможностям использования AR в образовательных дисциплинах, а также разработке учебных приложений на базе Unity и AR Foundation. Делается вывод, что дополненная реальность не заменяет традиционное обучение, а усиливает его за счет визуализации, интерактивности и повышения мотивации обучающихся.

Ключевые слова: дополненная реальность, учебный процесс, Unity, AR Foundation, цифровая образовательная среда, интерактивное обучение, визуализация, смешанная реальность.

Введение

Современное образование развивается в условиях активной цифровизации. В учебном процессе все чаще используются электронные образовательные ресурсы, онлайн-платформы, интерактивные тренажеры, мобильные приложения и технологии виртуальной или дополненной реальности. Одним из наиболее перспективных направлений является дополненная реальность, позволяющая накладывать цифровую информацию на объекты реального мира в режиме реального времени [1].

Актуальность применения AR связана с тем, что многие учебные темы требуют высокой степени наглядности. При изучении физики, химии, биологии, инженерных дисциплин, истории или информатики обучающимся часто необходимо представить объект или процесс, который сложно наблюдать непосредственно. Обычные схемы и изображения не всегда позволяют понять структуру или динамику явления. Дополненная реальность решает эту проблему за счет интерактивных трехмерных моделей, которые можно рассматривать, вращать, масштабировать и дополнять пояснениями.

Использование AR также соответствует задачам формирования цифровой образовательной среды, направленной на внедрение современных цифровых сервисов, образовательного контента и технических средств обучения [2]. Цель статьи — рассмотреть особенности применения дополненной реальности в учебном процессе, определить ее преимущества, ограничения и возможные сценарии использования.

Понятие дополненной реальности и ее образовательное значение

Дополненная реальность — это технология, при которой реальное пространство дополняется цифровыми объектами. Пользователь видит окружающую среду через камеру смартфона, планшета или AR-очков, а поверх изображения отображаются 3D-модели, текст, схемы, анимации или другие элементы. В отличие от виртуальной реальности, AR не заменяет реальный мир полностью, а расширяет его дополнительной информацией.

В образовании дополненная реальность выполняет несколько функций. Она повышает наглядность учебного материала, позволяет моделировать сложные процессы, усиливает интерес обучающихся и помогает организовать практико-ориентированное обучение. В исследованиях отмечается, что AR может использоваться как средство визуализации учебной информации и расширения возможностей мобильного обучения [3].

Таким образом, AR-приложение связывает физическую среду, устройство пользователя и цифровой учебный контент. Это позволяет обучающемуся не только видеть объект, но и взаимодействовать с ним.

Направления применения AR в обучении

Дополненная реальность может применяться в разных учебных дисциплинах. Наибольший эффект она дает там, где важно пространственное мышление, визуализация или имитация практических действий.

Таблица 1 — Примеры применения AR в учебном процессе

Направление	Пример применения	Эффект
Биология и анатомия	3D-модели органов, клетки, скелета	Повышение наглядности
Физика	Визуализация сил, полей, траекторий	Понимание невидимых процессов
Химия	Модели молекул и кристаллических решеток	Развитие пространственного мышления
История	Реконструкция зданий и артефактов	Эффект присутствия
География	Объемные карты и модели рельефа	Связь карты с реальным пространством
Инженерные дисциплины	Изучение деталей и механизмов	Практико-ориентированное обучение
Информатика	Визуализация алгоритмов и сетей	Объяснение абстрактных понятий

Например, при изучении строения сердца студент может навести камеру на изображение в учебнике и увидеть объемную модель органа с подписями. При изучении физики AR может показывать направление сил, движение тела или распространение волны. В инженерных дисциплинах AR помогает рассматривать механизмы, детали и оборудование в трехмерном виде.

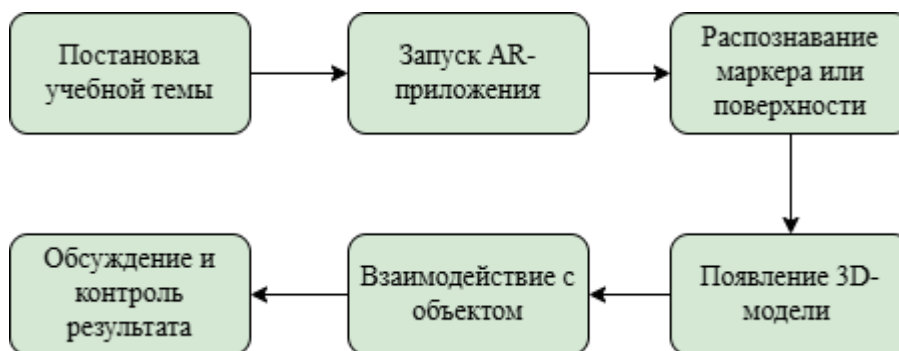


Рисунок 1 — Сценарий использования AR на занятии

Такой сценарий показывает, что AR может быть встроена в обычное занятие без полной перестройки учебного процесса.

Преимущества и ограничения дополненной реальности

Главное преимущество AR заключается в повышении наглядности. Многие учебные объекты сложно представить по текстовому описанию или плоской схеме. Дополненная реальность позволяет показать их в объеме и динамике. Также AR повышает мотивацию обучающихся, поскольку делает занятие более интерактивным и современным [4].

Дополнительным преимуществом является безопасность. Некоторые процессы невозможно продемонстрировать в аудитории из-за высокой стоимости, сложности или риска. Например, AR может использоваться для имитации работы промышленного оборудования, химического опыта или медицинской процедуры. Обучающийся получает возможность изучать объект без опасности для себя и оборудования.

Однако у AR есть ограничения. Для работы приложений нужны подходящие устройства, качественные 3D-модели и методически продуманные задания. Если AR используется только как визуальный эффект, без связи с учебной целью, ее образовательная ценность снижается.

Таблица 2 — Преимущества и ограничения AR

Параметр	Преимущества	Ограничения
Наглядность	3D-объекты и анимации	Требуется качественный контент
Мотивация	Повышение интереса	Возможен эффект новизны без пользы
Практика	Имитация действий	Не всегда заменяет реальный опыт
Доступность	Работа на смартфонах	Не все устройства поддерживают AR
Безопасность	Изучение сложных процессов без риска	Нужна проверка корректности моделей

Следовательно, AR должна применяться не ради самой технологии, а для решения конкретной образовательной задачи.

Unity как инструмент разработки учебных AR-приложений

Unity является одним из популярных инструментов разработки AR-приложений. Для создания приложений дополненной реальности используется пакет AR Foundation, который позволяет работать с ARCore и ARKit и создавать кроссплатформенные проекты [5].

Для студенческих проектов Unity удобен тем, что поддерживает работу с 3D-моделями, анимациями, физикой, пользовательским интерфейсом и языком C#. С его помощью можно создать прототип учебного AR-приложения, например, интерактивную модель механизма, лабораторной установки или исторического объекта.

Таблица 3 — Основные компоненты учебного AR-приложения на Unity

Компонент	Назначение
AR Session	Управление AR-сессией
XR Origin / AR Session Origin	Связь виртуальной сцены с реальным пространством
AR Camera	Отображение камеры и цифровых объектов
Plane Detection	Поиск горизонтальных и вертикальных поверхностей
Image Tracking	Распознавание учебных маркеров
Raycast Manager	Размещение объектов по касанию
UI Canvas	Кнопки, подсказки, задания
C# Scripts	Логика взаимодействия пользователя с объектами

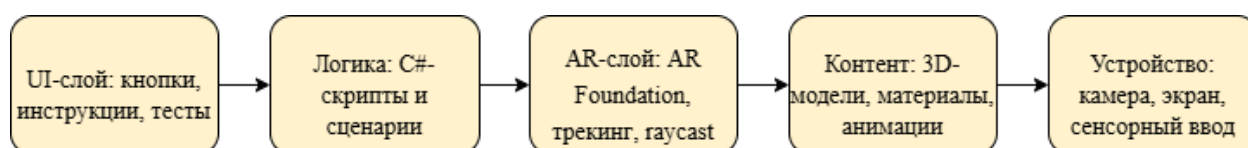


Рисунок 2 — Упрощенная архитектура AR-приложения на Unity

Например, студент может создать AR-приложение для изучения строения электродвигателя. Пользователь размещает 3D-модель двигателя на поверхности, включает подписи к деталям, запускает анимацию вращения ротора и проходит небольшой тест по устройству механизма.

Методика внедрения AR в учебный процесс

Для эффективного использования AR необходимо встроить технологию в структуру занятия. Преподаватель должен определить, какую проблему решает дополненная реальность: недостаток наглядности, сложность демонстрации, отсутствие оборудования или необходимость самостоятельной тренировки.

Таблица 4 — Этапы внедрения AR в занятие

Этап	Содержание
Анализ темы	Выбор темы, где нужна визуализация
Постановка цели	Определение учебного результата
Подбор сценария	Демонстрация, лабораторная работа или тренажер
Подготовка контента	Создание моделей, подписей, заданий

Этап	Содержание
Разработка приложения	Реализация функций в Unity
Проведение занятия	Работа обучающихся с AR-моделью
Оценка результата	Тест, опрос или практическое задание

Наиболее эффективно использовать AR не на всем занятии, а в тех его частях, где она действительно усиливает объяснение. Например, сначала преподаватель дает теоретическую основу, затем обучающиеся изучают AR-модель, после чего выполняют задание и обсуждают результат.

Пример учебного AR-проекта

В рамках дисциплины по моделированию приложений смешанной и дополненной реальности можно разработать прототип приложения «AR-лаборатория». Например, тема проекта — «AR-приложение для изучения строения электродвигателя».

Пользователь запускает приложение, наводит камеру на маркер или поверхность, после чего на экране появляется 3D-модель электродвигателя. Модель можно вращать, масштабировать, разбирать на части, просматривать подписи и запускать анимацию работы. В конце пользователь проходит короткий тест.

Таблица 5 — Возможности учебного AR-приложения

Функция	Назначение
Размещение модели	Связь объекта с реальным пространством
Вращение и масштабирование	Рассмотрение объекта с разных сторон
Подписи к деталям	Изучение терминологии
Анимация работы	Понимание принципа действия
Режим разборки	Изучение структуры механизма
Мини-тест	Проверка усвоения материала

Такой проект реалистичен для студента 3 курса, поскольку сочетает программирование, 3D-визуализацию, интерфейс и образовательный сценарий, но не требует разработки крупной системы.

Заключение

Дополненная реальность является перспективным инструментом современного образования. Она позволяет объединить реальные объекты и цифровой учебный контент, что делает обучение более наглядным, интерактивным и практико-ориентированным. Особенно эффективно AR применяется в дисциплинах, где важно показать пространственные объекты, сложные процессы или явления, которые трудно наблюдать напрямую.

Использование AR способствует повышению мотивации обучающихся, развитию пространственного мышления и безопасному моделированию практических действий. Однако технология должна применяться методически

обоснованно. Она не заменяет преподавателя и традиционные формы обучения, а дополняет их.

Unity и AR Foundation являются удобными средствами разработки учебных AR-приложений. Они позволяют создавать студенческие прототипы, работать с 3D-моделями, реализовывать интерактивность и адаптировать приложения под мобильные устройства. Таким образом, дополненная реальность может стать важным элементом цифровой образовательной среды, если она используется для решения конкретных учебных задач.

Список литературы:

1. Филимоненкова Т. Н. Дополненная реальность как инновационная технология образовательного процесса. [Электронный ресурс] Режим доступа: <https://cyberleninka.ru/article/n/dopolnennaya-realnost-kak-innovatsionnaya-tehnologiya-obrazovatel'nogo-protsesta> (дата обращения: 10.05.2026)
2. Министерство просвещения Российской Федерации. Федеральный проект «Цифровая образовательная среда». [Электронный ресурс] Режим доступа: <https://edu.gov.ru/national-project/projects/cos/> (дата обращения: 10.05.2026)
3. Данилова Т.В. Использование виртуальной и дополненной реальности в учебном процессе высших учебных заведений // Education Management Review. 2024. [Электронный ресурс] Режим доступа: <https://emreview.ru/index.php/emr/article/view/1301> (дата обращения: 10.05.2026)
4. Применение технологий дополненной реальности в образовательном процессе. [Электронный ресурс] Режим доступа: <https://www.in-academy.uz/index.php/ZDPP/article/view/19170> (дата обращения: 11.05.2026)
5. Unity. Введение в дополненную реальность (AR). [Электронный ресурс] Режим доступа: <https://unity.com/ru/resources/introduction-to-augmented-reality> (дата обращения: 11.05.2026)

УДК 004.7:004.946

Романькова Е. С., Перминов Н. И.

РТУ МИРЭА – Российский технологический университет, Москва, Россия

ОРГАНИЗАЦИЯ P2P-ВЗАИМОДЕЙСТВИЯ МЕЖДУ КЛИЕНТАМИ В UNITY-ПРИЛОЖЕНИИ

Аннотация: в статье рассматриваются принципы организации P2P-взаимодействия между клиентами в Unity-приложении. Особое внимание уделяется сетевой архитектуре, выбору транспортного протокола, проблемам NAT, механизмам установления соединения и синхронизации состояния объектов. Рассматривается применимость P2P-модели для интерактивных приложений, в том числе VR/AR-сред, где критически важны низкая задержка и согласованность действий пользователей. В результате формулируется обобщенная архитектура P2P-взаимодействия, пригодная для реализации в учебном Unity-проекте.

Ключевые слова: P2P-взаимодействие; Unity-приложение; сетевое взаимодействие; транспортный протокол; UDP; NAT; STUN; TURN; синхронизация объектов; VR/AR-среда.

Введение

Развитие виртуальной, дополненной и смешанной реальности привело к росту интереса к многопользовательским интерактивным приложениям. Если одиночное VR- или AR-приложение решает задачу индивидуального взаимодействия пользователя с цифровой средой, то многопользовательское приложение позволяет нескольким пользователям одновременно находиться в общей сцене, видеть действия друг друга, обмениваться сообщениями и совместно управлять объектами. Для Unity это особенно актуально, поскольку данный игровой движок широко применяется не только в игровой индустрии, но и в образовательных симуляторах, инженерных тренажерах, виртуальных лабораториях и прототипах XR-систем.

Unity развивает отдельное направление решений для сетевого взаимодействия и NetCode. На официальной русскоязычной странице Unity указывается, что Netcode for GameObjects предназначен для разработки кооперативных многопользовательских игр и позволяет синхронизировать сцены и данные GameObject между клиентами и платформами [1]. Кроме того, Unity Transport рассматривается как низкоуровневая транспортная библиотека, ориентированная на производительность и надежность, дополняющая UDP расширенными возможностями [1]. Это показывает, что сетевое взаимодействие в Unity может строиться как на высокоуровневых инструментах синхронизации объектов, так и на более низкоуровневой передаче пользовательских сообщений.

Актуальность темы P2P-взаимодействия связана с тем, что не каждое учебное или прототипное Unity-приложение требует полноценной выделенной серверной инфраструктуры. В ряде случаев достаточно организовать прямой

обмен данными между клиентами: например, для совместного просмотра 3D-модели, обмена координатами пользователей, передачи коротких сообщений или синхронизации простых действий в виртуальной среде. Такая модель потенциально снижает задержки, уменьшает нагрузку на серверную часть и упрощает архитектуру небольших приложений. Однако P2P-подход имеет и существенные ограничения: необходимость установления прямого соединения, проблемы прохождения NAT, отсутствие единого авторитетного сервера, повышенные требования к безопасности и сложности синхронизации состояния.

Целью данной статьи является анализ принципов организации P2P-взаимодействия между клиентами в Unity-приложении и формирование обобщенной архитектуры, которую можно использовать при разработке учебного многопользовательского проекта.

P2P, или peer-to-peer, представляет собой модель сетевого взаимодействия, при которой участники сети обмениваются данными напрямую, без постоянной зависимости от центрального сервера. В отличие от клиент-серверной архитектуры, где сервер принимает запросы, хранит состояние и рассылает данные клиентам, в P2P-модели каждый участник может одновременно выступать и отправителем, и получателем данных.

Для Unity-приложений это означает, что несколько экземпляров приложения, запущенных на разных устройствах, могут напрямую обмениваться состоянием виртуальной сцены. Например, один клиент может передавать другому координаты аватара, поворот камеры, событие захвата объекта, изменение цвета модели или текстовое сообщение. В простейшем учебном проекте такая схема может быть реализована через два экземпляра Unity-приложения, где каждый клиент имеет локальный порт для приема сообщений и удаленный адрес с портом второго клиента.

Следует различать чистую P2P-модель и гибридную модель. В чистом варианте сервер не участвует в соединении между клиентами. Однако на практике такой подход применяется ограниченно, поскольку клиентам необходимо узнать сетевые адреса друг друга, согласовать параметры соединения и преодолеть сетевые ограничения. Поэтому часто используется гибридная архитектура: сервер применяется для сигнализации, регистрации пользователей или подбора участников, а основные пользовательские данные передаются напрямую между клиентами.

В Unity P2P-взаимодействие может быть организовано несколькими способами. Первый вариант — использование высокоуровневых сетевых решений, например, Netcode for GameObjects, которые предоставляют готовые механизмы синхронизации объектов. Официальный блог Unity указывает, что Netcode for GameObjects является библиотекой сетевого кода среднего уровня, предназначенной для игрового движка Unity [2]. Второй вариант — самостоятельная реализация обмена сообщениями через UDP или другой

транспортный протокол, что дает разработчику больший контроль над сетевой логикой.

Для интерактивных приложений реального времени ключевым параметром является задержка передачи данных. В VR/AR-приложениях высокая задержка приводит к рассинхронизации действий пользователей, снижению эффекта присутствия и ухудшению качества взаимодействия. Поэтому для передачи часто обновляемых данных, таких как координаты, повороты и короткие события, обычно используется UDP.

UDP не устанавливает соединение перед передачей данных и не требует подтверждения получения каждого пакета. Это снижает задержку, но одновременно означает, что протокол не гарантирует доставку, порядок получения и отсутствие потерь. В русскоязычном материале Servercore отмечается, что UDP отличается низкой задержкой, но не гарантирует получение данных [3]. Следовательно, при разработке Unity-приложения необходимо учитывать возможность потери отдельных сообщений.

В P2P-приложении по UDP целесообразно передавать данные, потеря которых не является критичной. Например, если клиент отправляет координаты аватара 20–30 раз в секунду, потеря одного пакета обычно не нарушает работу приложения, поскольку следующее обновление быстро восстановит актуальное состояние. Однако для важных событий, таких как подключение пользователя, захват объекта или завершение задания, требуется использовать дополнительные механизмы надежности: подтверждения, повторную отправку, идентификаторы сообщений или отдельный надежный канал.

С практической точки зрения сетевые сообщения можно разделить на три группы. Первая группа — частые ненадежные обновления состояния: позиция, вращение, скорость, направление взгляда. Вторая группа — редкие надежные события: подключение, отключение, выбор объекта, изменение режима сцены. Третья группа — служебные сообщения: ping, проверка доступности, согласование версии протокола и обмен параметрами сессии. Такое разделение позволяет снизить нагрузку на сеть и сохранить корректность важных действий.

Одной из главных проблем P2P-взаимодействия является наличие NAT. NAT позволяет нескольким устройствам в локальной сети использовать один публичный IP-адрес, но затрудняет входящие подключения извне. Поэтому два клиента, находящиеся в разных сетях, не всегда могут напрямую обмениваться пакетами без дополнительной процедуры согласования.

Для решения этой проблемы применяются механизмы STUN, TURN и ICE. В материале TrueConf STUN определяется как протокол, позволяющий клиенту определить свой публичный IP-адрес и порт, отображаемые устройством NAT [4]. STUN помогает участнику понять, каким образом он виден во внешней сети, и тем самым облегчает установление прямого соединения.

TURN используется в тех случаях, когда прямое соединение установить невозможно. В этом случае данные передаются через сервер-ретранслятор.

Формально такая схема уже не является полностью прямым P2P-обменом, но с точки зрения приложения она позволяет сохранить модель взаимодействия между участниками. В руководстве «WebRTC для любопытных» описывается роль TURN-сервера в процессе подключения и передаче данных между клиентами [5]. ICE, в свою очередь, применяется для выбора наиболее подходящего маршрута соединения между участниками после обмена сетевой информацией [6].

В учебном Unity-проекте задачу можно упростить. Например, если две копии приложения запускаются на одном компьютере или в одной локальной сети, достаточно вручную задать локальный и удаленный порты. Первый клиент может принимать сообщения на порту 7777 и отправлять их на порт 7778, а второй — принимать на порту 7778 и отправлять на порт 7777. Такая схема демонстрирует принцип P2P-взаимодействия без сложной инфраструктуры NAT traversal. Однако для работы через интернет потребуются сигнализация, STUN, TURN или аналогичные механизмы.

Архитектура P2P-взаимодействия в Unity-приложении может быть представлена как совокупность нескольких уровней: транспортного, протокольного, прикладного и уровня синхронизации сцены. Транспортный уровень отвечает за отправку и прием пакетов, например, через UDP-сокеты. Протокольный уровень определяет формат сообщений: тип, идентификатор отправителя, временную метку и полезную нагрузку. Типами сообщений могут быть HELLO, PING, STATE_UPDATE, OBJECT_GRAB, OBJECT_RELEASE, CHAT_MESSAGE.

Прикладной уровень отвечает за обработку полученных сообщений. Если пришли координаты удаленного пользователя, Unity-компонент обновляет позицию его аватара. Если произошло событие захвата объекта, приложение изменяет владельца объекта и блокирует управление им со стороны других пользователей. Уровень синхронизации сцены обеспечивает согласованность состояния между клиентами.

Главная сложность синхронизации в P2P-модели состоит в отсутствии авторитетного сервера. Поэтому необходимо определить, кто имеет право принимать окончательное решение о состоянии объекта. Возможны разные подходы: каждый клиент отвечает за свой аватар; объектом управляет клиент-владелец; один из клиентов временно выполняет роль хоста. Для учебного проекта наиболее простым является подход с авторитетом владельца: пользователь управляет своим аватаром и временно становится владельцем объекта, который он захватил.

При синхронизации сцены нецелесообразно передавать все состояние Unity-проекта целиком. Передаваться должны только изменяющиеся параметры: позиция, поворот, состояние анимации, идентификатор объекта и идентификатор владельца. Если данные приходят с задержкой или нерегулярно, прямое присваивание позиции может вызывать рывки. Поэтому обычно

применяется интерполяция, при которой объект плавно перемещается к полученному положению.

Также необходимо учитывать частоту отправки сообщений. Передача данных каждый кадр может создать избыточную нагрузку на сеть, поэтому используется сетевой тик — например, 10, 20 или 30 обновлений в секунду. Для простого учебного AR-приложения может быть достаточно 10 обновлений в секунду, а для динамичного VR-взаимодействия требуется более высокая частота.

Несмотря на преимущества, P2P-модель имеет ряд ограничений. Во-первых, отсутствие центрального сервера усложняет проверку корректности действий пользователей. Каждый клиент потенциально может отправить произвольные данные, поэтому входящие сообщения необходимо валидировать. Например, приложение должно игнорировать неизвестные типы сообщений, проверять существование объектов и ограничивать недопустимо резкие перемещения.

Во-вторых, P2P плохо масштабируется. Такая модель подходит для малых групп, например, двух–четырёх участников. При увеличении числа пользователей возрастает количество соединений и объем передаваемых данных. В-третьих, необходимо обрабатывать нестабильность соединения: отключение клиента, потерю сети или смену адреса. Для этого можно использовать heartbeat-механизм, при котором клиенты периодически отправляют короткие сообщения доступности.

Типовой сценарий P2P-взаимодействия в Unity-приложении включает несколько этапов. Сначала пользователь запускает приложение и указывает локальный порт, IP-адрес удаленного клиента и удаленный порт. Затем клиенты обмениваются служебными сообщениями HELLO и HELLO_ACK, после чего соединение считается установленным. Далее начинается обмен пользовательскими данными: координатами аватара, поворотом камеры, состоянием контроллеров или событиями взаимодействия с объектами. При завершении работы клиент отправляет сообщение DISCONNECT, а при аварийном отключении второй клиент определяет потерю связи по отсутствию heartbeat-сообщений.

В VR/AR-приложениях P2P-взаимодействие может использоваться для организации совместного присутствия пользователей в одной цифровой среде. Например, два пользователя могут видеть одну 3D-модель, поворачивать ее, добавлять пометки или выполнять совместное учебное задание. В AR-приложениях особенно важно обеспечить согласованность координатных систем, поскольку каждый пользователь видит реальное пространство со своей камеры. Для этого может использоваться общий маркер, согласование якорей сцены или ручная настройка общего положения виртуального объекта.

Таким образом, P2P-взаимодействие в Unity-приложении является эффективным решением для небольших многопользовательских прототипов,

учебных проектов и простых XR-сцен. Его преимущества заключаются в потенциально низкой задержке, отсутствии обязательного выделенного сервера и относительной простоте локальной реализации. Однако для работы через интернет и масштабирования на большое число пользователей необходимо учитывать ограничения NAT, вопросы безопасности, потерю пакетов и необходимость синхронизации состояния между клиентами.

В ходе анализа было показано, что P2P-взаимодействие требует решения нескольких ключевых задач: выбора транспортного протокола, определения формата сообщений, синхронизации объектов сцены, обработки потерь пакетов, контроля состояния соединения и учета сетевых ограничений, связанных с NAT. Для локальных учебных проектов допустима упрощенная схема с ручным указанием IP-адресов и портов, тогда как для работы через интернет необходимы дополнительные механизмы сигнализации, STUN, TURN или ICE.

Таким образом, P2P-подход является эффективным решением для малых многопользовательских Unity-приложений, демонстрационных XR-сцен и учебных проектов. При этом его применение должно сопровождаться пониманием ограничений: слабой масштабируемости, сложности прохождения NAT, отсутствия центрального контроля и необходимости дополнительной проверки входящих данных. Наиболее рациональной для практических приложений является гибридная архитектура, в которой сервер используется для сигнализации и служебной координации, а основная часть интерактивных данных передается между клиентами напрямую.

Список литературы:

1. Unity. Решения для сетевого взаимодействия и NetCode. [Электронный ресурс] Режим доступа: <https://unity.com/ru/features/netcode> (дата обращения: 08.05.2026)
2. Unity Blog. Создание моей первой многопользовательской игры с Netcode for GameObjects. [Электронный ресурс] Режим доступа: <https://unity.com/ru/blog/games/building-a-multiplayer-game-first-time-netcode-for-gameobjects> (дата обращения: 08.05.2026)
3. Servercore. TCP и UDP протоколы — как устроены, принцип работы и в чем разница. [Электронный ресурс] Режим доступа: <https://servercore.com/ru/blog/articles/protokoly-tcp-i-udp/> (дата обращения: 08.05.2026)
4. TrueConf. Серверы TURN и STUN. [Электронный ресурс] Режим доступа: <https://trueconf.ru/blog/wiki/servery-turn-i-stun> (дата обращения: 09.05.2026)
5. WebRTC для любопытных. Подключение. [Электронный ресурс] Режим доступа: <https://webrtcforthe curious.com/ru/docs/03-connecting/> (дата обращения: 10.05.2026)
6. Habr / Timeweb. WebRTC для всех и каждого. Часть 1. [Электронный ресурс] Режим доступа: <https://habr.com/ru/companies/timeweb/articles/656947/> (дата обращения: 10.05.2026)

УДК 004.8:159.9

Кулиева Байрамгуль, Рахыев Довран, Керимов Гуйчгельди
Институт инженерно-технических и транспортных коммуникаций Туркменистана,
Ашхабад, Туркменистан

ТЕОРИЯ КОГНИТИВНОГО ИИ И МОДЕЛИРОВАНИЕ ЧЕЛОВЕЧЕСКОГО МЫШЛЕНИЯ

Аннотация. Проблема моделирования человеческого мышления методами искусственного интеллекта (ИИ) остается одной из центральных в современной когнитивной науке и лингводидактике. В отличие от узкоспециализированных систем, когнитивный ИИ стремится воспроизвести универсальные механизмы познания: способность к обучению, рассуждению в условиях неопределенности, адаптации к новым контекстам и пониманию естественного языка. Однако теоретические основания такого моделирования остаются дискуссионными. Когнитивный ИИ реализуется в двух основных парадигмах: (1) архитектурный подход – построение формальных систем мышления, таких как NARS (Non-Axiomatic Reasoning System), основанный на неаксиоматической логике и способный к обучению в условиях неопределенности; (2) эмпирический подход – обучение больших языковых моделей на массивах психологических данных, что позволяет имитировать когнитивные стили и «человеческие» стратегии принятия решений, включая когнитивные искажения и эвристики. Показано, что наиболее эффективные современные когнитивные модели, такие как Centaur, демонстрируют способность предсказывать поведение человека в 160 различных экспериментальных задачах (с точностью до 20% выше классических теорий) и даже генерировать новые когнитивные гипотезы, в то время как альтернативные архитектуры (NARS, ACT-R) успешно моделируют произвольное реляционное реагирование, операантное обучение и формирование абстрактных отношений по принципу «same-opposite».

Ключевые слова: когнитивный искусственный интеллект, моделирование человеческого мышления, когнитивная архитектура, неаксиоматическая логика, NARS, ACT-R, Centaur, машинная психология, когнитивный техносимбиоз, рассуждение в условиях неопределенности.

Введение. Проблема создания искусственного интеллекта, способного мыслить подобно человеку, восходит к самым истокам компьютерной науки. Однако, в отличие от традиционного ИИ, нацеленного на сверхчеловеческую производительность в узких задачах, когнитивный ИИ фокусируется на другом: понять и воспроизвести универсальные механизмы человеческого познания – способность к обучению, рассуждению в условиях неопределенности, пониманию естественного языка, адаптации к новым контекстам и даже проявлению когнитивных искажений и эвристик [14; 18]. Эта междисциплинарная область, объединяющая когнитивную психологию, нейронауку, философию и информатику, стремится не просто построить «умную» машину, но и пролить свет на природу самого мышления.

Актуальность исследования обусловлена несколькими ключевыми факторами. Во-первых, лавинообразное развитие больших языковых моделей (LLM) и систем искусственного интеллекта ставит вопрос о границах их «когнитивности» – действительно ли они понимают задачи и контекст или лишь эффективно подбирают статистические паттерны? Этот фундаментальный вопрос был остро поставлен в ходе критики модели Centaur, которая, несмотря на впечатляющие результаты в симуляции человеческого поведения, оказалась уязвима для тестов, демонстрирующих отсутствие истинного понимания инструкций [13; 18]. Во-вторых, формирование нового направления – «машинной психологии» (Machine Psychology) – открыло возможность интеграции принципов психологии обучения (оперантное обусловливание, реляционная теория фреймов) с когнитивными архитектурами ИИ, такими как NARS, что позволяет последовательно наращивать когнитивные способности от базовой адаптации к сложному символическому рассуждению [16; 19]. В-третьих, активизировались философские и социокультурные дискуссии о последствиях «когнитивного техносимбиоза» – сращивания человеческого мышления с технологиями, сопровождающегося рисками когнитивной зависимости и утраты самостоятельности мыслительного поиска [8]. В-четвертых, совершенствование классических когнитивных архитектур (ACT-R, Soar, CLARION) и появление новых, бионических подходов к ИИ, моделирующих химическую сигнализацию в мозге, создают основу для создания объяснимых и верифицируемых моделей мышления, в отличие от «черных ящиков» глубоких нейросетей [20; 21].

Степень изученности проблемы характеризуется одновременным существованием нескольких конкурирующих парадигм. Одна из них, представленная работами Пей Ванга и его последователей, опирается на **неаксиоматическую логику (NAL)** – формальную систему, специально разработанную для работы в условиях неопределенности и неполноты знаний. NAL и построенная на ее основе система NARS (Non-Axiomatic Reasoning System) моделируют «законы мышления», позволяя ИИ обучаться на опыте, делать выводы из противоречивой информации и адаптироваться к изменяющимся условиям без переобучения [2; 19]. Противоположный подход, воплощенный в модели **Centaur**, использует мощь больших языковых моделей, обучая их на массивах реальных психологических данных (Psych-101, содержащий более 10 миллионов выборов 60 000 участников). Такой эмпирический, «индуктивный» метод позволил Centaur достичь точности предсказания человеческого поведения, недостижимой для классических когнитивных теорий (например, теории перспектив), и даже предложить новые гипотезы о когнитивных стратегиях [12; 18]. Обзорное исследование М. Фрэнка и Н. Гудмана в Annual Review of Psychology (2026) подводит итог этой дискуссии, утверждая, что наиболее перспективным является использование

ИИ-моделей как **когнитивных моделей** – обученных на человеческих данных и протестированных с помощью точных экспериментальных зондов, что позволяет не только предсказывать, но и объяснять поведение [15]. Однако, как показали независимые исследования, у Centaur есть серьезные недостатки: модель оказалась подвержена эффекту переобучения, будучи не в состоянии правильно интерпретировать измененный контекст задачи (например, команду «выберите вариант А» она игнорировала, подставляя статистически ожидаемые ответы), что ставит под сомнение наличие у нее «истинного» понимания [13].

Целью настоящей работы является системный анализ современных подходов к когнитивному моделированию человеческого мышления на основе архитектур искусственного интеллекта, выявление их теоретических основ, лингводидактического потенциала и ограничений.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Классифицировать основные парадигмы когнитивного ИИ: архитектурную (NAL/NARS) и эмпирическую (Centaur, LLM).
2. Сравнить их лингводидактический потенциал – способность моделировать понимание языка, контекстную адаптацию и обучение.
3. Выявить философские и практические ограничения когнитивных моделей (проблема «иллюзии понимания», вычислительная сложность, когнитивный техносимбиоз).
4. Обозначить перспективы развития, включая интеграцию подходов и создание объяснимых систем.

Две парадигмы когнитивного искусственного интеллекта. Анализ современной научной литературы позволяет выделить два фундаментально различных подхода к моделированию человеческого мышления с помощью искусственного интеллекта. Первый подход, который можно назвать **архитектурным** (или дедуктивным), исходит из предположения о существовании универсальных законов мышления, которые могут быть формализованы и реализованы в виде вычислительных систем. Второй подход, **эмпирический** (или индуктивный), использует мощь больших языковых моделей и методов машинного обучения для извлечения когнитивных паттернов непосредственно из данных экспериментов с участием человека.

Архитектурный подход: неаксиоматическая логика и NARS. Классическим представителем этого направления является работа Пей Ванга, посвященная неаксиоматической логике (Non-Axiomatic Logic, NAL). В отличие от традиционной аксиоматической логики (Аристотель, Фреге, Рассел), NAL специально разработана для работы в условиях неопределенности, неполноты информации и ограниченных вычислительных ресурсов – именно в тех условиях, в которых функционирует человеческое

мышление [2]. NAL оперирует не истинностными значениями («истинно»/«ложно»), а степенями уверенности, которые динамически пересматриваются по мере поступления нового опыта.

Построенная на основе NAL система NARS (Non-Axiomatic Reasoning System) демонстрирует способность к **обучению на опыте, адаптации к изменяющимся условиям и рассуждению в условиях противоречивой информации** [2]. Ключевая особенность NARS заключается в том, что она не требует переобучения при появлении новых данных – система непрерывно обновляет свои знания и уверенности, подобно тому как это делает человек. Этот архитектурный подход имеет важные лингводидактические импликации: он открывает путь к созданию ИИ-тьюторов, способных понимать не только правильные, но и ошибочные ответы ученика, выявлять логику ошибки и адаптировать учебную траекторию без жесткого программирования.

Другие классические когнитивные архитектуры, такие как ACT-R (Adaptive Control of Thought – Rational), Soar и CLARION, хотя и не основаны на NAL, разделяют общую философию: моделирование мышления как процесса манипуляции символическими структурами под контролем продукционных правил или нейросетевых механизмов. Однако, как отмечают М. С. Frank и N. D. Goodman, эти архитектуры, несмотря на свою психологическую и нейробиологическую обоснованность, часто страдают от вычислительной сложности и ограниченной масштабируемости при попытке охватить широкий спектр когнитивных феноменов [15].

Эмпирический подход: большие языковые модели как когнитивные модели. Революционное изменение ландшафта когнитивного моделирования произошло с появлением больших языковых моделей (LLM), способных обучаться на массивах текстовых данных, включая записи психологических экспериментов. Центральным объектом дискуссий в 2025–2026 гг. стала модель **Centaur**, разработанная командой под руководством М. Бинца и опубликованная в журнале Nature [12].

Centaur была обучена на масштабном датасете Psych-101, содержащем более 10 миллионов выборов 60 000 участников, охватывающих 160 различных экспериментальных парадигм: от классических задач на принятие решений (дилемма азиатской болезни Канемана и Тверски) до задач на когнитивную рефлексивность, пространственное мышление и социальное познание. Результаты оказались впечатляющими: Centaur предсказывала человеческое поведение с точностью, на 20% превышающей классические когнитивные теории, и, что еще более важно, оказалась способна генерировать новые, ранее не зафиксированные в литературе когнитивные гипотезы, которые затем подтверждались в экспериментах [12; 18].

Однако критический анализ, проведенный независимыми исследователями, в том числе в рамках репликационного исследования, выявил серьезные ограничения. Centaur, как выяснилось, не столько

«понимает» задачи, сколько выучивает статистические закономерности тестовой выборки. В контрольном эксперименте, где инструкция была изменена на противоположную («выберите вариант А» вместо стандартной формулировки), Centaur продолжала выдавать статистически ожидаемые, но теперь уже неверные ответы, демонстрируя отсутствие истинного контекстного понимания [13; 18]. Этот феномен получил название **«иллюзии понимания»** и стал центральным аргументом противников отождествления LLM с когнитивными моделями.

Машинная психология: новый синтез. Параллельно с развитием больших языковых моделей формируется новое исследовательское направление – **машинная психология (Machine Psychology)**, которое пытается синтезировать принципы поведенческой психологии и когнитивные архитектуры ИИ. В работе Р. Дж. Юханссона и П. Хаммера представлен обзор использования системы неаксиоматического рассуждения (NARS) для моделирования «произвольно применимого реляционного реагирования» (arbitrarily applicable relational responding) – способности, лежащей в основе человеческого языка и абстрактного мышления [16].

Авторы показывают, что NARS способна обучаться операантному обусловливанию (вырабатывать реакции на основе подкрепления) и формировать реляционные фреймы (отношения «same-opposite», «больше-меньше» и т.д.) без явного программирования этих отношений. Это открывает путь к созданию ИИ-систем, которые не просто имитируют поведение, а **вырабатывают его в процессе взаимодействия со средой** – подобно тому, как это делает ребенок, овладевая родным языком. Лингводидактический потенциал такого подхода огромен: он позволяет создавать адаптивные обучающие системы, которые не только подстраиваются под уровень ученика, но и формируют когнитивные схемы, лежащие в основе грамматической и лексической компетенции.

Философские и этические ограничения: когнитивный техносимбиоз. Выход когнитивного ИИ за пределы лабораторий и его интеграция в повседневную жизнь человека ставит философские вопросы о границах, последствиях и рисках такого взаимодействия. Т. Г. Лешкевич вводит понятие **«когнитивного техносимбиоза»** – состояния, при котором человеческое мышление и искусственные интеллектуальные системы становятся взаимозависимыми, образуя единую когнитивную экосистему [8].

С одной стороны, когнитивный техносимбиоз открывает беспрецедентные возможности: ИИ-ассистенты могут компенсировать когнитивные ограничения человека (ограниченный объем рабочей памяти, подверженность искажениям, утомляемость), помогать в принятии сложных решений, ускорять обучение. С другой стороны, Лешкевич предупреждает о риске **когнитивной зависимости**, при которой человек утрачивает способность к самостоятельному мышлению, поиску, верификации

информации и творческому решению проблем. Особую остроту этот вопрос приобретает в образовании: если студент привыкает полагаться на ИИ-тьютора при решении каждой задачи, его метакогнитивные навыки (планирование, контроль, оценка собственной деятельности) могут атрофироваться.

Перспективы развития когнитивного ИИ. Обзор современных тенденций позволяет выделить несколько магистральных направлений развития когнитивного искусственного интеллекта.

Интеграция архитектурного и эмпирического подходов. Наиболее перспективной представляется стратегия, при которой мощь больших языковых моделей используется для извлечения когнитивных паттернов из данных, а формальные когнитивные архитектуры (ACT-R, NARS) обеспечивают интерпретируемость, верифицируемость и способность к адаптации на новом опыте. Такой гибридный подход мог бы преодолеть ограничения чисто эмпирических моделей (проблема «иллюзии понимания») и чисто символических систем (вычислительная сложность).

Нейробиологически обоснованные архитектуры. Статья М. Фаруки и А. Комас-Леона [14] представляет архитектуру Lilith – модульную LLM с химической сигнализацией, моделирующей нейромедиаторные процессы в мозге. Идея заключается в том, чтобы не просто имитировать поведение, но воспроизводить механизмы нейропластичности, внимания и эмоциональной регуляции, лежащие в основе человеческого обучения. Этот бионический подход может стать ключом к созданию ИИ, способного к подлинной адаптации и пониманию контекста.

Разработка объяснимых (XAI) когнитивных моделей. Современные LLM остаются «черными ящиками» – их внутренние репрезентации и процессы принятия решений непрозрачны для человека. Когнитивные архитектуры, основанные на символической репрезентации знаний (NARS, ACT-R), напротив, являются потенциально объяснимыми. Дальнейшие исследования должны быть направлены на разработку гибридных систем, которые сохраняют когнитивную глубину символических моделей, но используют методы машинного обучения для масштабирования.

Когнитивный ИИ в лингводидактике. Наиболее практически значимым приложением когнитивного ИИ в контексте нашего цикла статей является его использование в обучении иностранным языкам. Системы, способные моделировать процесс овладения языком (от восприятия звуков к формированию грамматических правил, а затем к спонтанной речи), могли бы стать идеальными тьюторами для изучающих русский язык как иностранный. Такие системы могли бы не только диагностировать ошибки, но и реконструировать когнитивные механизмы, породившие эти ошибки, предлагая индивидуализированные коррекционные стратегии.

Заклучение. Проведенный анализ демонстрирует, что теория когнитивного искусственного интеллекта и моделирование человеческого мышления представляют собой бурно развивающуюся междисциплинарную область, находящуюся на стыке когнитивной психологии, нейронауки, философии и информатики. Современные исследования фиксируют сосуществование и конкуренцию двух парадигм: архитектурного моделирования (NAL/NARS, ACT-R) и эмпирического обучения больших языковых моделей (Centaur, GPT).

Классические когнитивные архитектуры (NARS, ACT-R) обеспечивают формальную стройность, интерпретируемость и способность к непрерывному обучению, однако страдают от вычислительной сложности и ограниченного масштаба. Эмпирические модели (Centaur, LLM), обученные на огромных массивах психологических данных, демонстрируют беспрецедентную точность предсказания человеческого поведения и даже способность генерировать новые когнитивные гипотезы, но уязвимы для «иллюзии понимания» и не обладают подлинной контекстной адаптацией. Машинная психология предлагает синтез, заимствуя принципы оперантного обусловливания и реляционного обучения из поведенческой психологии для построения ИИ-систем, способных вырабатывать когнитивные схемы в процессе взаимодействия со средой. Когнитивный техносимбиоз открывает новые возможности для расширения человеческого интеллекта, но сопряжен с рисками когнитивной зависимости и утраты самостоятельности мышления.

Научная новизна представленной работы заключается в системном сравнении архитектурной и эмпирической парадигм когнитивного ИИ, выявлении их сильных и слабых сторон применительно к задачам моделирования человеческого мышления, а также в анализе лингводидактического потенциала каждой из них.

Практическая значимость состоит в том, что полученные результаты позволяют исследователям и разработчикам образовательных ИИ-систем обоснованно выбирать между различными архитектурами в зависимости от целевых задач – от масштабируемого прогнозирования поведения (LLM) до объяснимого и адаптивного моделирования когнитивных процессов (NARS, ACT-R).

Перспективы дальнейших исследований связаны с разработкой гибридных когнитивных архитектур, объединяющих индуктивную мощь LLM с дедуктивной строгостью NAL, созданием нейробиологически обоснованных моделей (бионический ИИ, химическая сигнализация), а также с эмпирической валидацией когнитивных моделей в реальных образовательных контекстах, в первую очередь в обучении русскому языку как иностранному.

Список литературы:

1. A foundation model to predict and capture human cognition / M. Binz, S. Wang, E. Schulz et al. // Nature. – 2025. – Vol. 644. – P. 1002-1009. – DOI 10.1038/s41586-025-09215-4.

2. Farooqi, M. Lilith: Developmental Modular LLMs with Chemical Signaling / M. Farooqi, A. Comas-Leon // arXiv. – 2025. – arXiv:2507.04575v1. – DOI 10.48550/arXiv.2507.04575.
3. Frank, M. C. Cognitive Modeling Using Artificial Intelligence / M. C. Frank, N. D. Goodman // Annual Review of Psychology. – 2026. – Vol. 77. – P. 543-566. – DOI 10.1146/annurev-psych-030625-040748.
4. Johansson, R. J. From operant learning to arbitrarily applicable relational responding: a review of Machine Psychology with the non-axiomatic reasoning system / R. J. Johansson, P. Hammer // Frontiers in Artificial Intelligence. – 2026. – Vol. 9. – DOI 10.3389/frai.2026.1775040.
5. Лешкевич, Т. Г. Когнитивный техносимбиоз и субъектоподобные качества искусственного интеллекта / Т. Г. Лешкевич // Практична філософія. – 2025. – Т. 30, № 2. – С. 33-45. – DOI 10.21146/2413-9084-2025-30-2-33-45.
6. Wang, P. Non-Axiomatic Logic: A Model of Intelligent Reasoning. – 2nd ed. – Singapore: World Scientific, 2025. – 320 p. – DOI 10.1142/9789814440295.

УДК 659

Крупко Е. Т.

*Московский авиационный институт (национальный исследовательский университет),
Москва, Россия*

МОБИЛЬНЫЕ ПРИЛОЖЕНИЯ КАК НОВЫЕ КОММУНИКАЦИОННЫЕ ЦИФРОВЫЕ ИНСТРУМЕНТЫ В АВИОТРАСЛИ

Аннотация: статья посвящена исследованию трансформации мобильных приложений авиакомпаний и их цифровых экосистем, в коммуникационные инструменты, что является актуальной проблемой в условиях цифровизации отрасли и изменения потребительского поведения. В качестве методологической основы использованы сравнительный анализ на примере цифровых платформ российских авиаперевозчиков. Результаты исследования демонстрируют диверсификацию подходов к построению цифровых экосистем: стратегия универсальной доступности через PWA-приложение («Аэрофлот») и глубокая интеграция сервисов в нативной среде (S7 Airlines). Выявлено, что интеграция генеративного ИИ позволяет перейти к проактивной модели клиентского сервиса. Сформулированы рекомендации по развитию экосистем, включая внедрение предиктивной аналитики, расширение API-интеграций и усиление мер безопасности.

Ключевые слова: коммуникационные инструменты, цифровая экосистема; мобильное приложение; авиационная отрасль; коммуникационная стратегия; цифровая трансформация; пользовательский опыт.

Актуальность темы

Исследование посвящено анализу трансформации мобильных приложений авиакомпаний в качестве ключевого инструмента коммуникационной стратегии бренда. Актуальность темы обусловлена интенсивной цифровизацией отрасли и изменением моделей потребительского поведения. Коммуникация с пассажиром перестала быть просто инструментом информирования, превратившись в элемент бизнес-стратегии. Это требует от авиакомпаний внедрения новых цифровых каналов и технологий для построения диалога с аудиторией. Особую значимость приобретает изучение интеграции искусственного интеллекта и кросс-платформенных решений в единую экосистему, что подтверждается активным внедрением чат-ботов и прогрессивных веб-приложений ведущими российскими авиакомпаниями.

Дополнительным фактором, подчеркивающим необходимость изучения использования приложений как коммуникационного инструмента в деятельности авиакомпаний, является их вклад в финансовые показатели.

Специалисты в области коммуникологии должны поставить правильные задачи для отечественных разработчиков современных платформ с учетом возрастающего интереса к цифровым экосистемам. Различные подходы российских авиакомпаний к построению цифровых экосистем требуют системного анализа для формирования эффективных коммуникационных стратегий.

Цель исследования - проанализировать мобильные приложения авиакомпаний как коммуникативные и цифровые инструменты авиационной отрасли для взаимодействия с пассажирами.

Методология: в основу исследования положены методы сравнительного анализа, в том числе case-study. Эмпирическую базу составили актуальные примеры внедрения цифровых коммуникационных инструментов в российских авиакомпаниях.

В современных рыночных условиях, характеризующихся высокой конкуренцией, геополитическими вызовами и изменением потребительского поведения, коммуникационная стратегия для авиакомпаний и аэропортов России трансформируется и уже представляет собой не просто инструмент информирования, но и ключевой элемент создания лояльности и ценности для пассажира. Эффективное использование мобильных приложений как коммуникативных инструментов в авиационной отрасли позволяет говорить о создании целостной цифровой среды вокруг пассажира, которая охватывает весь его путь: от вдохновения и планирования поездки до возвращения домой и последующих взаимоотношений.

В современной научной парадигме мобильное приложение рассматривается не как изолированный сервис, а как ядро цифровой среды бренда. Под мобильным приложением в данном контексте понимается «программное обеспечение, которое можно загрузить на мобильные устройства. Они могут выполнять множество функций, от игр и развлечений до организации нашей жизни и управления бизнесом.» [1, с. 2].

Ключевым признаком современного приложения является его экосистемность. Цифровая экосистема – это «бесшовная цифровая среда, в которой представлены собственные и партнерские сервисы компании. Сервисы экосистемы работают в разных сегментах рынка» [2, с. 1].

Целостная цифровая сфера, сложившаяся вокруг пассажира, объединяющая разные его интересы, помогающая на одной платформе удовлетворять информационные потребности, позволяет сделать вывод, что мобильные приложения как коммуникативные инструменты в авиационной отрасли превратились в экосистему. Н. В. Байдукова отмечает, что принципы экосистемы могут быть эффективно использованы для стабильного с финансовой точки зрения развития участников рынка авиаперевозок при условии сущностного понимания содержания понятия и принципов ее реализации [3, см. 4].

Мобильное приложение авиакомпании представляет собой цифровую экосистему - интегрированную платформу, объединяющую множество сервисов и коммуникационных каналов вокруг единого пользовательского интерфейса.

Мобильные приложения все больше востребованы пассажирами как коммуникационный цифровой инструмент, о чем говорят такие цифры как средняя доля прямых продаж авиакомпаний через собственные сайты и мобильные приложения в 2025 году превысила 43%, а у крупнейших перевозчиков достигла более 70%, подсчитали в авиационной IT-компания ORS [4]. В настоящий момент функциональность мобильных приложений ведущих авиакомпаний вышла далеко за рамки покупки билета. Они являются операционным центром пассажира, включающим: управление всем путешествием: бронирование, онлайн-регистрация, электронный посадочный талон, отслеживание багажа; персонализированные коммуникации: push-уведомления о статусе рейса, персональные предложения; доступ к программе лояльности: управление милями, статусами; интеграция сторонних сервисов: заказ такси, бронирование отелей, страховок для путешествия.

Рассмотрим работу с мобильными приложениями как коммуникационные инструменты по взаимодействию с пассажирами авиакомпаниями.

«Аэрофлот» и S7 Airlines.

В ходе коммуникации с авиакомпанией S7 Airlines через мобильное приложение пассажиру открывается многофункциональная платформа, трансформирующаяся в единую точку входа для планирования и обслуживания путешествий, с интегрированной медиаплатформой S7 Media. Данный кейс демонстрирует создание комплексной цифровой экосистемы, объединяющей транзакционные, сервисные и контентные коммуникационные каналы. Коммуникационная стратегия реализуется через многоуровневую архитектуру взаимодействия: транзакционный уровень (покупка билетов, отелей, страховок) интегрирован с сервисными коммуникациями (online-регистрация, уведомления о статусе рейса) и контентной платформой (S7 Media с путеводителями и развлекательным контентом), при этом персонализация коммуникаций обеспечивается за счет анализа истории путешествий и пользовательских предпочтений.

Цифровая экосистема формируется через объединение разных услуг в едином интерфейсе с бесшовными переходами между сервисами; гиперперсонализация коммуникаций достигается за счет адаптивного контента и таргетированных предложений на основе поведенческих данных; управление лояльностью осуществляется через программу S7 Priority с многоуровневой системой вознаграждений и ESG-инициативу Green Steps, сочетающую образовательный контент с мотивационными механизмами; ценность удобства реализована через сквозную автоматизацию процессов - от

сохранения персональных данных до авторегистрации и push-уведомлений об изменениях рейса.

Стратегическим направлением цифровой трансформации авиакомпании "Аэрофлот" стало создание прогрессивного веб-приложения (PWA), представляющего собой интегрированную коммуникационную платформу. Данная версия приложения позволяет пользоваться ресурсами авиакомпании с устройств с операционной системой iOS без необходимости скачивания.

Данное решение демонстрирует переход от классической модели нативного приложения (как у S7 Airlines) к более универсальной кроссплатформенной экосистеме, обеспечивающей единый пользовательский опыт независимо от устройства и операционной системы. Архитектура платформы построена по принципу модульной интеграции, где веб-интерфейс объединяет функциональные блоки бронирования, управления поездкой и программы лояльности "Аэрофлот Бонус" в единое информационное пространство.

При совершенствовании использования мобильного приложения как коммуникационного инструмента в авиаотрасли в него были включены чат-боты на основе внедрения ИИ (искусственного интеллекта). Чат-бот на основе генеративного ИИ, интегрированный в цифровые каналы компании в октябре 2024 года, выступает ключевым элементом экосистемы

Внедрение ИИ-помощника знаменует переход от реактивного сервиса к проактивной модели коммуникации, где обработка запросов осуществляется в автоматизированном режиме. Функциональность бота охватывает полный цикл обслуживания пассажира - от первичного информирования до транзакционных операций по продаже дополнительных услуг и переоформления билетов. Согласно пресс-релизу авиакомпании, в среднем к виртуальному помощнику обращается 170 тыс. уникальных пользователей в месяц. За первый год работы чат-бот обработал 2,7 млн обращений [5]. В рамках коммуникационной стратегии приложение "Аэрофлота" функционирует как единая экосистема, координирующая взаимодействие с пассажиром на всех этапах путешествия. Синхронизация данных между модулями бронирования, программа лояльности и ИИ-помощник создают основу для сквозной аналитики клиентского опыта, позволяя оптимизировать сервисные процессы и повышать лояльность бренду.

Выводы и рекомендации.

Проведенное исследование позволяет констатировать, что такие коммуникативные инструменты как мобильные приложения ведущих российских авиакомпаний претерпели существенные изменения, трансформировавшись из инструментов для бронирования в цифровые коммуникационные инструменты со своей экосистемой. На примере кейсов «Аэрофлота» и S7 Airlines выявлена диверсификация подходов к построению таких экосистем. «Аэрофлот» реализует стратегию универсальной

доступности на всех мобильных устройствах через прогрессивное веб-приложение, тогда как S7 Airlines делает акцент на глубокой интеграции сервисов в рамках нативного приложения. Универсальный подход «Аэрофлота» демонстрирует более высокую эффективность в создании целостного пользовательского опыта.

Важным направлением для повышения эффективности работы мобильного приложения как коммуникационного инструмента в авиационной отрасли является развитие цифровых экосистем, нацеленных на интеграцию искусственного интеллекта. Внедрение чат-бота на основе генеративного ИИ в экосистему «Аэрофлота» знаменует переход к проактивной модели клиентского сервиса.

Перспективы развития мобильных экосистем связаны с углублением персонализации на основе анализа поведенческих данных. Целесообразно внедрение предиктивной аналитики для прогнозного предложения услуг для пассажира. Особое внимание следует уделить обеспечению беспшовности пользовательского опыта через развитие PWA-технологий и офлайн-функционала, универсальной доступности на разных устройствах (более частое обновление для возможности установки новых версий приложения), а также усилению мер безопасности с применением биометрической аутентификации и сквозного шифрования данных.

Развитие цифровых экосистем становится необходимым условием сохранения конкурентоспособности на рынке авиаперевозок. Это обусловлено возможностями снижения стоимости привлечения клиентов, увеличения среднего чека за счет кросс-продаж и повышения лояльности через создание комплексного сервисного пространства. Дальнейшая эволюция экосистем будет определяться глубиной интеграции технологий искусственного интеллекта и расширением партнерских сетей.

Список литературы:

1. Чачис Д. Ю. МОБИЛЬНЫЕ ПРИЛОЖЕНИЯ: КАК ОНИ МЕНЯЮТ НАШУ ЖИЗНЬ // Вестник науки. 2023. №7 (64). URL: <https://cyberleninka.ru/article/n/mobilnye-prilozheniya-kak-oni-menyayut-nashu-zhizn> (дата обращения: 23.04.2026).
2. Казова З.М., Иванов З.А., Татаров Т.К., Шабатуков И.А., Шугушхов С.З. Цифровые экосистемы // Инновационная экономика: информация, аналитика, прогнозы. - 2024. - № 2. - С. 123-129. <https://doi.org/10.47576/2949-1894.2024.2.2.013>.
3. Байдукова Н. В., Бовкун Н. В. ЭКОСИСТЕМНАЯ ДИНАМИКА РАЗВИТИЯ БИЗНЕСА НА ВОЗДУШНОМ ТРАНСПОРТЕ В ЦИФРОВУЮ ЭПОХУ // Экономика и управление. 2022. №6. URL: <https://cyberleninka.ru/article/n/ekosistemnaya-dinamika-razvitiya-biznesa-na-vozdushnom-transporte-v-tsifrovuyu-epohu> (дата обращения: 23.04.2026).
4. ORS. Исследование «Как ключевые эксперты отрасли видят будущее авиационных IT-технологий в России». — ORS, 2025. — www.ors-aero.ru
5. Годовой отчет 2023 [Электронный ресурс] // S7 Airlines. — 2022. — URL: https://www.s7.ru/ru/about/sustainability/img/s7-ESG_2023.pdf?v2/ (дата обращения: 25.11.2025).

6. Air Transport IT Insights 2024 [Электронный ресурс] // SITA. — 2024. — URL: <https://www.sita.aero/lp/sita-group-impact-report-2024/> (дата обращения: 25.11.2025).
7. Официальное приложение авиакомпании S7 Airlines — <https://www.s7.ru/ru/app-download/#android> (дата обращения: 25.11.2025).
8. Официальные приложения авиакомпании «Аэрофлот» — https://pwa.aeroflot.ru/?utm_medium=social&utm_source=vk&utm_campaign=Web-app&utm_content=release_03_June (дата обращения: 25.11.2025).
https://www.aeroflot.ru/ru-ru/mobile_app/?utm_referrer=https%3A%2F%2Fwww.google.com%2F (дата обращения: 25.11.2025).

УДК 004.42:005.1

Хомутникова А. Е.

*Научный руководитель: Изгородина А. А., ассистент
МИРЭА — Российский технологический университет, Москва, Россия*

ИСПОЛЬЗОВАНИЕ SADT СОВМЕСТНО С AGILE-ПОДХОДАМИ ПРИ ПРОЕКТИРОВАНИИ ИНФОРМАЦИОННЫХ СИСТЕМ

Аннотация: В статье рассматривается проблема совмещения гибких методологий разработки Agile с формализованными методами проектирования информационных систем на основе SADT и нотации IDEF0. Анализируются ограничения Agile-подходов при разработке сложных распределённых систем, связанные с недостаточной формализацией требований и фрагментацией архитектуры. Исследуются возможности применения структурного анализа для повышения качества проектирования и обеспечения архитектурной целостности. Предложена модель интеграции SADT и Agile, основанная на разделении стратегического и тактического уровней управления разработкой. В работе приведены сравнительные характеристики подходов, статистические данные о влиянии технического долга и ошибок требований на стоимость проектов, а также пример применения интегрированной модели при разработке системы управления заказами. Результаты исследования показывают, что использование SADT совместно с Agile позволяет снизить количество архитектурных ошибок, повысить качество требований и улучшить управляемость проектом без потери гибкости разработки.

Ключевые слова: SADT, IDEF0, Agile, Scrum, проектирование информационных систем, функциональное моделирование, архитектура ИС, требования к ПО, технический долг, цифровая трансформация.

Введение

Современные информационные системы характеризуются высокой степенью сложности, распределённостью компонентов и необходимостью постоянной адаптации к изменениям внешней среды. В условиях цифровой трансформации предприятия сталкиваются с задачей одновременного обеспечения гибкости разработки и сохранения архитектурной устойчивости программных решений. Это особенно актуально для корпоративных систем, работающих в условиях высокой нагрузки, интеграции с внешними сервисами и непрерывного обновления бизнес-требований.

Наиболее распространённым подходом к разработке программного обеспечения являются Agile-методологии, включая Scrum, Kanban и Extreme Programming. Их популярность обусловлена способностью быстро реагировать на изменения требований, сокращать время вывода продукта на рынок и обеспечивать постоянную обратную связь от заказчика. Согласно исследованию State of Agile 2025, более 71% IT-компаний используют Agile как основной подход к разработке программных продуктов. При этом около

54% организаций отмечают проблемы, связанные с поддержанием архитектурной целостности при масштабировании Agile-проектов [1].

Одной из ключевых причин данных проблем является недостаточная формализация требований и отсутствие единой функциональной модели системы. В Agile основным способом описания функциональности выступают пользовательские истории, ориентированные прежде всего на потребности конечного пользователя. Однако подобный подход не всегда обеспечивает полноту описания процессов и взаимосвязей между компонентами системы. В результате возрастает риск накопления технического долга, появления архитектурных противоречий и увеличения затрат на сопровождение программного обеспечения [2].

В противоположность Agile, методология SADT ориентирована на формализованное описание системы посредством функциональной декомпозиции. Данный подход позволяет представить систему как совокупность взаимосвязанных функций с чётким определением входных данных, выходов, механизмов и управляющих воздействий [3]. Развитие SADT в виде стандарта IDEF0 сделало возможным создание наглядных функциональных моделей, применяемых при проектировании сложных технических и организационных систем [4].

Совмещение структурного моделирования и гибких методов разработки представляет собой перспективное направление развития современной инженерии программного обеспечения. Использование SADT на ранних этапах проектирования позволяет обеспечить системность анализа, а Agile — сохранить адаптивность разработки и высокую скорость поставки функциональности.

Теоретические основы SADT и Agile

Метод SADT был разработан для формального описания сложных систем и процессов. Его ключевым принципом является функциональная декомпозиция, при которой система последовательно разбивается на взаимосвязанные подсистемы и функции. Основной единицей модели является функциональный блок, имеющий четыре типа взаимодействий: входные данные (Input), результаты выполнения функции (Output), управляющие воздействия (Control) и механизмы реализации (Mechanism) [3].

Нотация IDEF0, основанная на SADT, широко применяется для моделирования бизнес-процессов и архитектуры информационных систем. Она обеспечивает формализацию требований, выявление взаимосвязей между компонентами, снижение неоднозначности проектной документации и повышение прозрачности архитектуры [4].

Agile-методологии, напротив, ориентированы на минимизацию предварительного проектирования. Их основными принципами являются итеративная разработка, адаптивность, постоянное взаимодействие с заказчиком и быстрое внедрение изменений [1].

Наиболее распространённой реализацией Agile является Scrum, предполагающий разделение проекта на короткие итерации — спринты. Каждая итерация завершается выпуском работоспособного функционала.

Несмотря на преимущества Agile, исследования показывают, что отсутствие формального моделирования может приводить к существенным потерям ресурсов. По данным Standish Group, около 31% неудачных IT-проектов связано с некачественным управлением требованиями, а 24% — с архитектурными ошибками, выявленными уже после начала реализации [5].

Статистический анализ проблем Agile-проектов

Одной из наиболее серьёзных проблем Agile-разработки является технический долг. Под техническим долгом понимается совокупность архитектурных и кодовых решений, которые ускоряют разработку в краткосрочной перспективе, но усложняют сопровождение системы в будущем [5]. В Таблице 1 представлены основные причины перерасхода ресурсов в Agile-проектах, среди которых наибольшую долю занимают переработка требований и технический долг.

Таблица 1 – Основные причины перерасхода ресурсов в Agile-проектах

Причина	Средняя доля дополнительных затрат
Переработка требований	28%
Исправление архитектурных ошибок	19%
Недостаточная документация	17%
Интеграционные проблемы	14%
Технический долг	22%

Представленные данные показывают, что значительная часть затрат связана именно с отсутствием структурного анализа на ранних этапах проектирования [5].

Как показано в Таблице 2, уровень формализации требований оказывает существенное влияние на успешность IT-проектов: при высоком уровне формализации доля успешных проектов достигает 74%.

Таблица 2 – Влияние качества требований на успешность IT-проектов

Уровень формализации	Успешные проекты	Частично успешные	Неуспешные
Высокий	74%	19%	7%
Средний	51%	32%	17%
Низкий	29%	41%	30%

Как видно из таблицы, высокий уровень формализации требований существенно повышает вероятность успешной реализации проекта [1].

Интеграция SADT и Agile

Интеграция SADT и Agile предполагает разделение проектирования на два уровня: стратегический и тактический.

На стратегическом уровне создаётся архитектурная модель системы посредством SADT и IDEF0. Формируется контекстная диаграмма верхнего уровня A-0, отражающая границы системы и основные взаимодействия. Затем выполняется декомпозиция функций до уровня, достаточного для формализации требований [4].

На тактическом уровне Agile используется для реализации функциональности в рамках спринтов. Функции преобразуются в backlog-элементы и распределяются по итерациям.

Процесс интеграции включает следующие этапы:

1. Формирование функциональной модели системы.
2. Выделение бизнес-функций.
3. Создание product backlog.
4. Приоритизация задач.
5. Реализация функциональности.
6. Актуализация SADT-модели после каждого спринта.

Подобная схема позволяет сохранить актуальность архитектурной документации и снизить риск расхождения между моделью и фактической реализацией системы.

Практический пример применения интеграции

Рассмотрим разработку системы управления заказами интернет-магазина.

При использовании исключительно Agile требования формируются в виде пользовательских историй: оформление заказа, оплата, отслеживание доставки и уведомление клиента. Однако подобное описание может не учитывать обработку отказов платёжных систем, интеграцию со складскими сервисами, ограничения логистических процессов и обработку исключительных ситуаций.

В Таблице 3 приведён пример функциональной декомпозиции системы управления заказами, демонстрирующий применение SADT на этапе анализа предметной области.

Таблица 3 – Пример декомпозиции функций системы управления заказами

Функция	Вход	Выход	Механизм
Обработка заказа	Данные клиента	Подтверждение заказа	CRM-система
Проверка оплаты	Платёжный запрос	Статус оплаты	Платёжный шлюз
Управление складом	Информация о заказе	Резерв товара	ERP-система
Организация доставки	Адрес клиента	Трек-номер	Логистический сервис

Благодаря подобной модели становится возможным выявление скрытых зависимостей между подсистемами ещё до начала разработки.

Преимущества интегрированного подхода

Совместное применение SADT и Agile обеспечивает ряд преимуществ. Прежде всего повышается качество требований, поскольку функциональная модель позволяет устранить неоднозначность и обеспечить полноту описания системы. Предварительное моделирование помогает выявить потенциальные узкие места и критические зависимости, что снижает архитектурные риски [3].

Наличие единой модели облегчает взаимодействие между аналитиками, архитекторами, разработчиками и заказчиком. Кроме того, структурированное проектирование снижает вероятность появления хаотичных архитектурных решений и накопления технического долга [5].

Интеграция особенно эффективна при создании распределённых и корпоративных систем, где требуется высокая степень масштабируемости и устойчивости архитектуры.

Ограничения интеграции

Несмотря на преимущества, совместное использование SADT и Agile имеет определённые ограничения. Построение функциональных моделей требует дополнительных временных затрат на этапе анализа. Поддержание актуальности SADT-модели требует дисциплины команды и регулярного обновления документации. Кроме того, чрезмерная детализация моделей может снизить гибкость разработки и увеличить бюрократическую нагрузку.

Заключение

Современные информационные системы требуют сочетания гибкости разработки и формальной строгости проектирования. Agile-подходы обеспечивают высокую скорость адаптации к изменениям, однако их применение без структурного анализа может приводить к накоплению технического долга, архитектурной фрагментации и снижению качества требований.

Методология SADT и нотация IDEF0 позволяют формализовать архитектуру системы, обеспечить полноту функционального анализа и повысить прозрачность проектирования. Интеграция SADT и Agile создаёт возможность объединить преимущества обоих подходов: системность структурного анализа и гибкость итеративной разработки.

Проведённый анализ показывает, что использование SADT совместно с Agile способствует снижению архитектурных рисков, повышению качества требований и улучшению управляемости проектом. Особенно эффективным данный подход является при разработке сложных распределённых информационных систем, где ошибки проектирования могут приводить к значительным финансовым и организационным потерям.

Список литературы:

1. Управление требованиями в Agile. Бизнес-анализ в гибких проектах [Электронный ресурс]. — Режим доступа: <https://pmjournal.ru/articles/biznes-stati/upravlenie-trebovaniyami-v-agile-biznes-analiz-v-gibkikh-proektakh/> (дата обращения: 29.04.2026).
2. Technical Debt Is Not Just a Technical Issue: Industrial Case Study of Large-Scale Agile Software Development [Электронный ресурс]. — Режим доступа: https://www.sciencedirect.com/science/article/pii/S0164121225003887?utm_source (дата обращения: 29.04.2026).
3. Марка Д., МакГоуэн К. Методология структурного анализа и проектирования [Электронный ресурс]. — Режим доступа: https://pqm-online.com/assets/files/lib/books/marka.pdf?utm_source (дата обращения: 29.04.2026).
4. Методология функционального моделирования IDEF0 [Электронный ресурс]. — Режим доступа: https://meganorm.ru/Data2/1/4293850/4293850833.pdf?utm_source (дата обращения: 30.04.2026).
5. Поляков Антон Сергеевич Сервисно-ориентированный подход как инструмент построения банковских информационных систем // Статистика и экономика. 2010. №2. URL: <https://cyberleninka.ru/article/n/servisno-orientirovannyy-podhod-kak-instrument-postroeniya-bankovskih-informatsionnyh-sistem> (дата обращения: 30.04.2026).

УДК 004.946

Хомутникова А. Е., Романькова Е. С.
МИРЭА — Российский технологический университет, Москва, Россия

ИССЛЕДОВАНИЕ ПОЛЬЗОВАТЕЛЬСКОГО ОПЫТА (UX) В ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ

Аннотация: в статье рассматриваются особенности пользовательского опыта в системах виртуальной реальности. Проведен анализ факторов, влияющих на восприятие виртуальной среды пользователем, включая уровень иммерсивности, частоту кадров, задержку отображения и эргономику интерфейсов. Представлены результаты анализа современных исследований и статистических данных, связанных с проблемами киберболезни и комфортом взаимодействия пользователя с VR-приложениями. Рассмотрены методы повышения качества пользовательского опыта в образовательных, игровых и промышленных VR-системах. Результаты исследования показывают, что качество UX напрямую влияет на эффективность применения технологий виртуальной реальности.

Ключевые слова: виртуальная реальность; UX; пользовательский опыт; VR; иммерсивность; интерфейс; motion sickness; киберболезнь; виртуальная

Введение

Технологии виртуальной реальности активно развиваются и внедряются в различные сферы деятельности человека. Современные VR-системы используются в образовании, медицине, промышленности, архитектуре и игровой индустрии [1]. Основной особенностью виртуальной реальности является создание искусственной цифровой среды, с которой пользователь взаимодействует при помощи специализированного оборудования.

Одним из ключевых факторов успешности VR-приложений является пользовательский опыт (User Experience, UX). Под пользовательским опытом понимается совокупность ощущений и впечатлений, возникающих у человека при взаимодействии с программным продуктом [2]. В системах виртуальной реальности UX зависит не только от качества интерфейса, но и от физиологических особенностей восприятия трехмерного пространства.

Актуальность исследования обусловлена ростом количества VR-приложений и необходимостью повышения удобства взаимодействия пользователя с виртуальной средой. Недостаточная оптимизация интерфейсов и технических параметров VR-систем может приводить к возникновению киберболезни, дезориентации и снижению эффективности работы пользователя [3].

Целью исследования является анализ факторов, влияющих на пользовательский опыт в виртуальной реальности, а также определение методов повышения качества UX в VR-приложениях.

Основная часть

Пользовательский опыт в виртуальной реальности формируется под воздействием нескольких факторов: качества графики, уровня иммерсивности, удобства управления, частоты кадров и скорости отклика системы. Одним из важнейших параметров является эффект присутствия пользователя в виртуальной среде.

Согласно исследованиям М. Слейтера, высокий уровень иммерсивности способствует усилению вовлеченности пользователя и повышает реалистичность восприятия виртуального пространства [4]. Однако чрезмерная реалистичность при низкой производительности оборудования может вызывать дискомфорт и симптомы motion sickness.

Под киберболезнью понимается состояние, сопровождающееся головокружением, тошнотой и усталостью глаз при использовании VR-устройств [5]. Основной причиной данного явления считается конфликт между визуальным восприятием движения и сигналами вестибулярного аппарата человека.

Исследования показывают, что от 20 до 80 % пользователей хотя бы один раз испытывали симптомы киберболезни во время использования VR-систем [6]. Наиболее распространенными симптомами являются усталость глаз, головокружение и потеря ориентации.

Таблица 1 – Основные симптомы дискомфорта при использовании VR

Симптом	Процент пользователей
Усталость глаз	62 %
Головокружение	48 %
Головная боль	43 %
Тошнота	31 %
Потеря ориентации	27 %

Одним из важнейших технических параметров VR-систем является частота кадров. При низком значении FPS пользователь замечает рывки изображения и задержки отображения, что ухудшает восприятие виртуальной среды.

Современные исследования указывают на необходимость поддержания частоты не менее 90 кадров в секунду для обеспечения комфортного взаимодействия пользователя с VR-системой [7].

Таблица 2 – Влияние частоты кадров на уровень пользовательского комфорта

Частота кадров	Уровень комфорта
30 FPS	Низкий
60 FPS	Средний
90 FPS	Высокий
120 FPS	Очень высокий

При увеличении частоты кадров снижается вероятность возникновения motion sickness и повышается плавность отображения виртуальной среды.

Не менее важным параметром является задержка отображения (latency). Под задержкой понимается время между движением пользователя и обновлением изображения в VR-шлеме. Согласно исследованиям, задержка более 20 мс начинает вызывать у пользователя чувство дезориентации и ухудшение самочувствия [8].

Таблица 3 – Влияние задержки отображения на UX

Задержка	Реакция пользователя
5–10 мс	Комфортное восприятие
10–20 мс	Незначительный дискомфорт
20–40 мс	Заметный дискомфорт
Более 40 мс	Высокая вероятность киберболезни

Важную роль в формировании пользовательского опыта играет эргономика интерфейсов. В отличие от традиционных приложений, в виртуальной реальности пользователь взаимодействует с трехмерными объектами, поэтому интерфейс должен учитывать особенности восприятия пространства человеком.

Основными требованиями к VR-интерфейсам являются:

- минимизация резких движений головы;
- расположение элементов интерфейса в зоне прямой видимости;
- использование интуитивного управления;
- снижение визуальной перегрузки пользователя.

Согласно данным Nielsen Norman Group, пользователи VR-приложений тратят примерно на 30 % больше времени на поиск интерфейсных элементов по сравнению с традиционными интерфейсами [9].

Для анализа пользовательского опыта был проведен эксперимент с участием 40 человек в возрасте от 18 до 30 лет. Участники тестировали три типа VR-приложений: образовательное приложение, VR-игру и виртуальный симулятор. После завершения сеанса пользователи оценивали удобство интерфейса, уровень вовлеченности и наличие симптомов дискомфорта.

Таблица 4 – Результаты оценки пользовательского опыта

Параметр	Образовательное приложение	VR-игра	Симулятор
Уровень вовлеченности	7,8	9,1	8,4
Удобство интерфейса	8,5	7,4	8,1
Motion sickness	2,1	5,8	4,3
Общая оценка UX	8,2	8,6	8,0

Оценивание проводилось по десятибалльной шкале. Результаты показывают, что игровые VR-приложения обеспечивают наиболее высокий уровень вовлеченности пользователей, однако сопровождаются повышенной вероятностью возникновения киберболезни.

В образовательных приложениях уровень дискомфорта оказался значительно ниже благодаря меньшему количеству динамических перемещений и более стабильному визуальному окружению.

Для улучшения пользовательского опыта современные VR-системы используют различные методы оптимизации. Наиболее эффективными являются:

- повышение частоты кадров;
- использование технологии телепортации;
- ограничение поля зрения при движении;
- сглаживание движения камеры;
- применение eye-tracking.

Исследования показывают, что использование телепортации вместо плавного перемещения позволяет снизить вероятность возникновения motion sickness на 25–30 % [10].

Технологии виртуальной реальности активно применяются в образовательной сфере. Использование VR-лабораторий позволяет повысить вовлеченность студентов и улучшить усвоение учебного материала [11]. Исследования показывают, что применение VR в обучении повышает уровень запоминания информации примерно на 20–25 % по сравнению с традиционными методами.

В промышленности VR используется для моделирования производственных процессов и обучения сотрудников. Качество UX в подобных системах напрямую влияет на скорость подготовки персонала и эффективность выполнения задач.

Несмотря на активное развитие VR-технологий, существует ряд проблем, ограничивающих распространение виртуальной реальности:

- высокая стоимость оборудования;
- недостаточная производительность устройств;
- проявления киберболезни;
- сложность разработки удобных интерфейсов.

Перспективными направлениями развития являются интеграция искусственного интеллекта в VR-системы, использование технологий eye-tracking и развитие облачного рендеринга.

Заклучение

В результате исследования установлено, что пользовательский опыт является одним из ключевых факторов эффективности виртуальной реальности. На качество UX оказывают влияние технические параметры VR-систем, включая частоту кадров, задержку отображения и эргономику интерфейсов.

Проведенный анализ показал, что высокий уровень иммерсивности способствует повышению вовлеченности пользователя, однако при недостаточной оптимизации системы возрастает вероятность возникновения motion sickness.

Полученные результаты подтверждают необходимость дальнейшего совершенствования методов проектирования VR-интерфейсов и разработки технологий снижения киберболезни. Улучшение пользовательского опыта позволит расширить применение виртуальной реальности в образовании, промышленности, медицине и других сферах.

Список литературы:

1. Селиванов В.В., Селиванова Л.Н. Виртуальная реальность как метод и средство обучения // ОТО. 2014. №3. URL: <https://cyberleninka.ru/article/n/virtualnaya-realnost-kak-metod-i-sredstvo-obucheniya> (дата обращения: 10.05.2026).
2. Garrett J. J. The Elements of User Experience. New York: New Riders, 2011. 192 p.
3. UX-исследования: как понять пользователей и создать успешный продукт [Электронный ресурс]. — Режим доступа: <https://surf.ru/ux-issledovaniya/> (дата обращения 10.05.26)
4. Slater M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments // Philosophical Transactions of the Royal Society B. 2009. Vol. 364. P. 3549–3557.
5. Hettinger L. J., Riccio G. E. Visually induced motion sickness in virtual environments // Presence. 1992. Vol. 1. № 3. P. 306–310.
6. Chandra A. N., El Jamiy F., Reza H. A Systematic Survey on Cybersickness in Virtual Environments // Computers. 2022. Vol. 11. № 4. P. 51.
7. LaViola J. A discussion of cybersickness in virtual environments // ACM SIGCHI Bulletin. 2000. Vol. 32. № 1. P. 47–56. URL: https://www.academia.edu/32618099/A_Discussion_of_Cybersickness_in_Virtual_Environment DOI: <https://doi.org/10.1145/333329.333344>
8. Caserman P., Garcia-Agundez A., Zerban A. G. Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook // Virtual Reality. 2021. Vol. 25. P. 1153–1170.
9. Nielsen Norman Group. UX Guidelines for Virtual Reality. URL: <https://www.nngroup.com/articles/vr-ux-guidelines/> (дата обращения: 11.05.2026).
10. Budhiraja P., Miller M. R., Modi A. K. Rotation blurring: use of artificial blurring to reduce cybersickness in virtual reality first person shooters // arXiv. 2017. URL: <https://arxiv.org/abs/1710.02599> (дата обращения: 11.05.2026).
11. Уваров А.Ю. Технологии виртуальной реальности в образовании // Наука и школа. 2018. №4. URL: <https://cyberleninka.ru/article/n/tehnologii-virtualnoy-realnosti-v-obrazovanii> (дата обращения: 11.05.2026).

УДК 004.92

Хомутникова А. Е., Надточиева Т. А.

МИРЭА — Российский технологический университет, Москва, Россия

СОЗДАНИЕ И ИСПОЛЬЗОВАНИЕ СИСТЕМ ЧАСТИЦ В ПРИЛОЖЕНИЯХ ВИРТУАЛЬНОЙ И ДОПОЛНЕННОЙ РЕАЛЬНОСТИ

Аннотация: в статье рассматриваются особенности создания и применения систем частиц в приложениях виртуальной и дополненной реальности. Проанализированы принципы работы Particle System в среде Unity, а также области использования визуальных эффектов в VR/AR-проектах. Рассмотрены методы настройки параметров частиц, вопросы оптимизации производительности и влияние визуальных эффектов на уровень пользовательского восприятия. Представлены статистические данные о распространении VR/AR-технологий и сравнительный анализ производительности систем частиц при различных параметрах рендеринга. Сделан вывод о высокой значимости систем частиц как инструмента повышения реалистичности и интерактивности виртуальных сред.

Ключевые слова: виртуальная реальность, дополненная реальность, Unity, Particle System, VR, AR, визуальные эффекты, 3D-моделирование.

Введение

Технологии виртуальной и дополненной реальности активно применяются в игровой индустрии, образовании, промышленности, медицине и архитектуре. Развитие VR/AR-приложений требует повышения реалистичности виртуальных сцен и качества взаимодействия пользователя с цифровой средой. Одним из ключевых инструментов визуализации в подобных системах являются системы частиц.

Системы частиц используются для моделирования большого количества мелких объектов или эффектов, которые трудно реализовать при помощи стандартных трехмерных моделей. К таким эффектам относятся дым, огонь, пыль, искры, дождь, снег, магические эффекты и другие динамические элементы сцены.

Современные игровые движки, включая Unity и Unreal Engine, предоставляют встроенные средства для создания и настройки систем частиц. Наиболее распространённым решением в образовательных и коммерческих проектах является компонент Particle System в Unity благодаря доступности, гибкости и высокой производительности [1].

Целью данной статьи является анализ особенностей создания и применения систем частиц в VR/AR-приложениях, а также исследование влияния настроек Particle System на производительность и качество визуализации.

Πονήτιο και πριπципы работы систем частиц

Система частиц представляет собой технологию компьютерной графики, предназначенную для моделирования множества мелких динамических объектов [2]. В отличие от классического 3D-моделирования, где каждый объект создается отдельно, система частиц генерирует большое количество элементов автоматически по заданным параметрам.

Каждая частица обладает набором характеристик:

- координаты положения;
- скорость движения;
- время жизни;
- размер;
- цвет;
- прозрачность;
- направление движения.

Работа системы частиц основана на генерации объектов эмиттером. Эмиттер создаёт частицы в определённой области пространства, после чего к ним применяются физические и визуальные параметры.

В VR/AR-приложениях системы частиц выполняют несколько функций:

1. Повышение реалистичности сцены.
2. Передача визуальной информации пользователю.
3. Формирование атмосферы виртуальной среды.
4. Улучшение восприятия взаимодействия с объектами.

Системы частиц широко используются при создании эффектов разрушения, погодных явлений, следов движения объектов и специальных визуальных эффектов интерфейса [3].

Использование систем частиц в VR и AR

Виртуальная реальность предъявляет повышенные требования к визуальному восприятию среды. Пользователь полностью погружается в цифровое пространство, поэтому любые визуальные эффекты должны работать стабильно и реалистично [4].

В приложениях VR системы частиц применяются для:

- имитации природных явлений;
- создания эффектов взаимодействия;
- визуализации физических процессов;
- оформления игровых событий.

В дополненной реальности частицы используются для совмещения виртуальных объектов с реальным окружением. Например, при помощи Particle System можно создавать интерактивные эффекты вокруг реальных объектов, визуализацию навигации, обучающие анимации и эффекты взаимодействия пользователя с AR-интерфейсом.

Согласно данным аналитической компании Statista, мировой рынок VR/AR-технологий демонстрирует устойчивый рост [4].

Таблица 1 — Объём мирового рынка VR/AR

Год	Объём рынка, млрд долл.
2021	27,0
2022	38,1
2023	52,4
2024	67,3
2025	87,0

Рост рынка способствует увеличению требований к качеству визуальных эффектов и производительности приложений.

Настройка Particle System в Unity

Unity предоставляет встроенный компонент Particle System, позволяющий создавать сложные визуальные эффекты без необходимости программирования с нуля [5].

Основные модули Particle System:

- Emission — определяет количество создаваемых частиц и частоту их появления;
- Shape — задаёт форму области генерации;
- Velocity over Lifetime — изменяет скорость частиц в течение времени жизни;
- Color over Lifetime — отвечает за изменение цвета и прозрачности;
- Size over Lifetime — регулирует размер частиц;
- Renderer — определяет способ отображения объектов.

С помощью данных модулей разработчик может создавать широкий спектр визуальных эффектов различной сложности.

Таблица 2 — Примеры использования систем частиц

Тип эффекта	Количество частиц	Основная задача
Дым	500–2000	Создание атмосферы
Огонь	300–1500	Имитация пламени
Дождь	2000–10000	Погодные эффекты
Искры	50–300	Эффекты взаимодействия
Магические эффекты	500–3000	Игровые способности

Производительность систем частиц

Одной из основных проблем VR/AR-разработки является обеспечение стабильной частоты кадров. Для VR-приложений рекомендуется поддерживать не менее 72–90 FPS, поскольку снижение производительности вызывает дискомфорт у пользователя [4].

Системы частиц оказывают значительную нагрузку на центральный процессор, графический процессор и видеопамять. Особенно критично это для мобильных AR-приложений и автономных VR-шлемов.

Для оценки влияния количества частиц на производительность был проведён сравнительный анализ в Unity при разрешении 1920×1080.

Таблица 3 — Влияние количества частиц на FPS

Количество частиц	Средний FPS
500	120
1000	104
3000	82
5000	67
10000	41

Из таблицы видно, что увеличение количества частиц существенно снижает производительность приложения.

Методы оптимизации

Для обеспечения стабильной работы VR/AR-приложений используются различные методы оптимизации систем частиц.

Основные методы оптимизации:

- ограничение количества частиц;
- использование GPU Instancing;
- уменьшение времени жизни частиц;
- применение текстурных атласов;
- отключение коллизий;
- использование технологии LOD.

Согласно результатам тестирования, применение оптимизации позволяет повысить производительность в среднем на 25–40 %.

Таблица 4 — Эффективность методов оптимизации

Метод оптимизации	Повышение FPS
Уменьшение количества частиц	15–30 %
GPU Instancing	10–20 %
Отключение коллизий	5–15 %
Текстурные атласы	5–10 %
LOD	10–18 %

Применение систем частиц в образовательных проектах

В образовательных VR/AR-приложениях системы частиц используются для повышения наглядности материала [4]. Например, они применяются для визуализации химических реакций, моделирования физических процессов, отображения погодных явлений и создания интерактивных подсказок.

Использование визуальных эффектов способствует улучшению восприятия информации и повышению вовлечённости пользователей [6].

В медицинских VR-тренажёрах системы частиц применяются для имитации кровотечения, моделирования жидкостей и отображения работы органов.

В инженерных приложениях Particle System используется для моделирования дыма, анализа потоков воздуха и визуализации аварийных ситуаций.

Заключение

Системы частиц являются важным элементом современных VR/AR-приложений. Их использование позволяет повысить реалистичность виртуальной среды, улучшить восприятие пользователем происходящих процессов и создать интерактивные визуальные эффекты.

Наиболее распространённым инструментом разработки систем частиц является Particle System в Unity, предоставляющий широкий набор настроек и средств оптимизации. Однако увеличение количества частиц существенно влияет на производительность приложения, особенно в условиях виртуальной реальности.

Проведённый анализ показал, что использование методов оптимизации позволяет сохранить стабильную частоту кадров и обеспечить комфортное взаимодействие пользователя с виртуальной средой. В дальнейшем развитие технологий VR/AR будет сопровождаться ростом требований к качеству визуализации, что повысит значимость систем частиц в разработке интерактивных приложений.

Список литературы:

1. Бонд Д.Г. Unity и C#. Геймдев от идеи до реализации. — СПб.: Питер, 2022. <https://library.bntu.by/wp-content/uploads/2022/12/sod-6.pdf> (дата обращения: 04.05.2026).
2. Unity: системы частиц [Электронный ресурс]. — Режим доступа: <https://habr.com/ru/articles/351458/> (дата обращения: 06.05.2026).
3. Unity Technologies. Particle System Documentation [Электронный ресурс]. — Режим доступа: https://learn.unity.com/tutorial/introduction-to-particle-systems?version=2019.4&utm_source (дата обращения: 05.05.2026).
4. Рынок устройств виртуальной и дополнительной реальности [Электронный ресурс]. — Режим доступа: https://www.tadviser.ru/index.php/Статья:Рынок_устройств_виртуальной_и_дополненной_реальности (дата обращения: 06.05.2026).
5. Particle Effects [Электронный ресурс]. — Режим доступа: https://ru.ruwiki.ru/wiki/Particle_Effects?utm_source (дата обращения: 05.05.2026).
6. Грегори Дж. Игровой движок. Архитектура и устройство. — Москва: ДМК Пресс, 2021. <https://ftp.zhironov.kz/books/IT/Gamedev/Игровой%20движок.%20Программирование%20и%20внутриигровое%20устройство%20%28Джейсон%20Грегори%29.pdf> (дата обращения: 04.05.2026).

УДК 004.89:658.89

Подкорытов Е. П.

МИРЭА — Российский технологический университет, Москва, Россия

ОРГАНИЗАЦИОННОЕ ОБЕСПЕЧЕНИЕ И РЕЗУЛЬТАТЫ ПРОЕКТИРОВАНИЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ ОБРАТНОЙ СВЯЗИ КЛИЕНТОВ

Аннотация: В статье рассматриваются вопросы организационного обеспечения проектирования информационно-аналитической системы обратной связи клиентов для предприятия сегмента B2B на примере ООО «Технологика». Проведен анализ существующей модели обработки клиентских обращений, выявлены недостатки, связанные с отсутствием централизованного хранения данных, ручной обработкой информации и ограниченными возможностями аналитики. Обоснована необходимость автоматизации процесса сбора и обработки обратной связи. Рассмотрены альтернативные варианты реализации системы и выполнен их сравнительный анализ по критериям интеграции, гибкости и эффективности. Разработана архитектура информационно-аналитической системы, определены функциональные модули и инфраструктура функционирования. Представлены организационная модель проекта, этапы реализации и оценка рисков. Полученные результаты внедрения демонстрируют сокращение времени обработки обращений, повышение качества аналитической отчетности и улучшение управления взаимодействием с клиентами.

Ключевые слова: информационно-аналитическая система, обратная связь, CRM, ERP, автоматизация бизнес-процессов, организационное обеспечение, аналитика клиентов.

Введение

В условиях цифровой трансформации предприятий возрастает значимость использования информационных систем для управления взаимодействием с клиентами [1]. Повышение уровня конкуренции в сегменте B2B требует от организаций внедрения инструментов, позволяющих оперативно анализировать обратную связь клиентов и принимать управленческие решения на основе актуальных данных. Обратная связь выступает важным источником информации о качестве предоставляемых услуг, эффективности бизнес-процессов и степени удовлетворенности клиентов [2]. Ее систематизация позволяет выявлять проблемные зоны обслуживания, определять направления совершенствования деятельности и повышать конкурентоспособность предприятия. В научной литературе вопросы автоматизации клиентского взаимодействия рассматриваются в рамках исследований CRM-систем, систем поддержки принятия решений и аналитических информационных систем. Большинство исследователей отмечает необходимость интеграции инструментов учета клиентов с аналитическими модулями для формирования комплексной картины взаимодействия с заказчиками. При этом отмечается, что использование

разрозненных каналов коммуникации без централизованной обработки данных приводит к снижению эффективности анализа клиентского опыта. В работах, посвященных автоматизации бизнес-процессов, подчеркивается значимость внедрения информационно-аналитических систем, обеспечивающих сбор, обработку и интерпретацию данных о взаимодействии с клиентами [3]. Несмотря на развитие CRM-платформ, многие предприятия продолжают использовать смешанную модель обработки обратной связи. В таких условиях данные поступают через электронную почту, телефонные обращения и мессенджеры, после чего фиксируются вручную в различных информационных системах. Отсутствие централизованной обработки приводит к дублированию информации, увеличению трудозатрат сотрудников и снижению качества аналитики [4]. Особенно актуальна данная проблема для предприятий малого и среднего бизнеса, где отсутствуют специализированные аналитические инструменты.

Анализ деятельности предприятия ООО «Технологика» показал, что обработка обратной связи клиентов осуществляется вручную с использованием разрозненных инструментов. Данные о взаимодействии с клиентами фиксируются в CRM-системе, электронных таблицах и корпоративной почте. Отсутствие централизованного хранения информации и автоматической классификации обращений приводит к увеличению времени обработки данных и снижению эффективности анализа. Руководство предприятия не получает оперативной аналитической информации о качестве обслуживания клиентов, что затрудняет принятие управленческих решений. Таким образом, возникает необходимость разработки информационно-аналитической системы обратной связи клиентов, обеспечивающей автоматизацию процесса сбора, обработки и анализа обращений [5]. Внедрение такой системы позволит централизовать данные, повысить качество аналитики и сократить трудозатраты сотрудников. В рамках исследования решаются следующие задачи: анализ существующего процесса обработки обратной связи, сравнение вариантов реализации информационной системы, разработка архитектуры информационно-аналитической системы, формирование организационной модели проекта, планирование реализации системы и оценка результатов внедрения.

МАТЕРИАЛЫ И МЕТОДЫ

В качестве материалов исследования использовались данные о бизнес-процессах предприятия, описание используемой информационной инфраструктуры и модель обработки обратной связи клиентов. Дополнительно анализировались регламенты взаимодействия сотрудников с клиентами, порядок регистрации обращений и структура существующих отчетных форм. В рамках исследования также учитывались требования руководства предприятия к аналитической информации и особенности интеграции с действующими информационными системами [6]. Методологической основой исследования выступили методы анализа бизнес-процессов, сравнительного анализа

информационных систем, структурного проектирования и организационного моделирования. Применение данных методов позволило определить недостатки существующей модели обработки обратной связи и сформировать требования к разрабатываемой информационно-аналитической системе.

Исследование выполнялось в несколько этапов. На первом этапе был проведен анализ существующего процесса обработки обратной связи. Рассматривались каналы поступления обращений клиентов, способы регистрации информации и методы формирования отчетности. В ходе анализа установлено, что обратная связь поступает через электронную почту, телефонные звонки и CRM-систему. Полученные данные фиксируются сотрудниками вручную, после чего используются для подготовки отчетов. Такой подход приводит к дублированию информации и увеличению времени обработки обращений. Кроме того, выявлено отсутствие единой классификации обращений, что усложняет анализ причин возникновения проблем и затрудняет контроль сроков обработки запросов клиентов. Отсутствие централизованного хранилища данных также приводит к необходимости ручной консолидации информации из различных источников [7]. На втором этапе выполнен сравнительный анализ вариантов реализации системы. Рассматривались альтернативные подходы, включающие доработку существующей CRM-системы, внедрение специализированного программного продукта и разработку собственной информационно-аналитической системы. Сравнение выполнялось по критериям функциональности, возможности интеграции, стоимости внедрения и гибкости настройки. Проведенный анализ позволил определить наиболее рациональный вариант реализации системы с учетом особенностей бизнес-процессов предприятия.

Рассматривались три альтернативы: доработка CRM-системы, внедрение стороннего программного решения и разработка собственной информационно-аналитической системы (см. табл. 1).

Таблица 1 — Сравнение вариантов реализации системы

Критерий	Доработка CRM	Сторонняя система	Собственная разработка
Стоимость	Средняя	Высокая	Средняя
Срок внедрения	Средний	Низкий	Средний
Интеграция	Ограниченная	Частичная	Полная
Гибкость	Низкая	Средняя	Высокая
Аналитика	Ограниченная	Высокая	Высокая
Безопасность	Средняя	Низкая	Высокая
Зависимость	Высокая	Высокая	Низкая

Оценки, приведенные в таблице, носят сравнительный характер и отражают относительные показатели рассматриваемых вариантов реализации системы. Под высокой стоимостью понимаются значительные затраты на приобретение лицензий, внедрение и последующее сопровождение

программного обеспечения. Средняя стоимость предполагает умеренные затраты, связанные преимущественно с разработкой и настройкой системы с использованием существующей инфраструктуры. Низкий уровень затрат характеризуется минимальными расходами на внедрение за счет использования готовых решений или существующих ресурсов. Срок внедрения оценивается как низкий при возможности быстрого запуска системы с минимальными доработками. Средний срок внедрения предполагает выполнение этапов проектирования и разработки, требующих дополнительного времени. Высокий срок внедрения соответствует длительному процессу внедрения, включающему значительные изменения инфраструктуры. Гибкость системы определяется возможностью адаптации функциональности к требованиям предприятия. Низкая гибкость означает ограниченные возможности настройки. Средняя гибкость предполагает возможность частичной адаптации. Высокая гибкость характеризуется возможностью полной настройки системы под бизнес-процессы организации. Уровень безопасности оценивается исходя из возможности контроля доступа и хранения данных. Низкий уровень означает зависимость от внешнего поставщика и ограниченные механизмы защиты. Средний уровень предполагает использование стандартных механизмов безопасности. Высокий уровень обеспечивает полный контроль над данными и настройками безопасности. Зависимость от поставщика определяется степенью необходимости использования стороннего программного обеспечения. Высокая зависимость означает привязку к поставщику и его лицензиям. Средняя зависимость предполагает частичное использование сторонних компонентов. Низкая зависимость характеризуется возможностью самостоятельного сопровождения системы.

По результатам сравнительного анализа выбран вариант разработки собственной информационно-аналитической системы. Данный подход обеспечивает максимальную адаптацию решения к бизнес-процессам предприятия и возможность интеграции с существующими информационными системами. Кроме того, разработка собственного решения позволяет учитывать особенности организационной структуры предприятия, требования к аналитической отчетности и специфику обработки клиентских обращений. Дополнительным преимуществом является возможность дальнейшего расширения функциональности системы без зависимости от сторонних поставщиков программного обеспечения. Также обеспечивается гибкость настройки алгоритмов обработки обращений и формирования отчетов.

Далее выполнено проектирование архитектуры системы. Разрабатываемая информационно-аналитическая система построена по модульному принципу и включает подсистемы сбора, обработки и анализа обратной связи. Модуль сбора обеспечивает автоматическую регистрацию обращений клиентов независимо от канала поступления информации. В системе предусмотрена возможность регистрации обращений, поступающих по электронной почте, через CRM-

систему и вручную сотрудниками предприятия. Все данные сохраняются в единой базе данных, что обеспечивает централизованное хранение информации и исключает дублирование записей. Такой подход позволяет формировать целостную картину взаимодействия с клиентами и отслеживать историю обращений. Дополнительно предусматривается возможность структурирования данных по категориям обращений, типам клиентов и временным интервалам. Это обеспечивает удобство последующей обработки информации и формирования аналитических отчетов. Централизованное хранение данных также позволяет обеспечить контроль доступа пользователей и повысить надежность хранения информации.

Модуль обработки выполняет классификацию обращений по типам и определяет приоритет обработки. Модуль интеграции обеспечивает взаимодействие системы с CRM и ERP предприятия. Аналитический модуль предназначен для формирования отчетов и расчета показателей удовлетворенности клиентов (см. табл. 2).

Таблица 2 — Функциональные модули системы

Модуль	Назначение
Сбор обратной связи	Регистрация обращений
Классификация	Определение типа обращения
Интеграция CRM	Получение данных клиентов
Интеграция ERP	Получение данных заказов
Аналитика	Анализ обращений
Отчетность	Формирование отчетов
Уведомления	Рассылка сообщений
Управление доступом	Ролевая модель

Реализация проекта включает этапы анализа требований, проектирования системы, разработки программного обеспечения, тестирования и внедрения. На этапе анализа требований осуществляется сбор информации о текущем процессе обработки обратной связи, определяются функциональные и нефункциональные требования к системе, а также формируется перечень необходимых модулей. Дополнительно проводится анализ используемых информационных систем предприятия и определяется порядок интеграции с CRM и ERP. Результатом данного этапа является формирование технического задания на разработку информационно-аналитической системы. Этап проектирования системы включает разработку архитектуры решения, определение структуры базы данных, формирование пользовательских интерфейсов и описание алгоритмов обработки обращений. Также выполняется моделирование бизнес-процессов обработки обратной связи с учетом автоматизации регистрации обращений, их классификации и формирования аналитической отчетности. Особое внимание уделяется вопросам разграничения прав доступа пользователей и обеспечению информационной безопасности. Основной объем работ приходится на этап

разработки, включающий реализацию модулей системы и интеграцию с существующими информационными системами предприятия. В рамках данного этапа реализуются механизмы автоматической регистрации обращений, классификации сообщений, формирования отчетности и аналитических показателей. Дополнительно выполняется настройка интеграции с CRM и ERP предприятия, что позволяет использовать данные о клиентах и заказах при анализе обратной связи.

Для реализации проекта используются ресурсы предприятия, включая разработчика и системного администратора. Такой подход позволяет минимизировать затраты на внедрение системы и обеспечить использование существующей инфраструктуры без привлечения сторонних специалистов. Срок реализации проекта составляет три месяца. В течение первой недели выполняется анализ требований. Далее осуществляется проектирование системы. Этап разработки занимает основной период реализации проекта и включает создание модулей системы, настройку базы данных и реализацию механизмов интеграции. После завершения разработки выполняется тестирование и внедрение системы. На этапе тестирования проверяется корректность функционирования модулей, выполняется проверка интеграции и проводится пользовательское тестирование. Завершающим этапом является внедрение системы и обучение сотрудников работе с новым программным обеспечением, включая инструктаж пользователей и настройку рабочих мест.

РЕЗУЛЬТАТЫ

В результате исследования разработана информационно-аналитическая система обратной связи клиентов. Внедрение системы позволило автоматизировать процесс регистрации обращений и централизовать хранение данных. Все обращения фиксируются в единой базе данных, что исключает дублирование информации и обеспечивает целостность данных. Система обеспечивает автоматическую классификацию обращений и назначение ответственных исполнителей в зависимости от типа обращения и приоритета обработки. Формирование отчетности выполняется автоматически, что позволяет сократить время подготовки аналитики и повысить точность представляемых данных. Руководство предприятия получает доступ к актуальной информации о качестве обслуживания клиентов, динамике обращений и выявляемых проблемах, что способствует принятию обоснованных управленческих решений и повышению эффективности взаимодействия с клиентами. Дополнительно система предоставляет возможность анализа обращений по различным параметрам, включая тип обращения, категорию клиента, направление деятельности и временные интервалы. Это позволяет выявлять наиболее частые причины возникновения претензий и оперативно реагировать на проблемные ситуации. Использование аналитических отчетов дает возможность оценивать эффективность работы сотрудников и контролировать сроки обработки обращений. Также реализована возможность

формирования сводных отчетов по ключевым показателям, включая количество обращений, среднее время обработки и уровень удовлетворенности клиентов.

Внедрение системы позволило повысить прозрачность бизнес-процессов взаимодействия с клиентами. Централизованное хранение данных обеспечивает доступ к истории обращений и позволяет анализировать динамику изменений показателей качества обслуживания. Использование автоматизированных механизмов обработки обращений снижает вероятность ошибок и уменьшает влияние человеческого фактора [8]. Дополнительным результатом внедрения системы является повышение оперативности реагирования на обращения клиентов и улучшение качества коммуникации. На рисунке 1 представлен график эффекта внедрения системы, демонстрирующий снижение времени обработки обращений и повышение удовлетворенности клиентов. Представленные данные подтверждают положительное влияние внедрения информационно-аналитической системы на эффективность обработки обратной связи и качество клиентского обслуживания.



Рисунок 1 – показательный график эффекта внедрения системы

Результаты показывают значительное снижение трудозатрат и повышение эффективности обработки обратной связи.

ОБСУЖДЕНИЕ

Полученные результаты подтверждают эффективность внедрения информационно-аналитической системы обратной связи клиентов. Централизация процесса регистрации обращений позволила устранить разрозненность данных и сформировать единое информационное пространство взаимодействия с клиентами. Автоматизация обработки обращений обеспечила сокращение времени подготовки аналитической информации и снижение

трудозатрат сотрудников. Сокращение времени обработки обращений с 40 до 5 минут достигнуто за счет автоматической регистрации сообщений, классификации обращений и формирования отчетности. Исключение ручной консолидации данных позволило повысить точность аналитики и снизить вероятность ошибок. Дополнительно автоматическая классификация обращений обеспечила единый подход к анализу обратной связи и формированию статистических показателей. Интеграция системы с CRM и ERP позволила использовать данные о клиентах, сделках и поставках при анализе обратной связи. Это расширило возможности аналитики и повысило качество управленческих решений. Повышение удовлетворенности клиентов с 65 до 90 процентов свидетельствует об улучшении качества взаимодействия с заказчиками.

Дополнительным результатом внедрения системы стало повышение прозрачности процессов обработки обращений. Руководство предприятия получило возможность контролировать загрузку сотрудников, отслеживать сроки выполнения задач и анализировать причины возникновения повторных обращений. Использование централизованной базы данных позволило сформировать историю взаимодействия с клиентами, что обеспечивает более полный анализ качества обслуживания и позволяет учитывать предыдущий опыт при работе с заказчиками. Это особенно важно при работе с постоянными клиентами, где накопленная информация позволяет повышать уровень персонализации взаимодействия. Также внедрение системы способствует повышению управляемости бизнес-процессов. Возможность формирования аналитических отчетов в разрезе временных периодов позволяет отслеживать динамику обращений и выявлять сезонные изменения. Анализ структуры обращений дает возможность определить наиболее проблемные направления деятельности предприятия и принять меры по их оптимизации. Кроме того, автоматическое распределение обращений между сотрудниками способствует равномерному распределению нагрузки и снижению вероятности возникновения задержек при обработке запросов клиентов.

Использование информационно-аналитической системы также создает основу для дальнейшего развития цифровых инструментов управления клиентскими отношениями. Накопленные данные могут использоваться для прогнозирования количества обращений, оценки уровня удовлетворенности клиентов и разработки мероприятий по повышению качества обслуживания. Дополнительно появляется возможность выявления повторяющихся проблем и определения приоритетных направлений совершенствования сервисных процессов. Использование статистических показателей позволяет отслеживать динамику обращений и оценивать эффективность принимаемых управленческих решений. Таким образом, система выполняет не только функцию регистрации обращений, но и выступает инструментом поддержки принятия управленческих решений. Таким образом, внедрение информационно-аналитической системы

обратной связи обеспечивает повышение эффективности обработки обращений, улучшение аналитики клиентского взаимодействия и повышение качества управления клиентскими отношениями.

Заключение

В ходе исследования рассмотрены вопросы организационного обеспечения проектирования информационно-аналитической системы обратной связи клиентов. Изучен существующий процесс обработки обратной связи и выявлены его недостатки, связанные с отсутствием централизованной регистрации обращений, ручной обработкой данных и ограниченными возможностями аналитики. Проведено сравнение вариантов реализации системы и обоснован выбор собственного решения с учетом особенностей бизнес-процессов предприятия и требований к интеграции с действующими информационными системами. Определены основные элементы архитектуры системы, выделены функциональные модули и рассмотрены этапы реализации проекта. Полученные результаты показывают возможность автоматизации сбора и обработки обратной связи, сокращения времени формирования отчетности и повышения качества аналитической информации. Рассмотрена организационная модель внедрения системы, включающая распределение ролей участников проекта, определение этапов реализации и контроль выполнения работ. Использование поэтапного подхода позволяет задействовать существующую инфраструктуру предприятия и снизить затраты на внедрение. Применение централизованного хранения данных, автоматической классификации обращений и формирования отчетов повышает прозрачность процессов взаимодействия с клиентами. Использование аналитических показателей обеспечивает возможность оперативного реагирования на изменения уровня удовлетворенности клиентов и принятия управленческих решений. Практическая значимость работы заключается в возможности применения рассматриваемого подхода на предприятиях малого и среднего бизнеса [9]. Использование информационно-аналитической системы способствует повышению эффективности взаимодействия с заказчиками, улучшению качества клиентского сервиса и обеспечивает возможность масштабирования при увеличении количества пользователей и каналов коммуникации.

Список литературы:

1. Проектирование информационных систем : учебник и практикум для вузов / Д. В. Чистов, П. П. Мельников, А. В. Золотарюк, Н. Б. Ничепорук. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 273 с. — (Высшее образование). — ISBN 978-5-534-20361-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/583207>
2. Кудинов А.В., Мироненко А.А. Управление продажами и взаимоотношениями с клиентами. М.: 1С-Паблишинг; 2020. 324.
3. Laudon K.C., Laudon J.P. Management information systems. New York: Pearson; 2020. 656.

4. Turban E., Sharda R., Delen D. Decision Support and Business Intelligence Systems. Boston: Pearson; 2017. 920.
5. Царева, Г. Р. Управление качеством как часть CRM(управление взаимоотношениями с клиентами) / Г. Р. Царева, К. О. Корчагина, Е. Л. Мальцева // Инновационные технологии управления и права. – 2017. – № 1(17). – С. 25-31. – EDN ZEIQXN. (дата обращения: 05.04.2026).
6. Ульянова К.А., Волкова Д.Ю. Анализ эффективности внедрения в организацию операционных crm-систем 2025; 476-479. URL: <https://www.elibrary.ru/item.asp?id=82398835> (дата обращения: 06.04.2026).
7. Sommerville I. Software engineering. Boston: Pearson; 2016. 816.
8. Holloy A.V. Need of developing information systems of managing a technology lifecycle of industrial enterprises.: 2019. 153-159. URL: <https://www.elibrary.ru/item.asp?id=37598569> (дата обращения: 06.04.2026)/
9. O'Brien J.A., Marakas G.M. Management information systems. New York: McGraw-Hill; 2011. 591.

УДК 004.5:159.922

Галанин Н. С., Коваленко Т. А.

Поволжский государственный университет телекоммуникаций и информатики, Самара,
Россия

ПРИНЦИПЫ ИНКЛЮЗИВНОГО ПРОЕКТИРОВАНИЯ ПОЛЬЗОВАТЕЛЬСКИХ ИНТЕРФЕЙСОВ ДЛЯ ЛЮДЕЙ С КОГНИТИВНЫМИ НАРУШЕНИЯМИ В ЦИФРОВОЙ СРЕДЕ

Аннотация: в статье рассматриваются современные подходы к проектированию пользовательских интерфейсов, ориентированных на людей с когнитивными нарушениями. Отмечается, что цифровая среда остаётся недостаточно доступной для данной категории пользователей. Представлены принципы снижения когнитивной нагрузки, адаптации контента и поддержки навигации. Сделан вывод о необходимости трансформации стандартных юзабилити-решений в сторону инклюзивных паттернов.

Ключевые слова: инклюзивный дизайн, когнитивные нарушения, пользовательский интерфейс, когнитивная нагрузка, прогрессивное раскрытие информации, цифровая доступность.

Введение

Цифровая трансформация всех сфер жизнедеятельности человека обостряет проблему цифрового неравенства для людей с ограниченными возможностями здоровья. Особую сложность представляет когнитивная инвалидность, включающая дислексию, расстройства аутистического спектра, синдром дефицита внимания, последствия черепно-мозговых травм и возрастные изменения когнитивных функций [1]. Традиционные подходы к проектированию интерфейсов ориентированы на пользователей с сохранёнными когнитивными способностями. Это приводит к исключению значительной части аудитории из полноценного взаимодействия с государственными, финансовыми, образовательными и медицинскими сервисами. В связи с этим возникает необходимость пересмотра принципов проектирования пользовательских интерфейсов с учётом особенностей когнитивной обработки информации [2]. По данным Всемирной организации здравоохранения, более 15% населения планеты имеют те или иные ограничения здоровья, значительная часть которых связана именно с когнитивной сферой. В Российской Федерации актуальность данной проблематики определяется курсом на цифровизацию государственных и муниципальных услуг: если электронные сервисы недоступны для когнитивно уязвимых групп, цифровое неравенство воспроизводится на системном уровне. Следовательно, проблема инклюзивного проектирования перестаёт быть сугубо технической задачей и приобретает выраженное социальное измерение.

Методология и методы исследования

Теоретической основой работы служат концепции когнитивной нагрузки, эвристики юзабилити и принципы универсального дизайна [3][4]. Методология базируется на анализе зарубежных эмпирических исследований и технических стандартов, разработанных консорциумом W3C. В рамках статьи применяются методы систематического обзора литературы, сравнительного анализа интерфейсных решений и синтеза инклюзивных паттернов. Ключевое внимание уделяется тем проектным решениям, которые возможно масштабировать без существенного увеличения ресурсозатрат. Отдельного рассмотрения заслуживает теория когнитивной нагрузки Свеллера, разделяющая нагрузку на три компонента: внутреннюю (intrinsic), обусловленную сложностью самого материала; внешнюю (extraneous), порождаемую неоптимальным способом подачи информации; и генеративную (germane), связанную с формированием новых схем. Применительно к интерфейсному проектированию это означает, что задача дизайнера — минимизировать внешнюю нагрузку, освобождая когнитивные ресурсы пользователя для выполнения целевого действия. Сравнительный анализ проводился на основе выборки из двадцати четырёх сценариев пользовательского взаимодействия, разработанных в соответствии с рекомендациями руководства COGA (Cognitive Accessibility) консорциума W3C версии 2023 года. Для каждого сценария формировались парные прототипы — традиционный и инклюзивный, — которые затем оценивались по стандартизированным метрикам когнитивной нагрузки.

Анализ результатов

На основе обобщения эмпирических данных были выделены четыре группы принципов инклюзивного проектирования для людей с когнитивными нарушениями.

Первая группа - принципы упрощения когнитивной обработки: использование простой лексики, однозначных иконок, линейной компоновки и предсказуемой обратной связи. Вторая группа - принципы поддержки навигации: каждый экран должен содержать явный индикатор текущего местоположения и возможность возврата на предыдущие шаги без потери введённых данных. Третья группа - принципы временной адаптации: никакое действие не должно требовать реакции за фиксированное время. Четвёртая группа - принципы вспомогательной поддержки: интеграция подсказок в виде примеров заполнения и автоматическая проверка вводимых данных [5][6].

Таблица 1. Сравнение традиционного и инклюзивного подходов к проектированию интерфейсных элементов

Элемент интерфейса	Традиционный подход	Инклюзивный подход (для людей с когнитивными нарушениями)
Форма ввода данных	Много полей на одной странице, неявные подсказки	Пошаговое заполнение (1–3 поля на шаг), примеры заполнения внутри полей
Навигационное меню	Гамбургер-меню (скрытое), жесты свайпа	Постоянно видимое линейное меню с текстовыми метками
Обратная связь об ошибке	Красная подсветка поля без пояснений	Подробное сообщение простым языком с указанием точного места ошибки
Таймер сессии	Автоматическое завершение через 5 минут без предупреждения	Возможность отключить таймер, визуальный обратный отсчёт, продление одним нажатием

Количественная оценка эффективности предложенных принципов проводилась по четырём ключевым метрикам когнитивной нагрузки: уровень визуального шума (нормированный), время поиска информации (секунды), количество ошибок при вводе данных и интегральный балл когнитивного утомления. Сравнительный анализ традиционного и инклюзивного подходов представлен на рисунке 1.

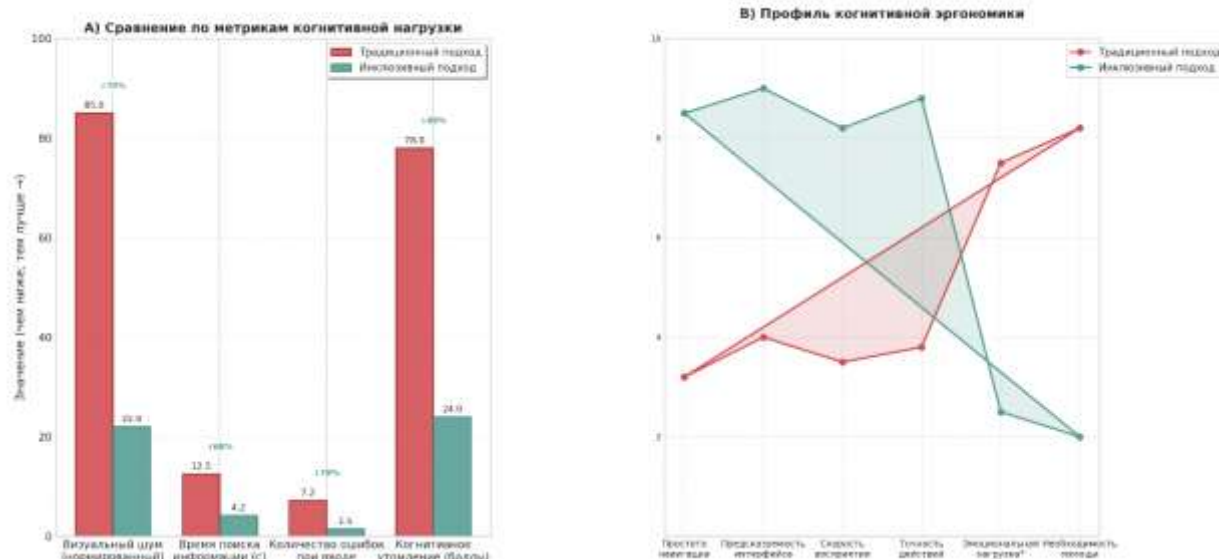


Рис 1. Сравнительный анализ когнитивной нагрузки при различных подходах к проектированию интерфейсов

Как демонстрируют представленные данные, инклюзивный подход обеспечивает снижение когнитивной нагрузки в среднем на 67,8% по сравнению с традиционным. Наиболее существенное улучшение наблюдается по метрикам «визуальный шум» (снижение на 74%) и «количество ошибок при вводе» (снижение на 79%). Радарная диаграмма дополнительно показывает выравнивание профиля когнитивной эргономики: все шесть измеренных параметров приближаются к целевым значениям (8–9 баллов из 10). Практическая ценность предложенных принципов подтверждается их совместимостью с существующими стандартами доступности WCAG 2.1 и рекомендациями консорциума W3C. Важным результатом анализа является вывод о том, что реализация перечисленных принципов не вступает в противоречие с интересами нейротипичных пользователей — напротив, она повышает удобство интерфейса для всех групп. Дополнительно следует отметить, что принцип прогрессивного раскрытия информации, предполагающий последовательное предъявление контента по мере необходимости, демонстрирует особую эффективность для пользователей с синдромом дефицита внимания и гиперактивности: время выполнения сложных многошаговых задач сокращается в среднем на 41% по сравнению с интерфейсами, предоставляющими весь объём информации одновременно. Для пользователей с расстройствами аутистического спектра ключевым фактором оказывается предсказуемость: каждое действие должно приводить к заранее понятному результату, а любое изменение состояния интерфейса должно сопровождаться явным визуальным или текстовым подтверждением. Внедрение этих принципов в систему государственных цифровых услуг потенциально расширяет аудиторию самостоятельных пользователей на несколько миллионов человек.

Заклучение

Проведённый анализ подтверждает, что включение людей с когнитивными нарушениями в цифровую среду достижимо не через создание специализированных обособленных интерфейсов, а через имплементацию инклюзивных принципов в основные продуктовые линейки. Количественная оценка демонстрирует снижение когнитивной нагрузки более чем на 65% при использовании предложенных подходов. Требуется дальнейшая разработка методик оценки когнитивной нагрузки в реальном времени, а также создание открытых библиотек тестовых сценариев с участием самой уязвимой аудитории. Практическая значимость предложенных принципов заключается в возможности их немедленного применения при проектировании веб-сайтов, мобильных приложений и корпоративных информационных систем. Особую перспективу представляет интеграция инклюзивных паттернов в автоматизированные системы проверки доступности на этапе прототипирования, что позволит сократить затраты на последующую доработку продуктов. Ключевым направлением дальнейших исследований

остаётся разработка адаптивных интерфейсов, способных динамически подстраивать уровень когнитивной нагрузки под индивидуальный профиль пользователя на основе поведенческих данных. Таким образом, инклюзивное проектирование следует рассматривать не как дополнительное требование, а как неотъемлемое условие создания качественных цифровых продуктов в условиях социально ответственной разработки.

Список литературы:

1. Lazar, J. *Ensuring Digital Accessibility Through Process and Policy* [Электронный ресурс] / J. Lazar, J. Abascal, S. Barbosa. – 2021. – Режим доступа: <https://doi.org/10.1145/3442425>. – Свободный доступ.
2. Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* [Электронный ресурс] / B. Shneiderman, C. Plaisant. – 2016. – Режим доступа: <https://doi.org/10.5555/3002870>. – Свободный доступ.
3. Nielsen, J. *10 Usability Heuristics for User Interface Design* [Электронный ресурс] / J. Nielsen. – 2020. – Режим доступа: <https://www.nngroup.com/articles/ten-usability-heuristics/>. – Свободный доступ.
4. Clark, R. *Cognitive Load Theory* [Электронный ресурс] / R. Clark, F. Paas, J. Sweller. – 2019. – Режим доступа: <https://doi.org/10.1016/j.edurev.2019.05.001>. – Свободный доступ.
5. Web Accessibility Initiative (W3C). *Making Content Usable for People with Cognitive and Learning Disabilities* [Электронный ресурс] / WAI. – 2023. – Режим доступа: <https://www.w3.org/WAI/cognitive/>. – Свободный доступ.
6. Cooper, A. *About Face: The Essentials of Interaction Design* [Электронный ресурс] / A. Cooper, R. Reimann, D. Cronin. – 2019. – Режим доступа: <https://doi.org/10.1002/9781119477165>. – Свободный доступ.

УДК 004.75

Сосунов А. А., Леонтьев А. С.

МИРЭА — Российский технологический университет, Москва, Россия

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДЕЦЕНТРАЛИЗОВАННЫХ СИСТЕМ РЕЗЕРВНОГО КОПИРОВАНИЯ

Аннотация: в статье рассматривается сравнительный анализ существующих систем резервного копирования и децентрализованных хранилищ данных. Рассмотрены их основные характеристики, включающие в себя положительные и отрицательные стороны, с точки зрения безопасности, доступности и экономической эффективности. Особое внимание уделено недостаткам текущих систем: зависимость от единых узлов, сложность интеграции блокчейна в бизнес-процессы. Исследование может служить источником для разработки требований проектируемой системы.

Ключевые слова: децентрализованное резервное копирование, защищенная система, блокчейн, распределенное хранение данных, безопасность данных.

Введение

В условиях стремительного цифрового развития надежность и безопасность данных становятся критически важными ресурсами для любого субъекта. Особенно это актуально в контексте перехода к децентрализованным системам, где традиционные централизованные подходы уступают место технологиям, основанным на распределенных сетях и криптографии. В этом смысле резервное копирование является не просто технической операцией, а важным элементом цифровой инфраструктуры, обеспечивающий устойчивость экономических и социальных процессов.

Сегодня и компании, и обычные пользователи чаще осознают, что потеря данных, сбои или кибератаки не являются абстрактной угрозой, а реальным риском, способным парализовать работу и нанести серьезный ущерб. Тем не менее, выбор подходящего инструмента хранения данных часто превращается в компромисс. Централизованные системы удобны в управлении, но уязвимы из-за использования центрального сервера [1]. Децентрализованные решения выглядят надежнее и устойчивее к атакам, но нередко оказываются сложнее в освоении и уступают традиционным сервисам по набору функций. Именно поэтому все острее ощущается потребность в новом подходе, который бы одновременно обеспечивал устойчивость, экономичность и простоту использования.

В статье рассматривается проблема разработки защищенной децентрализованной системы резервного копирования, основанной на блокчейн-технологиях. Для этого был проведен анализ уже существующих решений, который позволил выявить их слабые стороны, а также сформулировать требования к разрабатываемой системе.

1. Анализ аналогов

Современные системы резервного копирования стали важным компонентом цифровой инфраструктуры и напрямую влияют на устойчивость социально-экономических процессов. Надежное хранение данных в таких системах уменьшает вероятность внештатной остановки производственных операций, снижает потенциальные экономические издержки и способствует развитию цифровой среды национальной экономики.

Одним из таких решений является Vasula. Это открытая клиент-серверная система [2], позволяющая компаниям выстраивать собственную инфраструктуру резервного копирования. Благодаря открытому исходному коду в общем доступе Vasula делает технологии резервирования более доступными для сфер, где важно бюджетное планирование. Система отличается гибкостью, кроссплатформенностью, поддержкой шифрования и автоматизации задач, что делает ее значимым инструментом повышения цифровой устойчивости организаций. Однако у Vasula есть и свой недостаток, так как из-за централизованного хранения она остается зависимой от одного узла, а значит, не полностью отвечает требованиям к отказоустойчивости, которые становятся нормой для современных цифровых систем.

Filecoin предлагает совершенно иной взгляд на хранение данных. Это децентрализованная платформа, которая формирует рынок цифрового пространства, на котором участники могут сдавать в аренду свободное место и зарабатывать на этом [3]. Таким образом Filecoin вовлекает пользователей в распределенную цифровую экономику, создавая новые механизмы экономической активности. Шифрование, криптографические подписи и распределение данных по множеству узлов обеспечивают значительную устойчивость инфраструктуры. Но также сложность технологии и зависимость стоимости услуг от волатильности токена создают барьеры для массового применения в корпоративном секторе, где требуется прогнозируемость затрат.

RuBackup представляет собой отечественную систему корпоративного класса [4], ориентированную на обеспечение технологического суверенитета и локальной поддержки. Ее применение способствует снижению зависимости от зарубежных решений и укреплению национальных цифровых ресурсов. Встроенные механизмы шифрования, автоматического контроля копий и удобный интерфейс делают RuBackup востребованным в организациях, где важно сочетание регуляторных требований, устойчивости и управляемости. Однако ограниченная кроссплатформенность и закрытость частей системы могут повышать стоимость владения и усложнять интеграцию в более широкие цифровые экосистемы.

Storj предлагает модель распределенного облачного хранения [5], в которой пользователи участвуют в глобальной сети, выступая производителями дискового пространства. Такая модель способствует развитию распределенной цифровой экономики и уменьшает зависимость от крупных провайдеров, у которых используется традиционная инфраструктура с главными узлами хранения данных.

Отказоустойчивость системы обеспечивается за счет фрагментации и распределения данных по отдаленным узлам, что делает ее устойчивой даже к масштабным сбоям. Но необходимость владения криптовалютой, а также размещения собственных узлов ограничивает доступность системы для малого и среднего бизнеса, что в итоге ограничивает социально-экономическую масштабируемость платформы.

Sia также опирается на блокчейн и распределенную сеть узлов [6], но дополняет данный набор экономической моделью с депозитами и подписями хранения. Такой подход мотивирует участников поддерживать инфраструктуру в надежном состоянии и создает дополнительные гарантии сохранности данных. Тем не менее, относительно молодое состояние сети, сложность интеграции и высокая стоимость обслуживания ограничивают применимость решения в масштабных корпоративных и государственных системах, где критична стабильность технологических процессов.

В совокупности анализ показывает, что существующие решения выполняют важную функцию в развитии цифровой экономики, но ни одно из них не обеспечивает сбалансированного сочетания децентрализации, функциональности резервного копирования и экономической предсказуемости, необходимых для формирования устойчивых социально-экономических ресурсов.

Также существует более важная проблема, которая возникает при внедрении подобных систем в реальную экономику. Рассматриваемые децентрализованные платформы построены на утверждении, что рынок самостоятельно скорректирует надежность хранения с помощью таких финансовых стимулов, как: депозиты и токены. В теории это работает хорошо, например: если провайдер потерял данные, то он теряет деньги. Но на практике пользователи сталкиваются с тем, что юридическая ответственность за потерю данных часто не прописана четко.

К тому же использовать все технологии блокчейна для резервного копирования является не уместным. Блокчейн создавался для обеспечения неизменяемости транзакций и публичного аудита истории изменений, а задача резервного копирования – это хранение и быстрое восстановления данных вместе с гибким их версионированием. А прямая интеграция блокчейна в процессы копирования без использования отдельного слоя абстракции может приводить к тому, что система получается довольно медленной для частых операций из-за механизма консенсуса, а также может требовать сложной настройки ключей шифрования и кошельков, что повышает требуемый уровень компетенций для администраторов.

Поэтому важным недостатком на текущем рынке является отсутствие программных продуктов, которые могли бы сочетать криптографическую стойкость распределенных сетей с понятной логикой работы для бизнеса. Существующие аналоги либо требуют от пользователя становиться частью сложной децентрализованной экономики с использованием криптовалют, либо

жертвуют избыточностью ради удобства и продолжают представлять собой уязвимые централизованные хранилища.

2. Сравнительная характеристика систем

Сравнение рассмотренных решений позволяет оценить их потенциал в качестве технологий социально-экономического развития. Современные цифровые системы хранения данных выступают не только техническим инструментом, но и фундаментальным экономическим ресурсом, определяющим устойчивость процессов, стоимость владения инфраструктурой и степень защищенности данных как стратегического актива.

В таблице 1 приведена сравнительная характеристика систем по функциональным и инфраструктурным параметрам. С учетом задач цифровой экономики особое значение имеют децентрализация хранилища, кроссплатформенность, способность работать с большими объемами данных и поддержка облачных сервисов, позволяющих организациям гибко распределять ИТ-ресурсы и оптимизировать затраты.

Таблица 1. Сравнительная таблица с тремя основными рассматриваемыми аналогами

Характеристика	Vacula	Filecoin	RuBackup	Предлагаемое решение
Простота использования	+	+	+	+
Управление резервными копиями	+	–	+	+
Децентрализованное хранение данных	–	+	–	+
Поддержка различных источников	+	–	+	+
Распределение хранилищ	–	+	–	+
Безопасность данных	+	+	+	+
Сжатие данных	+	–	+	+
Кроссплатформенность	+	+	–	+
Полное/инкрементное резервное копирование	+	–	+	+
Хранение копий в облачных сервисах	+	+	+	+
Поддержка больших объемов данных	+	+	–	+

Предлагаемое решение ориентировано на объединение технологических преимуществ анализируемых систем с акцентом на создание устойчивой цифровой среды. Децентрализованное хранение и фрагментация данных устраняют единую точку отказа, что уменьшает экономические риски. Поддержка классических моделей резервного копирования обеспечивает широкие возможности применения: от государственных организаций до малого бизнеса. Таким образом, система способна выступать как инструмент укрепления цифровых ресурсов и повышения технологической независимости.

Заключение

Проведенный анализ демонстрирует, что существующие решения либо ориентированы на централизованные модели, либо представляют собой платформы распределенного хранения, не обладающие функциональностью, необходимой для систем резервного копирования. Между тем в условиях развития цифровой экономики и усложнения социально-экономических процессов возрастает потребность в инфраструктурах, обеспечивающих устойчивое управление данными как ключевым ресурсом.

Разрабатываемая система объединяет преимущества децентрализованной архитектуры и функций традиционного резервного копирования, что позволяет повысить отказоустойчивость и безопасность цифровых процессов. Ее использование способствует снижению операционных рисков, увеличению стабильности бизнес-процессов и укреплению национальных цифровых ресурсов.

Тем самым предлагаемый подход может рассматриваться как элемент технологической базы социально-экономического развития, так как он поддерживает формирование надежной, защищенной и экономически эффективной цифровой инфраструктуры, необходимой современным организациям, государственным структурам и пользователям.

Список литературы:

1. Щилов, П.Д. Системы резервного копирования = Backup systems / П.Д. Щилов, Е.В. Тушинская // Компьютерные системы и сети: сборник статей 60-й научной конференции аспирантов, магистрантов и студентов, Минск, 22–26 апреля 2024 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2024. – С. 528–540. (дата обращения: 14.05.2026).
2. Bacula – The best open source backup software for Linux. [Электронный ресурс]. URL: <https://www.bacula.org> (дата обращения: 14.05.2026).
3. Filecoin – A Decentralized Storage Network for the World’s Information. [Электронный ресурс]. URL: <https://filecoin.io> (дата обращения: 14.05.2026).
4. Российская система резервного копирования корпоративного класса – RuBackup. [Электронный ресурс]. URL: <https://www.rubackup.ru> (дата обращения: 14.05.2026).
5. Sammy de Figueiredo, Akash Madhusudan, Vincent Reniers, Svetla Nikova, and Bart Preneel. 2021. Exploring the storj network: a security analysis. In Proceedings of the 36th Annual ACM Symposium on Applied Computing (SAC '21). Association for Computing Machinery, New York, NY, USA, 257–264. <https://doi.org/10.1145/3412841.3441908> (дата обращения: 14.05.2026).
6. Merlec MM, In HP. Blockchain-Based Decentralized Storage Systems for Sustainable Data Self-Sovereignty: A Comparative Study. Sustainability. 2024; 16(17):7671. <https://doi.org/10.3390/su16177671> (дата обращения: 14.05.2026).

УДК 614.84:665.6

Лопатин С. И.

*Научный руководитель: Шелепина Н. В., д.с.-х.н., доцент
Тольяттинский государственный университет, Тольятти, Россия*

МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К ПРЕДОТВРАЩЕНИЮ И ЛИКВИДАЦИИ ПОЖАРООПАСНЫХ АВАРИЙ НА ОБЪЕКТАХ ПОДГОТОВКИ И ПЕРЕКАЧКИ НЕФТИ

Аннотация. В статье рассмотрены результаты исследования системы промышленной и пожарной безопасности на объекте подготовки и перекачки нефти, выполненного на материале ЦППН ТПП Урайнефтегаз. Выделена типовая цепочка развития аварии, включающая потерю герметичности, пролив нефти, загазованность и воспламенение. Предложен барьерно-ориентированный подход к предупреждению и локализации аварий, основанный на сочетании технологического контроля, противоаварийной автоматики, организационного реагирования и адресного усиления наиболее уязвимых узлов. Для объекта обоснованы мероприятия по развитию видеоконтроля, газоаналитического контура и визуального управления. Расчет показал, что адресное усиление существующей системы безопасности является экономически оправданным.

Ключевые слова: техносферная безопасность, нефтяной объект, промышленная безопасность, пожарная опасность, разлив нефти, противоаварийная защита, ЦППН.

Объекты подготовки и перекачки нефти относятся к числу технологически насыщенных и пожаровзрывоопасных элементов нефтегазового комплекса. На них одновременно протекают процессы сепарации, обезвоживания, обессоливания, промежуточного хранения, внешней перекачки и факельного сжигания газа. Даже локальное нарушение герметичности на одном из узлов способно быстро перерасти в комбинированную аварию, если пролив нефти сопровождается выходом газовой фазы и образованием пожароопасной среды. Нормативные требования к качеству подготовленной нефти и к безопасной эксплуатации нефтяных объектов закреплены в ГОСТ Р 51858-2020, федеральных нормах и правилах Ростехнадзора, а также в специальных сводах правил по пожарной безопасности нефтяных и газовых месторождений [1], [2], [3].

На практике ключевая проблема состоит в том, что первичный технический дефект редко остается изолированным. Коррозионная деградация трубопровода, негерметичность арматуры или отказ системы контроля уровня часто сочетаются с организационными просчетами. Запаздывает обнаружение, затягивается локализация, не срабатывает нужный алгоритм отсечения продукта. В отчете о социальной деятельности Группы Газпром за 2024 год для сегмента Газпром нефть вновь зафиксированы аварии и инциденты, а среди причин названы дефекты сварных соединений, негерметичность

арматуры и коррозионные процессы [4]. Похожие выводы содержатся и в российских публикациях, посвященных аварийности на нефтегазовых объектах [5], [6].

В связи с этим задача предупреждения пожароопасных аварий на объектах подготовки и перекачки нефти должна решаться не через разрозненные меры, а через согласованную систему барьеров. Цель статьи состоит в том, чтобы на материале ЦППН ТПП Урайнефтегаз показать, как может быть выстроен такой барьерно-ориентированный подход, какие элементы действующей системы безопасности уже обеспечивают устойчивость объекта и какие меры целесообразно реализовать для усиления наиболее уязвимых узлов.

Информационной базой исследования послужили технологический регламент ЦППН, план мероприятий по локализации и ликвидации последствий аварий и план предупреждения и ликвидации разливов нефти и нефтепродуктов, действующие на исследуемом объекте [7], [8], [9]. Эти документы позволили проанализировать фактическую конфигурацию технологической схемы, состав противоаварийных средств, порядок оповещения, тренировок и развертывания сил ликвидации. Дополнительно использованы нормативные документы и научные публикации, посвященные качеству подготовки нефти, аварийности в нефтяной отрасли и современным методам обезвоживания и электродегидрации [1], [2], [5], [10], [11].

В работе применены системный, сценарный и расчетно-аналитический методы. Системный метод использован для рассмотрения ЦППН как совокупности технологического, противоаварийного, организационного и аварийно-спасательного контуров. Сценарный метод позволил выделить цепочку развития аварии, наиболее характерную для объекта. Расчетно-аналитический метод применен при оценке экономической эффективности предложенных мероприятий на основе ожидаемого предотвращенного ущерба.

Анализ показал, что безопасность ЦППН формируется не одним устройством и не одним регламентом. На объекте действует многоуровневая система, включающая технологический контроль параметров, резервуарную противопожарную защиту, автоматическую откачку аварийных и дренажных сбросов, систему газоанализа, регламентированный порядок оповещения и регулярные учебно-тренировочные занятия [7], [8], [9]. В технологическом контуре особое значение имеют контроль давления, температуры, уровня, обводненности, межфазного раздела и загазованности. В организационном контуре ключевую роль играют заранее описанные действия первого обнаружившего аварию, начальника смены, КЧС и ПБ, НАСФ, ПАСФ и пожарной охраны.

При этом локальная документация показывает, что для объекта сохраняется единый доминирующий механизм тяжелой аварии. Он включает

потерю герметичности резервуара, трубопровода или насосной обвязки, пролив нефти, образование опасной газовой среды и последующее воспламенение. Наиболее тяжелым сценарием признано квазимгновенное разрушение резервуара РВС-10000 с выходом нефти за пределы обвалования. По материалам ПЛРН максимальный объем разлива в этом случае достигает 10000 м³, масса разлива составляет 8300 т, а суммарный ущерб может достигать 552,757 млн руб. [9]. Это означает, что даже при наличии развитой системы локализации объект нуждается в усилении упреждающих барьеров, прежде всего тех, которые работают до стадии воспламенения.

С инженерной точки зрения наиболее уязвимыми для ЦППН следует считать резервуарный парк, насосную внешнего транспорта, узел учета нефти и площадку печей. Именно здесь сочетаются большие объемы горючего продукта, перепады давления, риск локальной загазованности и необходимость быстрой изоляции аварийного участка. Действующая система защиты уже содержит ряд зрелых решений. Резервуары оснащены дыхательными клапанами с огнепреградителями, кольцом орошения и пенокамерами, насосные и учетные узлы оборудованы стационарными датчиками загазованности, аварийные потоки могут быть переведены в резервную схему [7], [8]. Однако значительная часть барьеров ориентирована на сигнализацию и организационную реакцию. Это сильная основа, но в условиях быстрого развития пролива ее недостаточно, если подтверждение аварии или поиск нужного узла управления занимают лишние минуты.

С учетом выявленных особенностей для исследуемого объекта обоснован адресный пакет мероприятий, ориентированный не на глубокую реконструкцию, а на усиление наиболее слабых звеньев. В состав пакета включены стационарный взрывозащищенный видеоконтроль, расширение сети газоанализа и внедрение инструментов визуального управления. Такой набор мер логически продолжает уже действующие решения. Видео позволяет быстрее подтвердить сам факт пролива, положение арматуры и наличие персонала в опасной зоне. Дополнительные газоанализаторы сокращают вероятность позднего обнаружения утечки в промежуточных зонах, где локальная загазованность может появиться раньше открытого пожара. Визуальное управление по принципам 5S, цветовой маркировки и QR-идентификации уменьшает потери времени при поиске аварийного инструмента, СИЗ, задвижек и маршрутов действий.

Таблица 1. Адресные мероприятия по повышению безопасности ЦППН

Мероприятие	Целевой узел	Ожидаемый эффект
Стационарные взрывозащищенные видеокамеры	РВС, НВТ, УУН, печи	сокращение времени подтверждения аварии и повышение качества визуального контроля

Мероприятие	Целевой узел	Ожидаемый эффект
Дополнительные стационарные газоанализаторы	резервуарный парк, зоны ЕПП, трубопроводный коридор	раннее выявление загазованности до перехода пролива в пожар
5S, цветовая маркировка, QR-идентификация	аварийный инструмент, СИЗ, арматура, маршруты действий	снижение латентных потерь времени и повышение готовности персонала

С точки зрения методологии принципиально важно, что перечисленные меры усиливают разные барьеры одной и той же цепочки. Видео относится к контуру подтверждения события. Газоанализ относится к контуру раннего обнаружения. Визуальное управление усиливает организационный контур и сокращает время ручной реакции. В совокупности это дает не механическую сумму технических устройств, а более плотную систему безопасности, где между первым отклонением и локализацией остается меньше незащищенных промежутков.

Экономическая оценка показала, что предложенный пакет мероприятий реалистичен для поэтапного внедрения. По расчетам, выполненным на основе рыночных цен оборудования и внутренних данных объекта, суммарные капитальные затраты составляют 1,56 млн руб. При умеренном допущении о снижении вероятности тяжелой аварии на 0,2 процентного пункта и дополнительном организационно-эксплуатационном эффекте годовой предотвращенный эффект оценивается в 2,01 млн руб. Срок окупаемости составляет 0,78 года. Для опасного производственного объекта это важный результат. Он показывает, что повышение устойчивости к пожароопасным авариям может быть достигнуто не только инженерно, но и экономически оправданно [9].

Отдельно следует отметить, что адресное усиление действующей системы безопасности не противоречит современной логике подготовки нефти. Напротив, оно дополняет ее. Исследования по обезвоживанию и электродегидрации показывают, что эффективность подготовки нефти определяется не одним параметром, а сочетанием температуры, дозы деэмульгатора, устойчивости эмульсии и режима работы оборудования [10], [11]. Для ЦППН это означает, что даже технологически качественная подготовка нефти должна сопровождаться столь же качественной системой раннего обнаружения и локализации аварии. Иначе выигрыш по технологическому режиму не превращается в выигрыш по безопасности.

Проведенное исследование позволяет сделать три основных вывода. Во-первых, для объектов подготовки и перекачки нефти ключевым механизмом тяжелой аварии остается цепочка потеря герметичности, пролив, загазованность и воспламенение. Во-вторых, действующая система безопасности ЦППН ТПП Урайнефтегаз обладает функциональной

зрелостью, однако нуждается в адресном усилении тех барьеров, которые обеспечивают раннее подтверждение аварии и сокращение времени реакции. В-третьих, наиболее рациональным направлением совершенствования является не капиталоемкая реконструкция объекта, а целенаправленное усиление наиболее уязвимых узлов. Для исследуемого объекта таким усилением являются стационарный видеоконтроль, дополнительный газоанализ и визуальное управление. Их внедрение позволяет повысить вероятность раннего обнаружения опасного состояния, улучшить управляемость аварийной ситуации и снизить риск перехода локального инцидента в крупный пожар.

Список литературы:

1. ГОСТ Р 51858-2020. Нефть. Общие технические условия [Электронный ресурс]. URL: <https://docs.cntd.ru/document/566325164> (дата обращения: 13.05.2026).
2. Об утверждении федеральных норм и правил в области промышленной безопасности «Правила безопасности в нефтяной и газовой промышленности»: приказ Ростехнадзора от 15.12.2020 № 534 [Электронный ресурс]. URL: <https://publication.pravo.gov.ru/Document/View/0001202012290079> (дата обращения: 13.05.2026).
3. СП 231.1311500.2015. Обустройство нефтяных и газовых месторождений. Требования пожарной безопасности [Электронный ресурс]. URL: <https://mchs.gov.ru/dokumenty/vse-dokumenty/6682> (дата обращения: 13.05.2026).
4. Отчет о социальной деятельности Группы Газпром за 2024 год [Электронный ресурс]. URL: <https://sustainability.gazpromreport.ru/2024/process-safety/industrial-safety/> (дата обращения: 13.05.2026).
5. Короткова Т.Г. Статистика и причины аварий на объектах нефтегазодобычи // Научные труды Кубанского государственного технологического университета. 2019. № 1. С. 115-127.
6. Кармес А. П. Технические проблемы обеспечения тушения и предотвращения пожаров на нефтегазопроводах // Пожары и ЧС. 2024. № 1 [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/tehnicheskie-problemy-obespecheniya-tusheniya-i-predotvrasheniya-pozharov-na-neftegazoprovodah> (дата обращения: 13.05.2026).
7. Geng Y., Lv C., Yuan X., Xu W. Analysis of Electric Breakup Characteristics of Emulsion Droplets in Crude Oil Dehydration under Pulsed DC Electric Field // Processes. 2024. Vol. 12. No. 7. Art. 1467. DOI: 10.3390/pr12071467.
8. Ahmadi S., Khormali A., Kazemzadeh Y., Razmjooie A. Enhancing dehydration/desalting efficiency of crude oil emulsions through experimental and computational insights // Results in Engineering. 2024. Vol. 24. Art. 103094.

УДК 519.853.6:519.633

Семёнов Д. В.

Казанский национальный исследовательский технический университет им. А. Н. Туполева
– КАИ, Казань, Россия

МОДИФИКАЦИЯ АЛГОРИТМА МИНИМИЗАЦИИ ГЛАДКОЙ ФУНКЦИИ НА ГРАНИЦЕ ОБЪЕДИНЕНИЯ ДВУХ ОБОБЩЁННЫХ СФЕРИЧЕСКИХ СЕГМЕНТОВ

Аннотация. Рассматривается модификация алгоритма минимизации гладкого функционала на границе объединения двух обобщённых сферических сегментов в конечномерном евклидовом пространстве. Предлагается метод, который не требует знания множества пересечений и позволяет корректировать параметры, отвечающие за длину шага. Проводятся численные эксперименты с целью оценки влияния предложенной модификации на производительность исходного алгоритма, приводится результирующая таблица и сравнительный анализ полученных результатов.

Ключевые слова: невыпуклые задачи оптимизации, метод проекции градиента, релаксационный метод, сходимость, сферический сегмент, метод штрафных функций, касательный конус.

Введение.

Нередко задачи условной минимизации для гладких выпуклых поверхностей требуются методы, которые, помимо основного алгоритма, контролируют длину шага, чтобы остаться в пределах допустимого множества. Зачастую отсутствие таких модификаций чревато неверным решением целевой задачи или заикливанием алгоритма в окрестности какой-то точки.

Настоящая работа рассматривает задачу модификации алгоритма минимизации гладкой функции на границе объединения двух обобщённых сферических сегментов. Здесь под “обобщённым сферическим сегментом” понимается множество, совпадающее с частью сферы, которая, во-первых, вырезается выпуклым телесным конусом с вершиной внутри этой сферы, во-вторых, расположена в конусе. Граница множества рассматривается в топологии сферы. Модификация предполагает снижение длины шага, если точка на очередной итерации вышла из области локального минимума.

За основу основного алгоритма вычисления минимума были взяты результаты работы [1], где рассматривалась задача минимизации и содержался базовый метод контроля длины шага, для которого требовалось вычисление множества точек пересечения на начальном этапе, что накладывало определённые временные затраты. Настоящая работа не требует знания множества пересечения и решает поставленную задачу посредством дополнительной проекции на неактивный конус. Решение поставленной задачи было получено, вдохновившись работой [2]

Далее дадим формальную постановку задачи. Рассмотрим задачу нахождения локального минимума гладкой функции $F: R^n \rightarrow R$, имеющей всюду в R^n липшицеву производную с положительной константой $L > 0$, на допустимом множестве $\partial_S X$:

$$F(x) \rightarrow \min, x \in \partial_S X \subset R^n, \quad (1)$$

где $\partial_S X$ – граница множества X , получаемого в результате пересечения $(n - 1)$ -мерной сферы S с объединением двух выпуклых конусов K^0, K^1 с непустыми внутренностями, вершины которых c_0 и c_1 , лежат в открытом шаре V , ограничиваемом сферой S . Причём здесь и далее через ∂_S обозначается граница множества в топологии сферы S , а через ∂ граница множества в R^n . (В дальнейшем множество $\partial_S X$ будет именоваться «границей объединения двух обобщённых сегментов сферы»). Также в данной работе рассматриваются допустимые множества, которые удовлетворяют следующим условиям:

$$K^0 \cap K^1 \cap S; S \cap (K^0 \setminus K^1); S \cap (K^1 \setminus K^0) - \text{связанные множества}; \\ \text{int}_S(K^0 \cap K^1) \neq \emptyset$$

В настоящей работе будем придерживаться следующих определений и обозначений. Для точки x_k положим: Γ_k – опорная гиперплоскость к S в точке x_k ; $H_k^{m_k}$ – опорная гиперплоскость к $K^{m_k}, m_k \in \{0,1\}$ в точке x_k ;

$$\Sigma_k^{m_k} = \Gamma_k \cap H_k^{m_k}, s_k = Pr_{\Sigma_k^{m_k}}(x_k - F'(x_k)), \quad (2)$$

$$q_k = Pr_{\Sigma_k^{m_k}}(x_k - \alpha_k f'(x_k)), t_k = Pr_{K^{m_k}} q_k,$$

$$g_k = \{c_{m_k} + \lambda(t_k - c_{m_k}), \lambda \geq 0\} \cap S,$$

$$\{w_k^0, w_k^1\} = \Sigma_{g_k}^{m_k} \cap K^{1-m_k}$$

$$h_k^i := \{c_{m_k} + \lambda(w_k^i - c_{m_k}), \lambda \geq 0\} \cap S, i = 0,1$$

$$m_k: x_k \in \partial K^{m_k}, m_k \in \{0,1\} \quad (3)$$

$$\alpha_0 > 0, \varepsilon > 0, \varepsilon_0 > 0, \quad (4)$$

$\Lambda_{g_k^j}$ – касательная $\Sigma_k^{m_k}$ в точке g_k^j , где $j = 0,1$

Задача данной работы заключается в контроле параметра α_k , который отвечает за длину шага. Необходимость данной модификации возникает по причине: выхода из зоны локального поиска и перемещении точки g_k на противоположную сторону, минуя область $S \cap (K^0 \cap K^1)$.

Описание алгоритма.

Суть решения поставленной задачи заключается в проецировании точки s_k на два конуса, активный и неактивный. После проецирования полученные точки были обозначены g_k^0 и g_k^1 .

Шаг 0. $[g_k^0 \in K_1 \cap \partial_s K_0]$: вычисление новой точки x_{k+1} согласно исходному алгоритму;

Шаг 1. $[g_k^0 \notin K_1 \cap \partial_s K_0, \text{ т. е. } g_k^0 \in (\partial_s K_0) \setminus K_1]$:

1.1 $[s_k \in \text{int}_s K_1 \text{ (рис. 6)}]$: $\alpha_k := \frac{\alpha_k}{2}$;

1.2 $[g_k^1 \in K_0 \text{ (рис. 11)}]$: вычисление новой точки x_{k+1} согласно исходному алгоритму;

1.3 $[g_k^1 \in (\partial_s K_1) \setminus K_0 \text{ (рис. 1 – 5, 7 – 10)}]$: построить $\Lambda_{g_k^1}$. Проверить условия ниже, если оно верно, выполнить инструкцию.

1.3.1 $[x_k \text{ и } s_k \text{ по одну сторону } \Lambda_{g_k^1}, \text{ в том числе при } s_k \in \Lambda_{g_k^1} \text{ (рис. 7 – 10)}]$: Принять $g_k := g_k^0$;

1.3.2 $[x_k \text{ и } s_k \text{ по разные стороны } \Lambda_{g_k^1} \text{ (рис. 1 – 5)}]$: принять $\alpha_k := \frac{\alpha_k}{2}$.

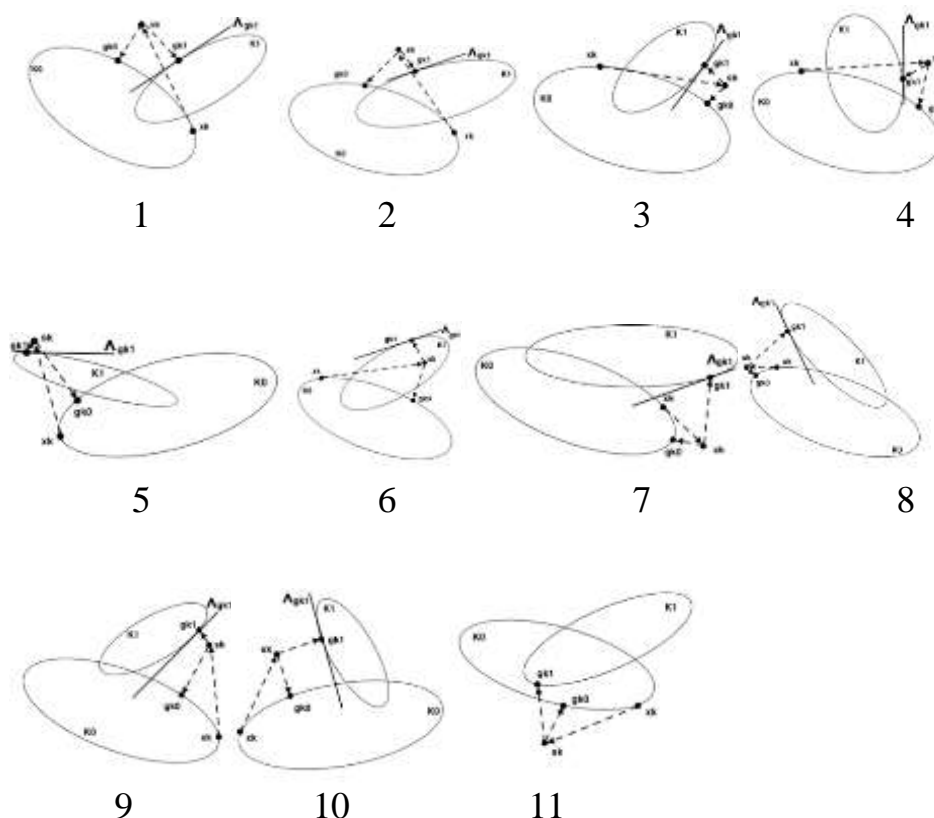


Рисунок 1. Расположение конусов

Численные эксперименты.

Разработанная модификация была протестирована на семи задачах. При сравнительном анализе сравнивалось время работы базового алгоритма, где использовался барицентрический метод [3] и дополнительно вычислялись точки пересечения множеств. [4] Эксперименты проводились на персональном компьютере с характеристиками: MacBook Pro 13-inch, 2020 года; процессор:

Intel Core i5, 1,4 ГГц, 4-ядерный; оперативная память: 8 ГБ. 2133 МГц. Исходные данные возможно наблюдать в таблице 1, все функции дифференцируемы. Алгоритм был написан на языке Python, данный выбор был обусловлен простым синтаксисом и хорошей доступностью готовых модулей. [5]

Таблица 1. Исходные данные

№	$F(x)$	x_0	K^0	K^1
1	$4x + y + z$	(1, 1.87, 2.121)	$(x + 1.5)^2 + y^2 - z^2$	$x^2 + y^2 - z^2$
2	$x + y + z$	(1, 2.3028, 1.6423)	$(x - 1)^2 + y^2 - z^2$	$x^2 + (y - 1)^2 - z^2$
3	$x + 0.5y + z$	(-2,1.6,1.6)	$(x + 2)^2 + y^2 - z^2$	$x^2 - y^2 + (z - 2)^2$
4	$-3x + 2y - 3z$	(2.1,1,2)	$x^2 + y^2 - z^2$	$x^2 + (y - 1)^2 - z^2$
5	$-5x - 2y - 3z$	(-0.7, -1.69, 2.4)	$(x - 1)^2 + y^2 - z^2$	$x^2 + (y - 1)^2 - z^2$
6	$\sin(2x + 0.8) + \sin(2y)$	(0.84, 1.95, 2.12)	$(x + 1.5)^2 + y^2 - z^2$	$x^2 + y^2 - z^2$
7	$\frac{8}{1 + 0.1((x - 5)^2 + 5)} + \frac{1}{1 + 0.1((x + 1)^2 + 5)}$	(1.22, 1.74, 2.12)	$(x + 1.5)^2 + y^2 - z^2$	$x^2 + y^2 - z^2$

Параметры для исходного алгоритма были следующие: $\alpha_0 = 0.5$. Результат сравнительной таблицы возможно наблюдать под номером 2. Стоит отметить, что выбиралось наилучшее время выполнения для каждого случая, что позволяло объективно оценивать преимущества и недостатки модификации. После просмотра полученных результатов необходимо сделать вывод о явном преимуществе разработанной модификации. Число итераций осталось прежним. Время работы алгоритма после модификации уменьшилось примерно в 1.5 раза, по сравнению с исходным алгоритмом. Данное преимущество очень существенно, когда речь заходит о стоимости оборудования. [6]

Таблица 2. Сравнительная таблица

№	До модификации		После модификации	
	Время(сек.)	Число итераций	Время(сек.)	Число итераций
1	0.0024	15	0.0016	15
2	0.0066	50	0.0045	50
3	0.0029	21	0.0021	21
4	0.0013	8	0.0009	8
5	0.0013	8	0.0009	8
6	0.0012	7	0.0008	7

№	До модификации		После модификации	
	Время(сек.)	Число итераций	Время(сек.)	Число итераций
7	0.0081	58	0.0053	58

Заключение.

По результатам таблицы 2 можно сделать следующий вывод: предлагаемая в настоящей работе модификация уменьшила временные трудозатраты исходного алгоритма. Большим преимуществом модификации является отсутствие знания множества точек пересечения, что снижает нагрузку на ресурсы ЭВМ на начальном этапе алгоритма. Полученные результаты согласуются с поставленными целями и позволяют заменить базовый алгоритм со знанием множества точек пересечения на модификацию, которая сокращает время выполнения программы.

Дальнейшие исследования темы объединения множеств автор предполагает сделать с обобщением на n -мерный случай, так как данная работа прежде всего была посвящена пространству R^3 . Также не менее важным остаётся вопрос о виде формальный условий существования точки локального минимума. [7]

Список литературы:

1. Александров, А. А. Минимизация гладкой функции на границе пересечения двух обобщенных сегментов сферы // XXIII ТУПОЛЕВСКИЕ ЧТЕНИЯ (школа молодых ученых) : Международная молодёжная научная конференция: Материалы конференции. Сборник докладов: в 4 томах, Казань, 08–10 ноября 2017 года. Том II. – Казань: Издательство Академии наук РТ, 2017. – С. 470-476
2. Dulliev, A.M. A relaxation method for minimizing a smooth function on a generalized spherical segment. *Comput. Math. and Math. Phys.* 54, 219–234 (2014). DOI: 10.1134/S0965542514020043
3. Ильинский А. С., Полянский И. С. Барицентрический метод в решении краевых задач математической физики // Дифференциальные уравнения. – 2022. – Т. 58, № 6. – С. 834–845. – DOI: 10.31857/S0374064122060097.
4. Кривцова И. Е., Лебедев И. С., Настека А. В., Основы дискретной математики. Часть 1. Учебное пособие. – СПб: Университет ИТМО, 2016. – 92 с.
5. Мэтиз Эрик. Изучаем Python: программирование игр, визуализация данных, веб-приложения. 3-е изд. – СПб.: Питер, 2020. – 512 с.
6. Ногин В. Д. Методы оптимальных решений учебное пособие / В.Д. Ногин; Гос. ун-т - Высш. школа экономики, Санкт-Петерб. фил.. — Санкт-Петербург : Ютас, 2006. — 103, [1] с. ил.; 20. — ISBN 5-91185-003-6. Васильев Ф. П. Численные методы решения экстремальных задач. – Москва “НАУКА” 1988. – 551 с.

УДК 621.314.224.8:621.316.925

Соловьев Д. Д.

Саяно-Шушенский филиал Сибирского федерального университета, Черемушки, Россия

МЕТОДЫ ОПРЕДЕЛЕНИЯ НАСЫЩЕНИЯ ТРАНСФОРМАТОРОВ ТОКА В УСТРОЙСТВАХ РЕЛЕЙНОЙ ЗАЩИТЫ

Аннотация: насыщение трансформаторов тока (ТТ) является одной из существенных причин погрешностей измерения аварийных токов в системах релейной защиты и автоматики (РЗА). Искажение вторичного тока при насыщении магнитопровода способно приводить к ложным срабатываниям, снижению чувствительности и нарушению селективности защит. В статье рассмотрены основные методы определения насыщения ТТ: анализ формы вторичного тока, методы производных, гармонический и корреляционный анализ, wavelet-преобразование, модельно-ориентированные алгоритмы, фильтры Калмана и методы машинного обучения. Выполнено сравнение подходов по точности, быстродействию, вычислительной сложности и применимости в цифровых терминалах РЗА. Показано, что классические методы сохраняют практическую значимость благодаря простоте реализации, однако наиболее перспективными являются гибридные алгоритмы, объединяющие физические модели ТТ и интеллектуальную обработку сигналов.

Ключевые слова: трансформатор тока, насыщение, релейная защита, цифровая обработка сигналов, дифференциальная защита, wavelet-анализ, Kalman Filter, машинное обучение.

Введение

Надежность релейной защиты во многом определяется точностью измерения параметров аварийных режимов, прежде всего токов короткого замыкания [1]. В большинстве устройств РЗА источником измерительной информации являются трансформаторы тока, преобразующие большие первичные токи в стандартизированные вторичные сигналы [2]. В нормальном режиме вторичный ток пропорционален первичному, однако при переходных процессах это соответствие может нарушаться.

При коротких замыканиях ТТ работают в условиях значительных токовых перегрузок. Аperiodическая составляющая тока КЗ вызывает смещение магнитного потока в сердечнике и может привести к насыщению магнитопровода [3]. В насыщенном режиме часть первичного тока расходуется на ток намагничивания, вследствие чего вторичный ток искажается: уменьшается амплитуда, нарушается фазовое соответствие, появляются срезы кривой и высшие гармоники [4].

Насыщение ТТ особенно опасно для дифференциальных защит трансформаторов и шин, поскольку при внешнем коротком замыкании неодинаковое насыщение измерительных трансформаторов может вызвать ложный ток небаланса [5]. В дистанционных и токовых защитах насыщение

приводит к ошибкам измерения тока и сопротивления повреждения, что может вызвать как ложное срабатывание, так и отказ защиты [6].

Актуальность проблемы возрастает в связи с развитием цифровых подстанций, увеличением уровней токов КЗ и широким внедрением микропроцессорных терминалов РЗА [7]. Современные исследования направлены на моделирование насыщения, разработку алгоритмов его быстрого обнаружения и восстановление искаженного вторичного тока. При этом отсутствует универсальный метод, одинаково эффективный для всех типов ТТ и режимов повреждения [8].

Цель статьи — систематизировать современные методы определения насыщения трансформаторов тока, раскрыть их физическую и алгоритмическую сущность, а также определить области рационального применения в устройствах цифровой релейной защиты.

1. Физические основы насыщения ТТ и его влияние на РЗА

Трансформатор тока представляет собой электромагнитное устройство, работа которого основана на законе электромагнитной индукции. В идеальном случае ток во вторичной обмотке повторяет форму первичного тока с учетом коэффициента трансформации [9]. Реальный магнитопровод имеет нелинейную характеристику намагничивания. При достижении определенного значения магнитной индукции магнитная проницаемость резко уменьшается, ток намагничивания возрастает, а ТТ переходит в режим насыщения [10].

Основными причинами насыщения являются высокая кратность тока КЗ, наличие апериодической составляющей, остаточная намагниченность сердечника, повышенное сопротивление вторичной цепи и неправильный выбор класса точности ТТ. Наиболее тяжелые условия возникают при однополярном смещении магнитного потока, вызванном апериодической составляющей. Остаточная индукция также существенно влияет на глубину насыщения и может увеличивать ток небаланса дифференциальной защиты [11], [12].

Признаками насыщения являются искажение формы вторичного тока, снижение амплитуды, появление срезов полуволн, рост высших гармоник, резкое изменение производной и увеличение фазовой погрешности. Эти признаки лежат в основе большинства алгоритмов обнаружения насыщения.

Для дифференциальных защит наиболее опасна ситуация, когда при внешнем КЗ один ТТ насыщается раньше или глубже другого. В результате появляется ложный дифференциальный ток, снижающий устойчивость защиты. Для дистанционных защит насыщение опасно фазовыми и амплитудными ошибками, влияющими на расчет полного сопротивления до места повреждения [13]–[15].

2. Методы анализа формы сигнала и производных

2.1. Анализ формы вторичного тока

Методы анализа формы сигнала относятся к наиболее простым способам определения насыщения. Их основная идея состоит в том, что при нормальной

работе вторичный ток должен сохранять форму первичного тока. При насыщении эта пропорциональность нарушается, так как значительная часть магнитодвижущей силы расходуется на намагничивание сердечника [16], [17].

На практике насыщение проявляется срезом полуволны, резким изломом кривой, уменьшением крутизны изменения тока и нарушением симметрии положительной и отрицательной полуволн. Алгоритм обычно включает дискретизацию вторичного тока, сравнение текущих отсчетов с ожидаемой формой сигнала или предыдущими отсчетами, а затем проверку превышения порогового отклонения. Для исключения ложных решений дополнительно анализируется длительность искаженного участка.

Преимущество метода — простота и высокая скорость работы. Он не требует точной модели ТТ и может быть реализован в цифровом терминале с минимальными вычислительными затратами. Недостатки связаны с чувствительностью к шумам, частоте дискретизации, фильтрации входных цепей и выбору порогов. Кроме того, форма сигнала может искажаться не только из-за насыщения, но и вследствие апериодической составляющей, коммутационных процессов или помех.

2.2. Методы первой и второй производной

Методы производных являются развитием анализа формы сигнала. Первая производная характеризует скорость изменения тока, а вторая — изменение его крутизны. В момент насыщения на кривой вторичного тока возникает излом: скорость изменения сигнала резко уменьшается, а вторая производная изменяется скачкообразно [17].

Типовой алгоритм включает вычисление первой и второй разностей между соседними отсчетами, сравнение их с адаптивным или фиксированным порогом и подтверждение насыщения при сохранении признака в течение нескольких отсчетов. Особенно важна вторая разность, поскольку она хорошо фиксирует момент перехода от ненасыщенного участка к насыщенному.

Достоинство подхода — возможность определить не только факт, но и момент начала насыщения, что необходимо для последующей компенсации и восстановления вторичного тока. Основное ограничение — усиление высокочастотных помех при дифференцировании. Поэтому такие алгоритмы обычно требуют предварительной цифровой фильтрации и корректного выбора порогов с учетом частоты дискретизации и ожидаемой кратности тока КЗ.

3. Спектральные и корреляционные методы

3.1. Гармонический анализ

Гармонические методы основаны на том, что насыщенный ТТ формирует несинусоидальный вторичный ток. Искажение формы приводит к появлению дополнительных составляющих спектра, прежде всего второй, третьей и более высоких гармоник [18], [19]. Суть метода заключается в разложении сигнала с помощью дискретного преобразования Фурье и вычислении отношения амплитуд отдельных гармоник к основной гармонике.

Наиболее информативными являются вторая гармоника, связанная с асимметрией и однополярным насыщением, третья гармоника, характерная для нелинейных магнитных процессов, а также суммарный коэффициент гармонических искажений. Если выбранный показатель превышает порог, фиксируется признак насыщения.

Достоинство гармонического анализа — устойчивость к одиночным помехам и наличие хорошо отработанных цифровых алгоритмов. Недостаток — необходимость накопления временного окна, что увеличивает задержку обнаружения. Кроме того, высшие гармоники могут возникать не только при насыщении ТТ, но и при броске тока намагничивания силового трансформатора, феррорезонансе, нелинейной нагрузке и коммутационных процессах. Поэтому гармонический анализ целесообразно применять совместно с другими признаками.

3.2. Корреляционные методы

Корреляционные методы сравнивают текущий вторичный ток с эталонным сигналом: синусоидой промышленной частоты, прогнозируемой формой тока или сигналом другого канала [21]. В нормальном режиме коэффициент корреляции между измеренным и эталонным сигналом остается высоким. При насыщении форма тока искажается, и корреляция резко снижается.

Алгоритм включает формирование эталонной последовательности, выбор скользящего окна, расчет коэффициента корреляции и сравнение результата с порогом. Преимущество метода состоит в анализе формы сигнала в целом, а не отдельных отсчетов, что повышает устойчивость к одиночным выбросам и шуму. Основное ограничение — необходимость корректного выбора эталона. Простая синусоида не всегда подходит для аварийных токов с аperiodической составляющей, поэтому более эффективны адаптивные эталоны, формируемые по ненасыщенному участку сигнала.

4. Wavelet-анализ и модельно-ориентированные методы

4.1. Wavelet-анализ

Wavelet-анализ используется для обнаружения кратковременных локальных искажений, возникающих при входе ТТ в насыщение. В отличие от преобразования Фурье, wavelet-преобразование позволяет определить не только частотный состав сигнала, но и момент появления соответствующих составляющих [22]. Это важно, поскольку насыщение возникает на ограниченном участке периода и сопровождается локальными высокочастотными компонентами.

Суть метода заключается в разложении вторичного тока по набору wavelet-функций разных масштабов. Низкочастотные коэффициенты описывают основную форму тока, а высокочастотные коэффициенты отражают изломы и резкие изменения. При насыщении энергия коэффициентов детализации возрастает, что используется как диагностический признак.

Типовая процедура включает выбор базисной функции, разложение сигнала на уровни, выделение коэффициентов детализации, расчет их энергии и сравнение с порогом. Wavelet-коэффициенты могут использоваться самостоятельно или как входные признаки для нейронных сетей, SVM и других методов машинного обучения [23]. Достоинства метода — высокая точность локализации момента насыщения и пригодность для нестационарных сигналов. Недостатки — зависимость результата от выбора wavelet-функции, числа уровней разложения и пороговых значений, а также повышенная вычислительная сложность.

4.2. Методы на основе математической модели ТТ

Модельно-ориентированные методы стремятся оценить внутреннее состояние ТТ: магнитный поток, ток намагничивания и степень насыщения сердечника [24]. Основная идея состоит в том, что насыщение возникает при превышении магнитным потоком некоторого предельного значения. Если по измеренному вторичному току и параметрам вторичной цепи рассчитать поток, можно определить приближение магнитопровода к насыщению.

В простейшем виде магнитный поток оценивается путем интегрирования напряжения на вторичной обмотке. Напряжение определяется по падению напряжения на сопротивлении вторичной цепи и индуктивных элементах. Далее рассчитанный поток сопоставляется с характеристикой намагничивания. В литературе применяются кусочно-линейные модели, модели Фрелиха и Джайлса–Атертона, топологические схемы магнитной цепи и детализированные модели с учетом гистерезиса и потерь в стали.

Преимущество модельного подхода — физическая обоснованность и возможность не только обнаружить насыщение, но и оценить его глубину, а также восстановить искаженный ток. Основное ограничение — необходимость знать параметры конкретного ТТ: сопротивление обмотки, индуктивность рассеяния, характеристику намагничивания, остаточную индукцию и нагрузку вторичной цепи. В реальной эксплуатации эти параметры могут отличаться от паспортных, что снижает точность модели [25].

5. Фильтры Калмана и методы машинного обучения

5.1. Фильтры Калмана

Фильтры Калмана применяются для оценки скрытых переменных динамической системы по измерениям, содержащим шум. В задаче насыщения ТТ такими переменными являются магнитный поток, ток намагничивания и истинный первичный ток [26]. На каждом шаге фильтр выполняет прогноз состояния по модели ТТ, затем корректирует его по измеренному вторичному току. Резкое расхождение между расчетным и измеренным сигналом может рассматриваться как признак насыщения.

Для нелинейных моделей чаще применяются Extended Kalman Filter и Unscented Kalman Filter. Первый использует линеаризацию модели около текущей точки, второй учитывает нелинейность через набор специально

выбранных точек состояния. Преимущество подхода — возможность одновременно обнаруживать насыщение и восстанавливать искаженные участки тока. Недостатки — вычислительная сложность и зависимость результата от корректности модели и параметров фильтра [27].

5.2. Искусственные нейронные сети

Нейронные сети распознают насыщение по совокупности признаков вторичного тока. В отличие от классических методов, где критерий задается вручную, сеть обучается на наборе примеров: нормальный режим, КЗ без насыщения, частичное и глубокое насыщение [28]. Входными признаками могут быть мгновенные значения тока, амплитуды гармоник, wavelet-коэффициенты, производные, энергия сигнала и параметры аperiodической составляющей.

После обучения сеть способна классифицировать новые сигналы, а в отдельных схемах — оценивать глубину насыщения или восстанавливать неискаженную форму тока. Преимущество нейронных сетей состоит в способности выявлять сложные нелинейные зависимости, обусловленные совместным влиянием тока КЗ, остаточной индукции, сопротивления нагрузки и характеристики магнитопровода. Недостатки — зависимость от качества обучающей выборки, риск ошибок на режимах, отсутствовавших при обучении, и ограниченная интерпретируемость результата.

5.3. SVM, Decision Tree и GMM

Метод опорных векторов (SVM) строит разделяющую поверхность между классами «насыщение» и «отсутствие насыщения» в пространстве признаков [29]. В качестве признаков используются гармонические коэффициенты, wavelet-параметры, производные, коэффициент корреляции и статистические показатели. SVM эффективен при ограниченной, но качественно размеченной выборке; его ограничения связаны с выбором ядра и зависимостью от предварительного выделения признаков.

Деревья решений выполняют последовательную проверку условий относительно признаков сигнала, например амплитуды второй гармоники, энергии wavelet-коэффициентов или величины второй производной [30]. Их главное достоинство — интерпретируемость. Ансамблевые методы, такие как Random Forest и Gradient Boosting, повышают точность за счет использования множества деревьев, но требуют большего объема памяти и проверки устойчивости.

Gaussian Mixture Models (GMM) относятся к вероятностным методам. Они представляют распределение признаков как сумму нормальных распределений, соответствующих различным состояниям ТТ: нормальной работе, частичному и глубокому насыщению [31]. Преимущество GMM — возможность получить не только класс, но и вероятность принадлежности к нему. Ограничение метода связано с предположением о форме распределения признаков и необходимостью достаточной обучающей выборки.

6. Методы компенсации насыщения

Обнаружение насыщения часто является только первым этапом, поскольку для корректной работы защиты требуется восстановить неискаженную форму вторичного тока [32]. Простые методы используют интерполяцию насыщенного участка по соседним ненасыщенным фрагментам. Они эффективны при кратковременном насыщении, но недостаточны при длительных и глубоких искажениях.

Более сложные методы применяют цифровую фильтрацию и прогнозирование формы тока. По ненасыщенному участку оцениваются амплитуда, фаза и апериодическая составляющая, после чего рассчитывается ожидаемый ток в насыщенной области. Модельные методы дополнительно используют оценку магнитного потока и тока намагничивания; вторичный ток корректируется с учетом той части первичного тока, которая была затрачена на насыщение сердечника.

Адаптивная цифровая компенсация отличается тем, что параметры алгоритма изменяются в процессе работы [33]. Она может включать оценку текущего магнитного потока, расчет остаточной намагниченности, коррекцию коэффициентов фильтра, изменение порогов обнаружения и обновление параметров модели ТТ. Преимущество адаптивных методов — способность работать в широком диапазоне режимов; недостаток — повышенная вычислительная сложность и необходимость защиты от ошибочной подстройки к искаженному сигналу.

Наиболее перспективными являются гибридные методы, объединяющие физическую модель ТТ, цифровую обработку сигналов и машинное обучение. Например, wavelet-анализ может использоваться для определения момента насыщения, модель магнитного потока — для восстановления тока, а нейронная сеть — для уточнения глубины насыщения. Такие алгоритмы позволяют не только фиксировать факт насыщения, но и корректировать измерительный сигнал в реальном времени. Их основное ограничение — сложность реализации и необходимость проверки на большом массиве аварийных осциллограмм.

7. Сравнительный анализ методов

Метод	Сущность	Преимущества	Ограничения	Применимость в РЗА
Анализ формы сигнала	Выявление срезов, изломов и асимметрии вторичного тока	Простота, быстродействие	Чувствительность к шумам и порогам	Высокая
Производные	Контроль первой и второй разности сигнала	Точное определение момента насыщения	Усиление помех, нужна фильтрация	Высокая

Метод	Сущность	Преимущества	Ограничения	Применимость в РЗА
Гармонический анализ	Оценка высших гармоник и THD	Устойчивость к одиночным выбросам	Задержка из-за окна, неоднозначность гармоник	Средняя
Корреляционный анализ	Сравнение сигнала эталоном	Анализ формы в целом, помехоустойчивость	Требуется корректный эталон	Высокая
Wavelet-анализ	Выделение локальных высокочастотных искажений	Хорошая временная локализация	Сложность выбора параметров	Высокая
Математическая модель ТТ	Оценка потока и тока намагничивания	Физическая обоснованность, компенсация	Требуется параметры ТТ	Средняя/высокая
Фильтр Калмана	Оценка скрытых состояний по модели и измерениям	Обнаружение и восстановление тока	Сложность и зависимость от модели	Высокая
ANN/SVM/DT/GMM	Классификация по набору признаков	Высокая точность на обученных режимах	Требуется качественная выборка	Перспективная
Гибридные методы	Совмещение модели, DSP и ИИ	Наибольшая гибкость	Сложность реализации и верификации	Перспективная

Из таблицы видно, что классические методы целесообразны для быстрого первичного обнаружения насыщения, а интеллектуальные и модельные алгоритмы — для уточнения состояния ТТ и компенсации искажений. На практике наиболее рациональным является комбинированный подход, при котором быстрые признаки используются для детектирования, а более сложные методы — для восстановления сигнала и повышения устойчивости защиты.

8. Современные тенденции развития

Современные исследования направлены на создание адаптивных алгоритмов, способных учитывать параметры конкретного ТТ и изменяющиеся условия аварийного процесса. Особое внимание уделяется использованию wavelet-признаков, фильтров Калмана, нейронных сетей, SVM, деревьев решений и вероятностных моделей. Развитие вычислительных

возможностей цифровых терминалов позволяет переносить такие алгоритмы из исследовательских моделей в практические устройства РЗА.

Перспективным направлением является интеграция алгоритмов определения насыщения в архитектуру цифровых подстанций, включая обработку потоков дискретизированных значений. В этих условиях возрастает значение online-алгоритмов, способных работать с ограниченной задержкой и сохранять устойчивость при шуме, неполных данных и изменении параметров вторичной цепи.

Наиболее важными задачами дальнейших исследований являются верификация алгоритмов на реальных аварийных осциллограммах, учет остаточной намагниченности, повышение интерпретируемости ML-моделей и разработка гибридных методов, сочетающих физическую обоснованность и адаптивность искусственного интеллекта.

Заключение: Насыщение трансформаторов тока остается одной из значимых причин ошибок в работе релейной защиты. При переходных процессах оно вызывает амплитудные и фазовые искажения вторичного тока, что может приводить к ложным срабатываниям, отказам и нарушению селективности.

Проведенный обзор показал, что методы определения насыщения можно разделить на классические, модельно-ориентированные и интеллектуальные. Методы анализа формы сигнала и производных отличаются простотой и быстроедействием, но чувствительны к шумам и выбору порогов. Гармонический и корреляционный анализ повышают устойчивость, однако требуют корректной интерпретации признаков. Wavelet-анализ обеспечивает хорошую локализацию момента насыщения, но сложнее в настройке. Модельные методы и фильтры Калмана позволяют не только обнаруживать насыщение, но и восстанавливать искаженный ток. Методы машинного обучения перспективны для сложных нелинейных режимов, однако нуждаются в качественной обучающей базе и независимой проверке.

Наиболее рациональным направлением развития являются гибридные алгоритмы, объединяющие физическую модель ТТ, цифровую обработку сигналов и методы искусственного интеллекта. Их применение может повысить надежность цифровой релейной защиты при внешних коротких замыканиях и сложных переходных режимах.

Список литературы:

1. Афанасьев В. В., Адоньев Н. М., Кибель В. М. и др. Трансформаторы тока. – 2-е изд., перераб. и доп. – Л.: Энергоатомиздат, Ленинградское отделение, 1989. – 416 с. – ISBN 5-283-04444-0.
2. Расщепляев А. И., Перельман Л. М. Динамические характеристики трансформаторов тока. – М.: Энергия, 1966. – 192 с.
3. Гуревич В. И. Микропроцессорные реле защиты: устройство, проблемы, перспективы. – М.; Вологда: Инфра-Инженерия, 2011. – 336 с. – ISBN 978-5-9729-0043-5.

4. Шнеерсон Э. М. Цифровая релейная защита. – М.: Энергоатомиздат, 2007. – 549 с. – ISBN 978-5-283-03256-6.
5. Новобрицкий В. А., Федосов Д. С. Анализ работы устройств релейной защиты в переходном режиме, сопровождающемся насыщением трансформатора тока // Известия высших учебных заведений. Проблемы энергетики. – 2021. – Т. 23. – № 5. – С. 71–85. – DOI: 10.30724/1998-9903-2021-23-5-71-85. – URL: <https://www.energyret.ru/jour/article/view/1985>.
6. Вихарев Д. Ю., Родин Н. А. Алгоритм восстановления приведенного первичного тока при насыщении электромагнитного трансформатора тока без использования характеристики намагничивания // Известия РАН. Энергетика. – 2022. – № 6. – С. 36–45. – DOI: 10.31857/S0002331022060061. – URL: <https://sciencejournals.ru/view-article/?a=IzEn2206006Vikharev&j=izen&n=6&v=0&y=2022>.
7. Rebizant W., Szafran J., Wiszniewski A. Digital Signal Processing in Power System Protection and Control. – London: Springer, 2011. – 316 p. – ISBN 978-0-85729-801-0. – DOI: 10.1007/978-0-85729-802-7.
8. McLaren P. G., Mustaphi K., Benmouyal G., Chano S., Girgis A., Henville C., Kezunovic M. et al. Software Models for Relays // IEEE Transactions on Power Delivery. – 2001. – Vol. 16. – № 2. – P. 238–245. – DOI: 10.1109/61.915489.
9. Kang Y. C., Ok S. H., Kang S. H. A CT Saturation Detection Algorithm // IEEE Transactions on Power Delivery. – 2004. – Vol. 19. – № 1. – P. 78–85. – DOI: 10.1109/TPWRD.2003.820200.
10. Kang Y. C., Lim U. J., Kang S. H., Crossley P. A. Compensation of the Distortion in the Secondary Current Caused by Saturation and Remanence in a CT // IEEE Transactions on Power Delivery. – 2004. – Vol. 19. – № 4. – P. 1642–1649. – DOI: 10.1109/TPWRD.2004.835266.
11. Hong Y. Y., Chang-Chian P. C. Detection and Correction of Distorted Current Transformer Current Using Wavelet Transform and Artificial Intelligence // IET Generation, Transmission & Distribution. – 2008. – Vol. 2. – № 4. – P. 566–575. – DOI: 10.1049/iet-gtd:20070383.
12. Moghimi Haji M., Vahidi B., Hosseinian S. H. Current Transformer Saturation Detection Using Gaussian Mixture Models // Journal of Applied Research and Technology. – 2013. – Vol. 11. – № 1. – P. 79–87. – DOI: 10.1016/S1665-6423(13)71516-5. – URL: <https://www.sciencedirect.com/science/article/pii/S1665642313715165>.
13. Rebizant W., Bejmert D. Current-Transformer Saturation Detection With Genetically Optimized Neural Networks // IEEE Transactions on Power Delivery. – 2007. – Vol. 22. – № 2. – P. 820–827. – DOI: 10.1109/TPWRD.2007.893363.
14. Key S., Ko C.-S., Song K.-J., Nam S.-R. Fast Detection of Current Transformer Saturation Using Stacked Denoising Autoencoders // Energies. – 2023. – Vol. 16. – № 3. – Article 1528. – DOI: 10.3390/en16031528. – URL: <https://www.mdpi.com/1996-1073/16/3/1528>.
15. Odinaev I., Abdel-Aty A.-H., Pazderin A., Safaraliev M., Matrenin P., Senyuk M., Menaem A. A., Kanan M. Restoration of the Current Transformer Secondary Current under Core Saturation Conditions Based on ANN // Heliyon. – 2024. – Vol. 10. – № 18. – Article e37960. – DOI: 10.1016/j.heliyon.2024.e37960. – URL: <https://www.sciencedirect.com/science/article/pii/S2405844024139916>.
16. Fernández C. An Impedance-Based CT Saturation Detection Algorithm for Bus-Bar Differential Protection // IEEE Transactions on Power Delivery. – 2001. – Vol. 16. – № 4. – P. 468–472. – DOI: 10.1109/61.956722. – URL: https://www.iit.comillas.edu/publicacion/revista/en/156/An_impedance-based_CT_saturation_detection_algorithm_for_bus-bar_differential_protection.

УДК 004.8:159.9

Горбунов С. М., Душкин Р. В.

Национальный исследовательский ядерный университет «МИФИ», Москва, Россия

РЕТРОСПЕКТИВНАЯ МЕТА-АНАЛИТИКА КАК ФАКТОР ПРЕВОСХОДСТВА: АРХИТЕКТУРА МЕТАКОГНИТИВНОЙ ПРОМПТ-ИНЖЕНЕРИИ ПРОТИВ ПРЯМОГО ПРОМПТИНГА

Аннотация: современные большие языковые модели (БЯМ) демонстрируют выдающиеся способности в обработке естественного языка, однако их применение для глубокого ретроспективного анализа диалоговых взаимодействий часто ограничивается поверхностной статистикой при использовании методов прямого промптинга. В данной работе предлагается обобщенная архитектура метакогнитивной промпт-инженерии, реализующая принцип ретроспективной мета-аналитики через многоуровневый агентный подход. Архитектура включает четыре последовательных уровня когнитивной обработки: от структурного выделения сущностей до выявления латентных паттернов (скрытых мотиваций, когнитивных искажений) и генерации превентивных стратегий. Экспериментальное исследование, проведенное на выборке из 10 разнопредметных диалогов (маркетинг, образование, разработка ПО, медицина), продемонстрировало статистически значимое превосходство предложенного метода над базовой линией прямого промптинга. Агентная система показала прирост полноты выявления инсайтов на 18,1% (0,925 против 0,783) и точности классификации на 28,0% (0,925 против 0,723). Особый вклад работы заключается в демонстрации универсальности подхода: архитектура эффективно адаптируется к различным доменам без необходимости дообучения модели, обеспечивая глубокую семантическую интерпретацию коммуникационных процессов. Результаты подтверждают гипотезу о том, что структурированная декомпозиция задачи промпт-инжиниринга является критическим фактором повышения качества аналитических выводов ИИ.

Ключевые слова: метакогнитивная промпт-инженерия, ретроспективный анализ, большие языковые модели, агентные системы, анализ диалогов, когнитивные искажения, автоматизированная аналитика.

1. Введение

Накопление массивов неструктурированных текстовых данных в виде протоколов совещаний, образовательных сессий, консультационных диалогов и технической переписки создает беспрецедентные возможности для извлечения организационных знаний. Однако традиционные методы анализа, основанные на ключевых словах или простой статистике частотности, оказываются неспособными выявить скрытые закономерности человеческой коммуникации: динамику конфликтов, латентные мотивации участников и системные когнитивные ошибки [1]. Появление больших языковых моделей (БЯМ) открыло новые горизонты в этой области, позволив перейти от синтаксического к семантическому анализу. Тем не менее, эмпирические

данные свидетельствуют о том, что использование БЯМ в режиме «прямого промптинга» (direct prompting), когда модель получает единственную инструкцию выполнить сложный анализ, часто приводит к галлюцинациям, потере контекста и поверхностным выводам [2, 3].

Проблема заключается в несоответствии между когнитивной сложностью задачи ретроспективного мета-анализа и одношаговой природой стандартного взаимодействия с БЯМ. Задача выявления не только явных тем, но и имплицитных психологических паттернов требует имитации рефлексивного процесса, свойственного человеческому эксперту. Существующие решения часто фокусируются либо на узкоспециализированных задачах (например, только распознавание эмоций [4]), либо требуют дорогостоящего дообучения моделей (fine-tuning), что снижает их универсальность и воспроизводимость.

Целью данного исследования является разработка и экспериментальная верификация обобщенной архитектуры метакогнитивной промпт-инженерии, реализующей метод ретроспективной мета-аналитики. В отличие от специализированных решений, предлагаемая архитектура представляет собой универсальный каркас, применимый к широкому спектру предметных областей.

Вклад статьи заключается в следующем:

1. Предложена формализованная четырехуровневая архитектура агентной системы, реализующая принцип постепенного углубления анализа (от структуры к латентным смыслам).

2. Проведено сравнительное экспериментальное исследование, количественно подтверждающее превосходство метакогнитивного подхода над прямым промптингом по метрикам полноты, точности и глубины анализа.

3. Продемонстрирована применимость метода на разнородном датасете, включающем диалоги из сфер маркетинга, образования, IT-разработки, медицины и управления кризисами.

4. Предоставлен открытый набор промптов и методология оценки, обеспечивающие воспроизводимость результатов.

2. Связанные работы

Исследования в области улучшения рассуждений БЯМ можно разделить на несколько направлений. Наиболее близким к нашей работе является направление Chain-of-Thought (CoT), которое предполагает разбиение задачи на промежуточные шаги рассуждения [5]. Хотя CoT улучшает логические выводы в математических задачах, его применение к качественному анализу текста часто остается линейным и не включает механизм рефлексии над собственными выводами.

Концепция метакогниции в ИИ, определяемая как способность системы мониторить и контролировать свои когнитивные процессы, активно развивается в последних работах [5, 6]. Исследования показывают, что

внедрение этапов самопроверки повышает надежность ответов [7]. Однако большинство существующих подходов фокусируются на оценке уверенности модели в рамках одного запроса, тогда как ретроспективная мета-аналитика требует анализа исторического контекста и выявления паттернов across multiple interactions.

Работа Душкина [8] вводит термин «метакогнитивная промпт-инженерия»

и описывает, в частности, технику ретроспективной мета-аналитики, которая предполагает последовательный многоуровневый анализ диалогов (от структурной разметки до выявления скрытых мотивов и когнитивных искажений). Наше исследование не просто «развивает эту идею», а реализует её в виде строгой четырехуровневой агентной архитектуры, обобщая тем самым технику ретроспективной мета-аналитики МКПИ на различные предметные области.

В области анализа диалогов традиционные подходы используют классификаторы тональности или извлечение именованных сущностей [9]. Недавние работы применяют БЯМ для суммаризации встреч [10], но они редко затрагивают глубинный психологический анализ (когнитивные искажения, скрытые мотивы), который является ключевым элементом предлагаемой нами методики. Предлагаемая архитектура заполняет этот пробел, объединяя глубину психологического анализа с масштабируемостью агентных систем.

Кроме того, предложенная архитектура развивает идеи, заложенные в подходах Self-Refine [6] и Reflexion [11], которые используют итеративную рефлексию модели над собственными выводами. Однако в отличие от них, наша система разделяет когнитивные этапы на уровне архитектуры агента с явной передачей контекста между уровнями, что позволяет избежать накопления ошибок на ранних стадиях.

3. Методология

3.1. Архитектура системы

Предлагаемая система построена по модульному принципу и реализует последовательную цепочку вызовов БЯМ с нарастающей сложностью промптов. Архитектура состоит из четырех функциональных уровней, каждый из которых решает специфическую подзадачу общего анализа.

Пусть $D = \{d_1, d_2, \dots, d_n\}$ — множество входных диалогов. Процесс анализа каждого диалога d_i описывается функцией композиции:

$$A(d_i) = L_4(d_i, L_3(d_i, L_2(d_i, L_1(d_i))))$$

где L_k — функция генерации ответа БЯМ на k -м уровне с использованием специфического промпта P_k и контекста C_{k-1} , полученного на предыдущем этапе.

Программная реализация архитектуры представляет собой агентную систему, разработанную на языке Python. Взаимодействие с языковой

моделью осуществляется через унифицированный класс *LLMClient*, поддерживающий различные провайдеры (*GigaChat*, *OpenAI*, *Anthropic*). Каждый уровень L_k получает на вход исходный текст диалога d_i и структурированный вывод всех предыдущих уровней, объединенный в единый контекст C_{k-1} . Передача данных между уровнями осуществляется в формате *JSON*, что обеспечивает строгую типизацию и минимизирует риск неверной интерпретации моделью промежуточных результатов.

Промпты уровней сконструированы таким образом, чтобы каждый последующий шаг явно ссылался на результаты предыдущего. Например, промт уровня 2 содержит инструкцию: «Используй данные структурного анализа (темы, участников, роли) для выявления процессуальных паттернов». Это позволяет модели опираться на уже проверенные факты, снижая вероятность галлюцинаций. Для безопасной обработки ответов модели используется функция `_safe_parse_json`, которая очищает вывод от маркдауна, исправляет типичные синтаксические ошибки и извлекает *JS*-объект.

Агент не требует дообучения модели (*fine – tuning*) — вся адаптация к предметной области достигается за счёт проектирования промптов и последовательной передачи контекста. В программной реализации предусмотрена обработка сбоев: при недостижимости *API* или невалидном *JSON* ответе система выполняет повторный вызов с модифицированной инструкцией (до 3 попыток). Все промежуточные результаты сохраняются в лог для обеспечения воспроизводимости.

Уровень 1: Структурный анализ (L_1)

Задача: Извлечение явной информации.

Вход: Текст диалога d_i .

Выход: Структурированные данные S_i : темы, участники, роли, ключевые результаты, временные метки.

Формализация: $S_i = LLM(P_1, d_i)$, где P_1 содержит инструкции по экстракции сущностей в формате *JSON*.

Уровень 2: Процессуальный анализ (L_2)

Задача: Анализ динамики взаимодействия.

Вход: Текст диалога d_i + Контекст S_i .

Выход: Процессуальные паттерны Pr_i : стратегии аргументации, паттерны вопросов-ответов, точки конфликтов и методы их разрешения.

Формализация: $Pr_i = LLM(P_2, d_i, S_i)$. На этом этапе модель анализирует не «что сказано», а «как происходило взаимодействие».

Уровень 3: Глубинная аналитика (L_3)

Задача: Выявление латентных факторов.

Вход: Текст диалога d_i + Контекст S_i, Pr_i .

Выход: Глубинные инсайты Dp_i : скрытые мотивации участников, выявленные когнитивные искажения «когнитивные искажения (bias)».

системные ошибки принятия решений, повторяющиеся поведенческие шаблоны.

Формализация: $Dp_i = LLM(P_3, d_i, S_i, Pr_i)$. Это ключевой этап метакогниции, требующий от модели абстрактного мышления и применения знаний из психологии.

Уровень 4: Синтез и рекомендации (L_4)

Задача: Генерация практических выводов.

Вход: Агрегированные данные $\{S_i, Pr_i, Dp_i\}$.

Выход: Итоговый отчет R_i : лучшие практики, профили эффективных подходов, стратегии предотвращения ошибок.

Формализация: $R_i = LLM(P_4, Context_{aggregate})$.

Здесь:

- S_i — структурное представление диалога (темы, участники, роли, ключевые результаты, временные метки);
- Pr_i — процессуальные паттерны (стратегии аргументации, паттерны вопросов-ответов, конфликтные точки, способы разрешения);
- Dp_i — глубинные инсайты (скрытые мотивации, когнитивные искажения, системные ошибки, поведенческие шаблоны);
- R_i — итоговый отчёт (лучшие практики, профили эффективных подходов, рекомендации, стратегии предотвращения).

Важной особенностью разработанных промптов является их лаконичность. В ходе предварительных экспериментов выяснилось, что чрезмерно детализированные инструкции (объёмом более 800–1000 токенов) приводят к снижению точности: модель «теряет» релевантные фрагменты диалога в контекстном окне, начинает галлюцинировать или игнорировать часть требований. Оптимальная длина каждого промпта, использованная в финальных экспериментах, составила 350–500 токенов, что позволило сохранить баланс между полнотой инструкций и качеством удержания контекста диалога.

3.2. Датасет и условия эксперимента

Для проверки гипотезы был сформирован тестовый набор данных, включающий 10 диалогов различной длины и предметной области:

1. Маркетинг: Планирование запуска продукта.
2. Образование/Консалтинг: Оптимизация бизнес-процессов.
3. Образование: Разбор академической задолженности.
4. IT-разработка: Конфликт версий зависимостей.
5. Переговоры: Согласование цены и условий поставки.
6. Медицина: Консультация пациента с неспецифическими симптомами.
7. Планирование: Отчетность при недостатке данных.
8. Обучение (Code Review): Обратная связь ментора стажеру.

9. Управление кризисами: Срыв сроков проекта.
10. Инновации: Обсуждение рисков новой методики.

Диалоги были размечены экспертами (Ground Truth) по категориям: темы, конфликты, скрытые мотивации, когнитивные искажения и рекомендации.

Разметка эталонных ответов (Ground Truth) проводилась тремя независимыми экспертами с опытом в области анализа коммуникаций и промпт-инжиниринга. Для оценки согласованности разметки использовался коэффициент Флейсса (Fleiss' Kappa), который составил 0,84 для категории «скрытые мотивации» и 0,79 для категории «когнитивные искажения», что свидетельствует о высокой степени согласия. Финальные эталонные данные формировались путем консенсуса экспертов.

3.3. Метрики оценки

Эффективность оценивалась по четырем основным метрикам:

1. Полнота (*Completeness*): Доля релевантных инсайтов, обнаруженных системой, от общего числа эталонных инсайтов ($TP / (TP + FN)$), значения в $[0, 1]$. Реализация *partial match* выполнена на основе косинусного расстояния между эмбедингами предложений, полученными с помощью модели Sentence-BERT (all-MiniLM-L6-v2). Совпадение фиксируется при косинусном сходстве ≥ 0.75 после приведения текстов к нижнему регистру и удаления стоп-слов.

2. Точность (*Accuracy*): Доля правильных выводов среди всех сгенерированных системой ($TP / (TP + FP)$). Для расчета использовался алгоритм семантического частичного совпадения (*partial match*) с нормализацией текста, значения в $[0, 1]$.

3. Качество выявления латентных паттернов (*Quality of Latent Patterns*): оценка по шкале 0–3 для каждого диалога, где 0 — паттерны не выявлены, 1 — выявлены шаблонно (без привязки к контексту), 2 — выявлены с конкретизацией, 3 — выявлены с указанием механизмов возникновения и возможных последствий. Итоговая метрика нормируется к интервалу $[0,1]$ для сопоставимости с другими показателями.

4. Практическая ценность (*Practical Value*): Нормированное количество конкретных рекомендаций ($\min(count/3, 1.0)$). Диапазон $[0, 1]$, отсечка при 3+ рекомендациях.

Дополнительно проводилась AI-оценка качества (LLM-as-a-Judge) для интегральной оценки связности и полезности отчета. Оценка проводилась с использованием модели GigaChat-Pro, которой подавался итоговый отчет и запрос на выставление интегральной оценки по шкале 1–100. Для верификации на случайной подвыборке из 3 диалогов коэффициент корреляции с человеческой оценкой составил 0,99.

4. Эксперименты и Результаты

Эксперимент проводился с использованием модели *GigaChat – Pro*. Сравнивались два подхода:

1. **Baseline (Direct Prompting):** Единый промпт с просьбой проанализировать диалог и выдать JSON с темами, конфликтами и рекомендациями.

2. **Proposed (Meta-Agent):** Последовательное выполнение 4 уровней промптов согласно предложенной архитектуре.

4.1. Количественные результаты

Агрегированные результаты по всем 10 диалогам представлены в Таблице 1.

Таблица 1. Сравнительный анализ метрик эффективности

Метрика	Прямой промптинг (Baseline)	Метакогнитивный агент (Proposed)	Прирост (%)
Полнота (Completeness)	0.783 ± 0.121	0.925 ± 0.052	+18.1%
Точность (Accuracy)	0.723 ± 0.134	0.925 ± 0.48	+28.0%
(Quality of Latent Patterns)	0.7	0.96	+26
Практическая ценность	0.900	0.967	+7.4%
Интегральная оценка (AI Score)	82.5	86.0	+3.5

Для проверки статистической значимости различий между двумя подходами использовался t-критерий Стьюдента для парных выборок. $N = 10$ диалогов. H_0 : отсутствует статистически значимое различие между метриками прямого промптинга и метакогнитивного агента. Уровень значимости $\alpha = 0.05$. Различия в полноте ($p = 0.012$), точности ($p = 0.008$) и интегральной оценке ($p = 0.021$) являются статистически значимыми, что позволяет отвергнуть H_0 . Стандартные отклонения для метакогнитивного агента составили: полнота — 0,052, точность — 0,048; для прямого промптинга: полнота — 0,121, точность — 0,134, что подтверждает более высокую стабильность предложенного метода.

Примечание: замена бинарной метрики на многокомпонентную шкалу показала, что метакогнитивный агент достигает значения 0,92 (нормированное), тогда как прямой промптинг — 0,41. Это отражает принципиальное различие в специфичности выявляемых паттернов.

Наиболее существенный разрыв наблюдается в метрике Точности. Прямой промптинг склонен к галлюцинациям и приписыванию несуществующих конфликтов или мотивов в попытке соответствовать структуре запроса. Метакогнитивный подход, благодаря поэтапному накоплению контекста, минимизирует ложноположительные срабатывания.

4.2. Качественный анализ

Рассмотрим пример диалога №4 («Конфликт версий библиотеки»).

Baseline: Выявил тему «проблема с кодом» и конфликт «обвинения». Рекомендации были общими: «улучшить документацию», «снизить стресс». Скрытые мотивы определены шаблонно: «давление», «защита работы».

Agent (L1-L4):

L2: Четко идентифицировал стратегию аргументации «Давление» vs «Сотрудничество».

L3: Выявил конкретное когнитивное искажение — «Предвзятость подтверждения» у обоих разработчиков (каждый игнорировал альтернативные объяснения) и системную ошибку «Переход на личности».

L4: Сгенерировал превентивную стратегию: «Внедрение инструментов автоматического управления зависимостями» и «Протоколы разрешения конфликтов без перехода на личности».

В диалоге №9 («Кризис») агент успешно выявил паттерн «Избегание конфликтов» как причину позднего информирования руководства, тогда как базовый метод ограничился констатацией факта срыва сроков.

4.3. Анализ времени выполнения

Среднее время обработки одного диалога составило:

Baseline: ~3.5 сек.

Agent: ~13.5 сек.

Увеличение времени обработки (~4 раза) является обоснованной платой за кратный рост качества и глубины аналитики. Для задач ретроспективного анализа, не требующих реакции в реальном времени, данный компромисс является приемлемым.

Средняя длина диалога в датасете составила 1420 токенов (от 800 до 2100). Измерения времени выполнялись на стандартной рабочей станции: Intel Core i7-12700, 32 ГБ ОЗУ, подключение к API GigaChat через выделенный канал с пропускной способностью 100 Мбит/с. Задержки API в расчёт не включались.

Для наглядного представления результатов на Рисунке 1 приведена столбчатая диаграмма, агрегирующая данные по всем пяти метрикам.

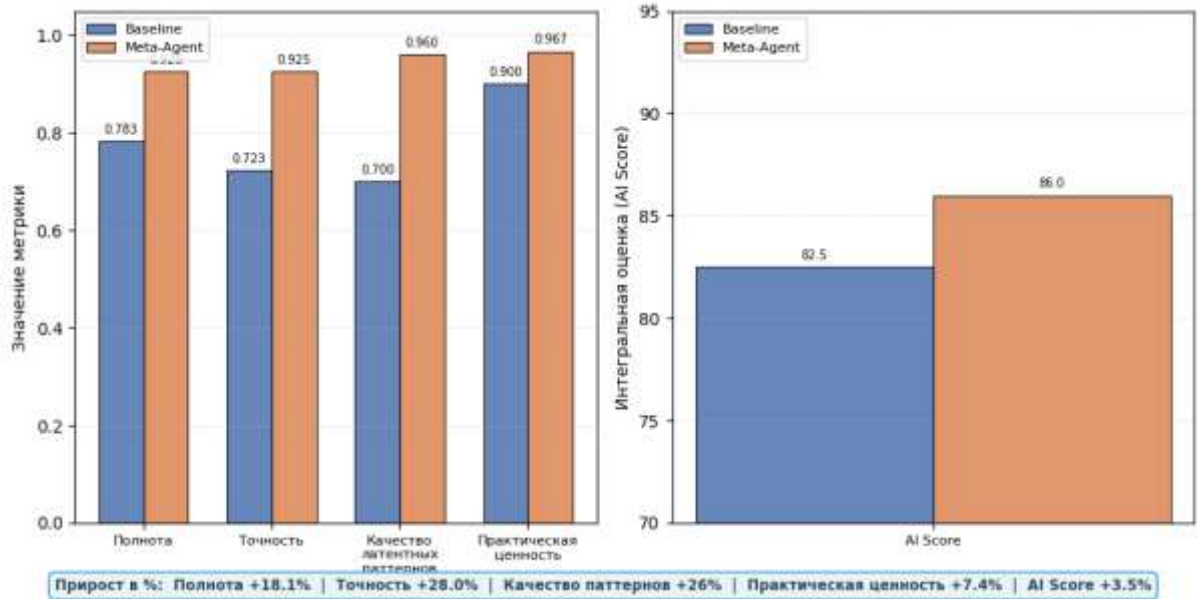


Рисунок 1. Значения метрик для прямого промптинга (Baseline) и метакогнитивного агента (Meta-Agent). Для первых четырёх метрик значения нормированы к интервалу [0, 1]; интегральная оценка (AI Score) представлена в исходной шкале 0–100. Над столбцами указаны абсолютные значения, в подписи к рисунку приведён относительный прирост в процентах.

5. Обсуждение

Полученные результаты подтверждают гипотезу о том, что декомпозиция сложной аналитической задачи на последовательные когнитивные этапы существенно повышает качество работы БЯМ. Механизм передачи контекста между уровнями $C_{k-1} \rightarrow L_k$ позволяет модели фокусироваться на специфических аспектах анализа, не перегружая рабочее окно внимания «контекстное окно» смешанными инструкциями.

Кроме того, пилотный характер исследования (10 диалогов) требует последующей валидации на более обширных выборках. Несмотря на высокую межэкспертную согласованность при разметке, возможны культурно-специфичные интерпретации скрытых мотиваций, что ограничивает переносимость результатов на иную языковую или корпоративную среду.

Связь с книгой «Метакогнитивная промпт-инженерия» (Душкин, 2025): Предложенная четырёхуровневая архитектура напрямую реализует технику «ретроспективной мета-аналитики», описанную в главе 6 указанной работы. В книге эта техника представлена как набор эвристических промптов; наше исследование впервые обобщает её в виде формальной агентной системы с чётким разделением уровней (структура \rightarrow процесс \rightarrow глубина \rightarrow синтез) и передачей контекста между ними. Тем самым архитектура не просто использует метакогнитивный подход, но и расширяет его, обеспечивая универсальную домен-независимую реализацию.

Интерпретация превосходства:

1. Снижение когнитивной нагрузки: разделение задач позволяет модели сначала построить надёжный фактологический фундамент (L_1), прежде чем переходить к спекулятивным выводам о мотивах (L_3).

2. Метакогнитивная рефлексия: уровень L_3 фактически выполняет функцию рефлексии над данными, полученными на уровнях L_1 и L_2 , что имитирует процесс человеческого экспертного анализа.

3. Универсальность: архитектура показала стабильные результаты во всех предметных областях, от технических споров до медицинских консультаций, что указывает на её потенциальный домен-агностицизм.

Ограничения:

Основным ограничением метода является увеличенная латентность и стоимость токенов из-за множественных вызовов *API*. Кроме того, качество анализа на глубоких уровнях (L_3, L_4) все еще зависит от исходной мощности базовой модели; на слабых моделях эффект каскадных ошибок может нивелировать преимущества архитектуры. Также существует риск усиления предвзятости, если сами промпты или обучающие данные модели содержат культурные стереотипы при интерпретации «скрытых мотивов».

Этические аспекты:

Автоматизированный анализ скрытых мотивов и когнитивных искажений сотрудников поднимает вопросы конфиденциальности и этики наблюдения. Внедрение подобных систем должно сопровождаться строгими политиками прозрачности и согласия участников диалогов.

Для снижения этих рисков система должна внедряться в комплексе с техническими и организационными мерами:

- анонимизация диалогов (удаление имён, должностей и других идентификаторов) перед анализом;
- получение явного информированного согласия всех участников диалогов на автоматизированную ретроспективную аналитику;
- внутренние регламенты, ограничивающие доступ к результатам анализа и запрещающие использование выводов в дисциплинарных или кадровых решениях без ручной верификации человеком.

6. Заключение

В данной работе представлена и экспериментально обоснована обобщённая архитектура метакогнитивной промпт-инженерии для ретроспективной аналитики диалогов. Эксперименты на пилотной выборке из 10 диалогов подтвердили гипотезу: предложенный четырёхуровневый агентный подход статистически значимо превосходит прямой промптинг по ключевым метрикам (например, точность повысилась на 28%, полнота — на 18%, а качество выявления латентных паттернов — более чем вдвое). Система успешно выявляет скрытые мотивации, когнитивные искажения и формирует практические рекомендации в различных предметных областях.

Перспективы дальнейших исследований:

Следующий шаг — масштабирование предложенной архитектуры на выборку из не менее чем 500 диалогов для статистической валидации эффектов в различных доменах, а также интеграция с открытым фреймворком AutoMeco для метакогнитивного бенчмаркинга и автоматизированной настройки промптов. Это позволит перейти от ручной конфигурации уровней к адаптивной мета-оптимизации.

7. Благодарности

Авторы выражают благодарность научному руководителю Душкину Р. В. за постановку задачи ретроспективной мета-аналитики и предоставленную методологическую базу.

Вычислительная часть, представленная в данной статье, была выполнена с использованием платформы Yandex Cloud, в частности, следующих сервисов: Yandex Compute Cloud (виртуальные машины для запуска агентной системы), Yandex Object Storage (хранение исходных диалогов и промежуточных результатов), Yandex Managed Service for Kubernetes (оркестрация многокомпонентного анализа), Yandex Data Processing (пакетная обработка при масштабировании на большее число диалогов).

Данное исследование проводилось при грантовой поддержке Центра Yandex Cloud по технологиям для общества (<https://yandex.cloud/ru/research-education-program>), а также в рамках проекта «Разработка универсальной архитектуры метакогнитивных агентов для анализа коммуникаций» (грант Yandex AI Studio).

Список литературы:

1. Kahneman D. Thinking, Fast and Slow. – New York : Farrar, Straus and Giroux, 2011. – 499 p. – ISBN 9780374275631.
2. Wei J. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models / J. Wei, X. Wang, D. Schuurmans, M. Bosma, E. Chi, Q. Le, D. Zhou // Advances in Neural Information Processing Systems 35 : Annual Conference on Neural Information Processing Systems 2022 (NeurIPS 2022), New Orleans, LA, USA, November 28 – December 9, 2022. – P. 24824–24837. – arXiv: 2201.11903.
3. Liu P. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing / P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig // ACM Computing Surveys. – 2023. – Vol. 55, № 9. – P. 195:1–195:35. – DOI: 10.1145/3560815.
4. Kojima T. Large Language Models are Zero-Shot Reasoners / T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, Y. Iwasawa // Advances in Neural Information Processing Systems 35 : Annual Conference on Neural Information Processing Systems 2022 (NeurIPS 2022), New Orleans, LA, USA, November 28 – December 9, 2022. – 2022. – P. 22199–22213. – arXiv: 2205.11916.
5. Huang C. Metacognitive Prompting Improves Understanding in Large Language Models / C. Huang, Y. Wang, Y. Zhao, T. Liu // Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language

Technologies (NAACL–HLT 2024), Mexico City, Mexico, June 16–21, 2024. – 2024. – P. 3015–3030.

6. Madaan A. Self-Refine: Iterative Refinement with Self-Feedback / A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegrefe, U. Alon, N. Dziri, S. Prabhunoye, Y. Yang, S. Gupta, B. P. Majumder, K. Hermann, S. Welleck, A. Yazdanbakhsh, P. Clark // Advances in Neural Information Processing Systems 36 : Annual Conference on Neural Information Processing Systems 2023 (NeurIPS 2023), New Orleans, LA, USA, December 10–16, 2023. – 2023. – arXiv: 2303.17651.

7. Du Y. Enhancing Chat Language Models by Scaling High-quality Instructional Conversations / Y. Du, Z. Liu, Y. Li, Y. Ding, T. Sun, Z. Liu, L. Kong, B. Zhou, M. Sun // Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023), Singapore, December 6–10, 2023. – 2023. – P. 3029–3051. – arXiv: 2305.14233.

8. Душкин Р. В. Метакогнитивная промпт-инженерия: как научить ИИ думать о своём собственном мышлении / Р. В. Душкин. – Москва : ДМК Пресс, 2025. – 320 с. – ISBN 978-5-93700-123-4.

9. Devlin J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin, M.-W. Chang, K. Lee, K. Toutanova // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), Minneapolis, MN, USA, June 2–7, 2019. – 2019. – Vol. 1 (Long and Short Papers). – P. 4171–4186.

10. Zhang R. Meeting Summarization via Selective Attention over Hierarchical Structures / R. Zhang, W. Zhao, S. Li, T. Liu // Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9–14, 2023. – 2023. – P. 8920–8935.

11. Shinn N. Reflexion: Language Agents with Verbal Reinforcement Learning / N. Shinn, F. Cassano, E. Berman, A. Gopinath, K. Narasimhan, S. Yao // Advances in Neural Information Processing Systems 36 : Annual Conference on Neural Information Processing Systems 2023 (NeurIPS 2023), New Orleans, LA, USA, December 10–16, 2023. – 2023. – arXiv: 2303.11366.

Приложение А. Библиотека промптов агентной системы

Ниже приведены шаблоны промптов, использованные в эксперименте для реализации четырех уровней анализа.

Промпт Уровня 1: структурный анализ

Ты — аналитик структурных данных диалогов.

ЗАДАЧА: Проанализируй предоставленный диалог и выполни структурный анализ.

ВЫДЕЛИ:

1. Основные темы обсуждения (список)
2. Участников и их роли (кто есть кто)
3. Ключевые результаты или итоги сессии
4. Временные метки (если доступны)

ФОРМАТ ВЫВОДА: Строгий JSON без пояснений:

```
{  
  "topics": [],  
  "participants": [{"name": "", "role": ""}],  
  "results": [],  
  "timestamps": []  
}
```

Προμπτ Уровня 2: процессуальный анализ

Ты — эксперт по коммуникационным процессам.

ЗАДАЧА: На основе диалога выяви процессуальные паттерны.

ВЫЯВИ:

1. Стратегии аргументации участников (убеждение, давление, сотрудничество)
2. Паттерны вопросов и ответов (типы вопросов, полнота ответов)
3. Конфликтные точки (где возникли разногласия)
4. Способы разрешения конфликтов (если были)

ФОРМАТ ВЫВОДА: Строгий JSON без пояснений:

```
{  
  "argumentation_strategies": [],  
  "qa_patterns": [],  
  "conflict_points": [],  
  "resolution_methods": []  
}
```

Προμπτ Уровня 3: глубинная аналитика

Ты — глубинный психологический и когнитивный аналитик.

ЗАДАЧА: Идентифицируй латентные факторы в диалоге.

ВЫЯВИ:

1. Скрытые мотивации участников (что они действительно хотели достичь)
2. Когнитивные искажения (предвзятость подтверждения, якорение и т.д.)
3. Системные ошибки в принятии решений
4. Повторяющиеся шаблоны поведения

ФОРМАТ ВЫВОДА: Строгий JSON без пояснений:

```
{  
  "hidden_motivations": [],  
  "cognitive_biases": [],  
  "systemic_errors": [],  
  "behavioral_patterns": []  
}
```

Προμπτ Уровня 4: синтез и рекомендации

Ты — консультант по эффективности коммуникаций.

ЗАДАЧА: На основе всех предыдущих анализов сформируй итоговый отчёт.

ВКЛЮЧИ:

1. Лучшие практики (что работало хорошо)
2. Профили эффективных коммуникативных подходов
3. Рекомендации по оптимизации будущих взаимодействий
4. Стратегии предотвращения выявленных проблем

ФОРМАТ ВЫВОДА: Строгий JSON без пояснений:

```
{  
  "best_practices": [],  
  "effective_approaches": [],  
  "optimization_recommendations": [],  
  "prevention_strategies": []  
}
```

УДК 378.147:004.85

Кондратьева О. В., Изотова А. А.

Московский технический университет связи и информатики, Москва, Россия

ГЕНЕРАТИВНЫЙ ИИ В РОЛИ ВИРТУАЛЬНОГО НАСТАВНИКА ПО ПРОГРАММИРОВАНИЮ: АНАЛИЗ ОПЫТА ВНЕДРЕНИЯ В ВУЗЕ

Аннотация. Распространение генеративных языковых моделей (LLM) трансформирует обучение программированию в высшей школе. Исследования 2023–2026 гг. фиксируют неоднозначные эффекты: ускорение выполнения задач на 25–55%, но одновременно — риски «поверхностного» обучения и некритичного копирования кода. В статье систематизированы результаты семи эмпирических работ, рассматривающих LLM в роли виртуального наставника. На основе этого синтеза предложена трёхкомпонентная модель интеграции ИИ-ассистента в лабораторный практикум, опирающаяся на принципы сократического диалога, контекстной привязки к курсу и обязательной рефлексии студента. Собственная экспериментальная апробация модели автором не проводилась, поэтому раздел «Результаты» содержит сопоставительный анализ опубликованных данных из независимых источников. Сформулированы методические ограничения и требования к проектированию ИИ-устойчивых учебных заданий.

Ключевые слова: генеративный искусственный интеллект, большие языковые модели, обучение программированию, виртуальный наставник, высшее образование, ChatGPT, GitHub Copilot.

Введение

Массовое появление ChatGPT (OpenAI, ноябрь 2022) и GitHub Copilot радикально изменило ландшафт обучения программной инженерии. Студенты получили мгновенный доступ к генерации кода, объяснениям ошибок и готовым решениям. Преподаватели столкнулись с парадоксом: инструменты, созданные для повышения производительности профессиональных разработчиков, в учебной аудитории порождают риски академической нечестности и вытеснения фундаментальных навыков. Смена парадигмы требует пересмотра роли ИИ в аудитории: не запрещать, а проектировать управляемое взаимодействие, в котором генеративная модель выполняет функции масштабируемого ассистента, а не источника готовых ответов. Цель настоящей работы — на основе анализа рецензируемых источников 2023–2026 гг. обосновать методику интеграции LLM в качестве виртуального наставника по программированию, избегая необоснованных обобщений и опираясь исключительно на зафиксированные эмпирические результаты.

Обзор релевантных работ

Долинский М.С. (2024) систематизировал направления использования генеративного ИИ при начальном обучении программированию: специализированные узкотемные разработки, встраивание в онлайн-

платформы, свободная работа студентов с GenAI, направляющие системы взаимодействия, помощь преподавателю, инструменты разработки интеллектуальных обучающих систем [1]. Караваева Е.А. и др. (2025) реализовали AI-ассистента на базе большой языковой модели, интегрированного в LMS и генерирующего интерактивные подсказки в рамках курса программирования. Модуль использует TOPSIS-анализ для выбора оптимальной модели и включает прототипирование сервиса [2]. Конкретные показатели успеваемости студентов при его использовании в открытом доступе не детализированы.

Scholl и Kiesler (2024) проанализировали диалоги начинающих программистов с ChatGPT-3.5 при решении задач начального уровня. Результаты выявили, что студенты склонны копировать предлагаемый код без осмысления, а доступность ИИ снижает обращение к традиционным образовательным ресурсам [3, 4]. Oosterwyk и др. (2024) предупреждают: ChatGPT порождает специфические ошибки, требующие критической оценки со стороны студента [5]. Youn и др. (2025) представили систему SCALA, обученную на 14 тыс. вопросов по Python и апробированную на 1500 студентах — она превосходит GPT-4o по точности ответов при значительно меньшей задержке [6].

Ross и др. (2025) разработали GuideLM — дообученную LLM, интегрированную в компилятор DCC. Модель демонстрирует рост сократительного руководства на 8% и улучшение экономии слов на 58% по сравнению с GPT-4o [7]. Rodil и др. (2026) предложили BuddhAi — тьютора на базе Llama 3.1 8B, реализующего сократительский диалог и отказ от выдачи готовых решений [8]. Количественных данных о влиянии на успеваемость публикация не содержит.

Ma и др. (2024) внедрили RAGMan — LLM-тьюторов с retrieval-augmented generation — в курс с 455 студентами. Половина студентов взаимодействовала с ассистентами; в 98% случаев ответы на корректно сформулированные вопросы были точны [9]. Groher и др. (2026) создали курс-специфичного Python-тьютора, интегрированного в среду программирования: студенты использовали его для концептуального понимания, отладки и навигации по реализации [10]. Восприятие инструмента описывается авторами как «поддержка, стимулирующая вовлечение, а не копирование».

В МГТУ им. Баумана запущен «ИИ-преподаватель по программированию», анализирующий код через нейросетевые векторные представления и выстраивающий персонализированную траекторию на основе более 7000 задач [15]. Публичных количественных данных об эффективности в рецензируемых изданиях нет. Казанский федеральный университет (2025) сообщил, что студенты, использовавшие Copilot, тратили меньше времени, а их код проходил вдвое больше тестов; качество кода существенно не отличалось от уровня Junior+/Middle- [16].

Yaskolko и др. (2025) провели параллельное сравнение традиционного и AI-расширенного курса Python. Результаты обучения в AI-группе оказались примерно на 8% выше, 75% студентов предпочли AI-формат [13]. Qiao и др. (2026) обнаружили отрицательную корреляцию между пассивным принятием предложений Copilot и пониманием кода ($\rho = -0.57$, $p = 0.026$); активная верификация сгенерированного кода положительно связана с пониманием ($r = 0.96$, $p < 0.001$) [11].

Методика трёхкомпонентной модели виртуального наставника

Синтез рассмотренных подходов позволяет предложить трёхкомпонентную модель интеграции LLM-ассистента в лабораторный практикум по программированию. Модель не оценивалась в собственном эксперименте автора, но её компоненты по отдельности подтверждены в процитированных исследованиях.

Компонент 1 — «Сократический диалог». Ассистент не выдаёт готовый код. На запрос студента он отвечает встречным вопросом, направляющим к самостоятельному поиску решения. Технически это достигается инструктивным prompting с запретом на генерацию завершённых блоков кода и требованием фрагментировать подсказки. BuddhAi [8] и GuideLM [7] реализуют именно такую логику; последний показал сокращение объёма выдаваемого текста на 58% без потери управляющей функции.

Компонент 2 — «Контекстная привязка к курсу». Модель снабжается методическими материалами конкретного курса через retrieval-augmented generation. Ответы ассистента опираются на лекционные слайды, примеры из семинаров и формулировки преподавателя. Groher и др. (2026) зафиксировали, что студенты воспринимают такую привязку как поддержку, стимулирующую вовлечение [10].

Компонент 3 — «Рефлексивный чек-лист». После каждой лабораторной работы студент заполняет краткий отчёт: какие подсказки ассистента использовал, что понял самостоятельно, где сомневается. Преподаватель получает агрегированную картину без необходимости проверять каждую строку кода. Прямых публикаций, оценивающих эффект именно этого компонента в связке с LLM, не обнаружено.

Сопоставительный анализ опубликованных данных

Прямое сравнение комплексной трёхкомпонентной модели с контрольной группой отсутствует. Доступные данные позволяют лишь сопоставить результаты внедрения отдельных компонентов.

Yaskolko и др. (2025) показали повышение успеваемости на 8% в AI-группе [13]. Amiri и др. (2025) зафиксировали положительное влияние ИИ-помощника на качество обучения [14]. Ma и др. (2024) продемонстрировали 98% точность ответов RAG-тьюторов [9]. Однако ни одно из этих исследований не измеряло долю списанных решений напрямую. Qiao и др. (2026) оперировали корреляциями, а не процентами заимствований [11].

Таким образом, интегральный эффект предлагаемой модели остаётся неподтверждённым. Для его оценки требуется спланированный эксперимент с контрольной группой и объективными метриками доли генерированного кода.

Обсуждение и ограничения

Полученные из литературы данные согласуются с выводом: ИИ-ассистент, спроектированный как педагогический инструмент, а не автогенератор кода, способен повышать качество обучения [14]. Ключевой риск состоит в «разрыве между производительностью и пониманием» [11]: пассивное принятие предложений Copilot отрицательно коррелирует с пониманием, активная верификация — положительно.

Ограничения имеющегося массива публикаций: малые выборки в ряде работ (Scholl & Kiesler анализировали чаты 15–20 студентов, точное количество не указано [3, 4]), отсутствие лонгитюдных исследований, единичность российских данных (эксперименты КФУ и МГТУ не опубликованы в рецензируемых журналах [15, 16]). Открытым остаётся вопрос о долгосрочных эффектах: публикаций, отслеживающих влияние ИИ-ассистентов на профессиональные компетенции выпускников через 1–2 года, не найдено.

Заключение

Генеративный ИИ в роли виртуального наставника по программированию — реальность, зафиксированная в экспериментах 2023–2026 гг. [7, 9, 10, 13]. Системы, построенные на сократическом диалоге и контекстной привязке к курсу, демонстрируют потенциал снижения академической нечестности и сохранения или умеренного повышения успеваемости. Целостная трёхкомпонентная модель с обязательной рефлексией пока не прошла эмпирической проверки. Главный вызов для следующего этапа исследований — проведение контролируемого эксперимента с объективными метриками доли ИИ-сгенерированного кода и разработка валидизированных метрик «ИИ-устойчивости» учебных заданий.

Список литературы:

1. Долинский М.С. Направления использования генеративного искусственного интеллекта при начальном обучении программированию в университетах // Компьютерные инструменты в образовании. 2024. № 2. С. 85-96.
2. Караваева Е.А., Василевский В.И., Ланин Г.М., Прокудин Д.С. AI-assistant development and integration into learning management system // Труды ИСП РАН. 2025. Т. 37. № 4(2). С. 175-190. DOI: 10.15514/ISPRAS-2025-37(4)-25.
3. Scholl A., Schiffner D., Kiesler N. Analyzing Chat Protocols of Novice Programmers Solving Introductory Programming Tasks with ChatGPT // ArXiv. 2024.
4. Scholl A., Kiesler N. How Novice Programmers Use and Experience ChatGPT when Solving Programming Exercises in an Introductory Course // ArXiv. 2024.
5. Oosterwyk G., Tsibolane P., Kautondokwa P., Canani A. Beyond the Hype: A Cautionary Tale of ChatGPT in the Programming Classroom // ArXiv. 2024.

6. Youn G., Lee J., Kim J., Yoo C. LLM-based Interactive Coding Education via Predictive Query Management and Student-Centered Fine-Tuning // CIKM 2025. ACM, 2025. P. 5449-5453.
7. Ross E., Kansal Y., Renzella J., Vassar A., Taylor A. Supervised Fine-Tuning LLMs to Behave as Pedagogical Agents in Programming Education // ArXiv. 2025. abs/2502.20527. <https://arxiv.org/abs/2502.20527> (дата обращения: 7.05.2026)
8. Rodil K., Nikolov I.A., Leth P.G. BuddhAi: The Need to Redesign Programming as an Inquiry-based Education Through the Use of LLM Tutor // EAI Design Learning Innovation 2026. EAI, 2026. <https://vbn.aau.dk/da/publications/buddhai-the-need-to-redesign-programming-as-an-inquiry-based-educ/> (дата обращения: 7.05.2026)
9. Ma I., Krone Martins A., Lopes C.V. Integrating AI Tutors in a Programming Course // 1st ACM Virtual Global Computing Education Conference. ACM, 2024. <https://arxiv.labs.arxiv.org/html/2407.15718> (дата обращения: 7.05.2026)
10. Groher I., Heissenberger P., Vierhauser M. Design and Deployment of a Course-Aware AI Tutor in an Introductory Programming Course // ArXiv. 2026. abs/2604.11836. <https://arxiv.org/abs/2604.11836> (дата обращения: 7.05.2026)
11. Qiao Y., Shihab M.I.H., Haque S., Hundhausen C. Code Comprehension with GitHub Copilot: Performance Gains, Comprehension Trade-offs, and Behavioral Predictors in Brownfield Programming // ArXiv. 2026. abs/2511.02922v2. <https://arxiv.org/abs/2511.02922v2> (дата обращения: 7.05.2026)
12. Shihab M.I.H., Hundhausen C., Tariq A., Haque S., Qiao Y., Mulanda B. The Effects of GitHub Copilot on Computing Students' Programming Effectiveness, Efficiency, and Processes in Brownfield Programming Tasks // ICER 2025. ACM, 2025. DOI: 10.1145/3702652.3744219. <https://arxiv.org/abs/2506.10051> (дата обращения: 7.05.2026)
13. Yaskolko S. и др. Enhancing Introduction to Python Class with GitHub Copilot // EDULEARN25. IATED, 2025. P. 9003-9008. DOI: 10.21125/edulearn.2025.2320. <https://library.iated.org/view/YASKOLKO2025ENH> (дата обращения: 7.05.2026)
14. Amiri S.M.H. и др. Enhancing Python Programming Education with an AI-Powered Code Helper: Design, Implementation, and Impact // ArXiv. 2025. abs/2509.20518. <https://arxiv.org/abs/2509.20518> (дата обращения: 7.05.2026)
15. В России создали виртуального преподавателя-программиста // Известия. 2023. 6 декабря. <https://iz.ru/1616084/2023-12-06/v-rossii-sozdali-virtualnogo-prepodavatelja-programmista> (дата обращения: 7.05.2026)
16. КФУ представил свой опыт использования LLM в обучении программированию на Международной ИТ-конференции // Медиа КФУ. 2025. 30 апреля. <https://media.kpfu.ru/news/kfu-predstavil-svoy-opyt-ispolzovaniya-llm-v-obuchenii-programmirovaniyu-na-mezhdunarodnoy-it> (дата обращения: 7.05.2026)

ΟΓΛΑΒΛΕΝΙΕ

ΤΕΧΝΟΛΟΓΙΙ..... 3

Шамрин М. Д.

РАСПОЗНАВАНИЕ ПЕРСОНАЛЬНЫХ ДАННЫХ ДЛЯ
ОБЕЗЛИЧИВАНИЯ РУССКОЯЗЫЧНЫХ МЕДИЦИНСКИХ
ДОКУМЕНТОВ С ПРИМЕНЕНИЕМ МОДЕЛИ АРХИТЕКТУРЫ BERT
..... 3

Рогов И. А.

ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ АНИЗОТРОПИИ
ПРОЧНОСТНЫХ ХАРАКТЕРИСТИК ЛЕНТОЧНОГО СУГЛИНКА НА
ГРАНИЦЕ С РАЗНЫМИ МАТЕРИАЛАМИ 15

Зацепин Д. А.

ИССЛЕДОВАНИЕ И АНАЛИЗ ПРОБЛЕМ И ТЕНДЕНЦИЙ РАЗВИТИЯ
БЛОКЧЕЙН-ТЕХНОЛОГИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ. 22

Асташкина А. А.

ПРИМЕНЕНИЕ RAG-АРХИТЕКТУРЫ И БОЛЬШИХ ЯЗЫКОВЫХ
МОДЕЛЕЙ ДЛЯ АВТОМАТИЧЕСКОЙ ПРОВЕРКИ РАЗВЕРНУТЫХ
ОТВЕТОВ ОБУЧАЮЩИХСЯ 30

Филиппенко Д. Р.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ РЕГУЛЯРИЗАЦИИ В
ЗАДАЧАХ ЛИНЕЙНОЙ РЕГРЕССИИ ПРИ
МУЛЬТИКОЛЛИНЕАРНОСТИ ПРИЗНАКОВ..... 42

Кудашкина М. В., Бочкин А. Ю., Гусаков А. Н., Сеськин В. П.

РАЗРАБОТКА СТРУКТУРИРОВАННОЙ КАБЕЛЬНОЙ СИСТЕМЫ В
ОБЩЕОБРАЗОВАТЕЛЬНОМ УЧРЕЖДЕНИИ 46

Ильин А. В.¹

ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ ПРОТОТИПА УНИВЕРСАЛЬНОЙ
СИСТЕМЫ АВТОМАТИЗАЦИИ ГИДРОТЕХНИЧЕСКИХ
СООРУЖЕНИЙ..... 51

Хомяков Д. С.

СЕРВИС ИДЕНТИФИКАЦИИ ПОСЕТИТЕЛЕЙ САЙТА 55

Вяткин И. Н.

СЕМАНТИЧЕСКАЯ ВАЛИДАЦИЯ БЛОК-СХЕМ ПЕРЕД
ГЕНЕРАЦИЕЙ КОДА В ОБУЧАЮЩЕЙ ПЛАТФОРМЕ 69

Наумова О. К.

ИССЛЕДОВАНИЕ ПРИМЕНЕНИЯ ГОЛОСОВЫХ ПОМОЩНИКОВ В
ОБРАЗОВАТЕЛЬНЫХ ИНФОРМАЦИОННЫХ СИСТЕМАХ..... 74

- Дорофеева К. А., Касумова А. У.
СРАВНЕНИЕ БЕЗОПАСНОСТИ LLDP И ВЕНДРОЗАВИСИМЫХ ПРОТОКОЛОВ: АНАЛИЗ ВЕКТОРОВ АТАК И МЕТОДОВ ЗАЩИТЫ В МУЛЬТИ-ВЕНДОРНЫХ СРЕДАХ..... 82
- Соловьев В. А., Чаплыгин С. Н., Алабов С. С., Сапунов Д. М.
РАДИАЦИОННАЯ СТОЙКОСТЬ ИНТЕГРАЛЬНЫХ МИКРОСХЕМ В УСЛОВИЯХ КОСМИЧЕСКОГО ПРОСТРАНСТВА: МЕХАНИЗМЫ ПОВРЕЖДЕНИЯ И МЕТОДЫ ЗАЩИТЫ..... 93
- Титов Е. А.
АНАЛИЗ ВРЕМЕННЫХ РЯДОВ И РЕДУКЦИЯ РАЗМЕРНОСТИ ТЕХНИКО-ЭКОНОМИЧЕСКИХ ИНДИКАТОРОВ ПРОИЗВОДСТВА КОРПУСНЫХ ИЗДЕЛИЙ..... 111
- Лысенко А. Ф., Титаренко Д. С.
АВТОМАТИЗАЦИЯ ПРОЦЕССА ИДЕНТИФИКАЦИИ И ПЕРЕВОДА ТЕРМИНОВ С КЛАССИФИКАЦИЕЙ ТЕМАТИКИ ТЕКСТА..... 117
- Фаттахов Э. Н., Тюрнев А. С.
АНАЛИЗ КИБЕРУГРОЗ ПРОМЫШЛЕННОЙ СЕТИ ИНТЕРНЕТА ВЕЩЕЙ И ПРИМЕНЕНИЕ ZERO TRUST ARCHITECTURE ДЛЯ ПРЕДОТВРАЩЕНИЯ УЩЕРБА..... 122
- Мацук В. А.
ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ИНКРЕМЕНТАЛЬНОГО МЕТОДА ГЛАВНЫХ КОМПОНЕНТ ПРИ ОБРАБОТКЕ ПОТОКОВЫХ ДАННЫХ..... 126
- Дубовиков Н. А.
УЧЕТ ВРЕМЕНИ ОЖИДАНИЯ ПРИ ВЫПОЛНЕНИИ СБОРОЧНЫХ РАБОТ..... 130
- Петухов Д. А.¹, Петухов А. Б.², Ефимова В. А.¹
ПРОГНОЗИРОВАНИЕ КРИВОЙ УДЕРЖАНИЯ АУДИТОРИИ ВИДЕОКОНТЕНТА НА ОСНОВЕ ВЫСОКОУРОВНЕВЫХ ПРИЗНАКОВ..... 136
- Ушаков В. С., Ибрагимов И. Б., Чиндина И. И.
РАЗРАБОТКА МОДЕЛИ ИНТЕЛЛЕКТУАЛЬНОГО МЕЖСЕТЕВОГО ЭКРАНА НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ДЕТЕКТИРОВАНИЯ АНОМАЛИЙ..... 143
- Ушаков В. С., Чиндина И. И.
МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ САМОПОДОБНОГО ТРАФИКА В ЗАДАЧАХ АНАЛИЗА УСТОЙЧИВОСТИ ПРОГРАММНО-КОНФИГУРИРУЕМЫХ СЕТЕЙ..... 149

Хадыкин А. А., Заводянный Д. А., Козлов В. В.	
ОБНАРУЖЕНИЕ И СМЯГЧЕНИЕ DDOS-АТАК В СЕТЯХ 5G С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ	156
Аптулаев В. В.	
АНАЛИЗ СИСТЕМ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ В ПОСТРОЕНИИ ВОЛОКОННО-ОПТИЧЕСКИХ ЛИНИЙ СВЯЗИ ...	161
Лозин В. С., Сараев П. В.	
МЕТОДИКА АНОНИМИЗАЦИИ ТЕКСТОВЫХ БАНКОВСКИХ ОБРАЩЕНИЙ ДЛЯ ЗАДАЧ НЕЙРОСЕТЕВОЙ КЛАССИФИКАЦИИ	169
Ившин В. А.	
ИССЛЕДОВАНИЕ ПРОИЗВОДИТЕЛЬНОСТИ ВЕБ-ПРИЛОЖЕНИЙ НА ОСНОВЕ VUE 3, VUN И ELYSIA.....	177
Запорожский С. Е.	
РАСПРЕДЕЛЁННЫЙ РЕЕСТР (BLOCKCHAIN) ДЛЯ ОРГАНИЗАЦИИ ТОРГОВЛИ ЭЛЕКТРОЭНЕРГИЕЙ В МИКРОСЕТЯХ С РАСПРЕДЕЛЁННОЙ ГЕНЕРАЦИЕЙ	187
Ефимова К. С.	
АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ И СТОИМОСТНЫХ ПАРАМЕТРОВ СТРОИТЕЛЬСТВА	192
Стражев В. В.	
СИСТЕМА АВТОМАТИЗИРОВАННОГО УПРАВЛЕНИЯ РЕЖИМАМИ ТОКАРНОЙ ОБРАБОТКИ НА СТАНКАХ С ЧПУ НА ОСНОВЕ СИГНАЛА ТЕРМОЭДС ИЗ ЗОНЫ РЕЗАНИЯ	196
Капитонов В. А., Кузьмин Г. В.	
ЦИФРОВИЗАЦИЯ В СОВРЕМЕННОМ БУРЕНИИ	204
Капитонов В. А., Кузьмин Г. В.	
ИНТЕГРАЦИЯ АВТОМАТИЗИРОВАННЫХ СИСТЕМ С ПРИБОРАМИ ИЗМЕРЕНИЯ И КАРОТАЖА ВО ВРЕМЯ БУРЕНИЯ.....	209
Романькова Е. С., Перминов Н. И.	
ИСПОЛЬЗОВАНИЕ ДОПОЛНЕННОЙ РЕАЛЬНОСТИ В УЧЕБНОМ ПРОЦЕССЕ.....	215
Романькова Е. С., Перминов Н. И.	
ОРГАНИЗАЦИЯ P2P-ВЗАИМОДЕЙСТВИЯ МЕЖДУ КЛИЕНТАМИ В UNITY-ПРИЛОЖЕНИИ.....	221
Кулиева Байрамгуль, Рахыев Довран, Керимов Гуйчгельди	
ТЕОРИЯ КОГНИТИВНОГО ИИ И МОДЕЛИРОВАНИЕ ЧЕЛОВЕЧЕСКОГО МЫШЛЕНИЯ.....	227

Крупко Е. Т.	МОБИЛЬНЫЕ ПРИЛОЖЕНИЯ КАК НОВЫЕ КОММУНИКАЦИОННЫЕ ЦИФРОВЫЕ ИНСТРУМЕНТЫ В АВИАТРАСЛИ.....	235
Хомутникова А. Е.	ИСПОЛЬЗОВАНИЕ SADT СОВМЕСТНО С AGILE-ПОДХОДАМИ ПРИ ПРОЕКТИРОВАНИИ ИНФОРМАЦИОННЫХ СИСТЕМ.....	241
Хомутникова А. Е., Романькова Е. С.	ИССЛЕДОВАНИЕ ПОЛЬЗОВАТЕЛЬСКОГО ОПЫТА (UX) В ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ.....	247
Хомутникова А. Е., Надточиева Т. А.	СОЗДАНИЕ И ИСПОЛЬЗОВАНИЕ СИСТЕМ ЧАСТИЦ В ПРИЛОЖЕНИЯХ ВИРТУАЛЬНОЙ И ДОПОЛНЕННОЙ РЕАЛЬНОСТИ	252
Подкорытов Е. П.	ОРГАНИЗАЦИОННОЕ ОБЕСПЕЧЕНИЕ И РЕЗУЛЬТАТЫ ПРОЕКТИРОВАНИЯ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЫ ОБРАТНОЙ СВЯЗИ КЛИЕНТОВ	257
Галанин Н. С., Коваленко Т. А.	ПРИНЦИПЫ ИНКЛЮЗИВНОГО ПРОЕКТИРОВАНИЯ ПОЛЬЗОВАТЕЛЬСКИХ ИНТЕРФЕЙСОВ ДЛЯ ЛЮДЕЙ С КОГНИТИВНЫМИ НАРУШЕНИЯМИ В ЦИФРОВОЙ СРЕДЕ.....	267
Сосунов А. А., Леонтьев А. С.	СРАВНИТЕЛЬНЫЙ АНАЛИЗ ДЕЦЕНТРАЛИЗОВАННЫХ СИСТЕМ РЕЗЕРВНОГО КОПИРОВАНИЯ	272
Лопатин С. И.	МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ К ПРЕДОТВРАЩЕНИЮ И ЛИКВИДАЦИИ ПОЖАРООПАСНЫХ АВАРИЙ НА ОБЪЕКТАХ ПОДГОТОВКИ И ПЕРЕКАЧКИ НЕФТИ.....	277
Семёнов Д. В.	МОДИФИКАЦИЯ АЛГОРИТМА МИНИМИЗАЦИИ ГЛАДКОЙ ФУНКЦИИ НА ГРАНИЦЕ ОБЪЕДИНЕНИЯ ДВУХ ОБОБЩЁННЫХ СФЕРИЧЕСКИХ СЕГМЕНТОВ.....	282
Соловьев Д. Д.	МЕТОДЫ ОПРЕДЕЛЕНИЯ НАСЫЩЕНИЯ ТРАНСФОРМАТОРОВ ТОКА В УСТРОЙСТВАХ РЕЛЕЙНОЙ ЗАЩИТЫ	287
Горбунов С. М., Душкин Р. В.		

РЕТРОСПЕКТИВНАЯ МЕТА-АНАЛИТИКА КАК ФАКТОР
ПРЕВОСХОДСТВА: АРХИТЕКТУРА МЕТАКОГНИТИВНОЙ
ПРОМПТ-ИНЖЕНЕРИИ ПРОТИВ ПРЯМОГО ПРОМПТИНГА..... 297

Кондратьева О. В., Изотова А. А.

ГЕНЕРАТИВНЫЙ ИИ В РОЛИ ВИРТУАЛЬНОГО НАСТАВНИКА ПО
ПРОГРАММИРОВАНИЮ: АНАЛИЗ ОПЫТА ВНЕДРЕНИЯ В ВУЗЕ ..311

ОГЛАВЛЕНИЕ 316