

Цаплин Сергей Тимофеевич

**ГИБРИДНАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ
СОЦИАЛЬНЫХ ПОКАЗАТЕЛЕЙ НА ОСНОВЕ КОРОТКИХ ВРЕМЕННЫХ
РЯДОВ**

2.3.5. Математическое и программное обеспечение вычислительных систем,
комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание учёной степени кандидата технических наук

Работа выполнена на кафедре «Компьютерные системы и сети» Федерального государственного автономного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана).

Научный руководитель: **БЕРЕЗКИН Дмитрий Валерьевич**
кандидат технических наук, доцент кафедры
«Компьютерные системы и сети»

Официальные
оппоненты: **МАКСИМЫЧЕВ Олег Игоревич**
доктор технических наук, доцент, профессор кафедры
«Автоматизированные системы управления»
Федерального государственного бюджетного
образовательного учреждения высшего образования
«Московский автомобильно-дорожный государственный
технический университет (МАДИ)», Институт
транспортных технологий и логистики (ИТТЛ)

МАРТЫШКИН Алексей Иванович
кандидат технических наук, доцент, заведующий
кафедрой «Программирование», декан факультета
автоматизированных информационных технологий
Федерального государственного бюджетного
образовательного учреждения высшего образования
«Пензенский государственный технологический
университет»

Ведущая организация: Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский технологический
университет «МИСИС»»

Защита состоится «17» декабря 2026 года в 13:00 на заседании диссертационного совета 24.2.331.19 в МГТУ им. Н.Э. Баумана по адресу: 105005, г. Москва, ул. 2-я Бауманская, д. 5, стр. 1, зал Учёного совета ГУК МГТУ им. Н.Э. Баумана.

С диссертацией можно ознакомиться в библиотеке МГТУ им. Н.Э. Баумана и на сайте bmstu.ru.

Отзывы на автореферат, заверенные печатью учреждения, просьба направлять по адресу: 105005, г. Москва, ул. 2-я Бауманская, д.5, стр.1, Учёному секретарю диссертационного совета 24.2.331.19, кафедра ИУЗ.

Автореферат разослан «__» _____ 2026 года

Ученый секретарь
диссертационного совета 24.2.331.19
кандидат технических наук, доцент

С.А. Сакулин

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Современные социальные показатели Российской Федерации характеризуется высокой динамичностью. Социальные процессы сопровождаются структурными переключениями, изменением волатильности и нарушением стационарности, что существенно осложняет задачу краткосрочного прогнозирования и принятия решений в социкиберфизических системах. Особенно остро эта проблема проявляется при работе с короткими временными рядами, длина которых часто ограничена 10–30 наблюдениями.

Ограничение длины рядов обусловлено несколькими факторами. Во-первых, для новых административных образований и недавно введённых статистических индикаторов (например, показателей цифровой трансформации социальной сферы) историческая база данных принципиально мала. Во-вторых, при анализе локальных социальных явлений — уровня безработицы в отдельных муниципалитетах, спроса на узкоспециализированные профессии, демографической ситуации в малых регионах - объём доступных наблюдений также оказывается недостаточным для применения классических эконометрических методов, требующих длительных временных рядов. В-третьих, разнообразная частотность социальных данных создаёт дополнительную сложность: ежемесячные показатели рынка труда необходимо согласовывать с поквартальными оценками доходов и ежегодными демографическими отчётами, что требует разработки специализированных методов интерполяции и совместного моделирования.

Традиционные методы прогнозирования - авторегрессионные модели, методы экспоненциального сглаживания, классические регрессионные методы - базируются на допущениях о линейности зависимостей, стационарности процессов и достаточной длине обучающей выборки. В условиях коротких нестационарных рядов эти допущения нарушаются, что приводит к неустойчивости оценок и низкой прогностической способности моделей. Одновременно с этим возникает потребность в методах, способных эффективно извлекать сигнал из ограниченного объёма наблюдений, учитывать возможные структурные сдвиги и обеспечивать допустимую точность прогнозов при высокой волатильности данных.

Решением задачи прогнозирования социальных показателей занимались Кукушкина С.Н., Смирнова Ю.А., Белоусов Д.Р. и Сабельникова Е.М. При этом в качестве основных решений были предложены экспертные методы, что ограничивает скорость принятия решений, объём обрабатываемой информации и накладывает субъективность на итоговый прогноз.

Альтернативой экспертным методам выступают методы, имеющие математическое обоснование. В частности, такими являются методы экстраполяции, моделирования и нейросетевые методы. Развитию и применению этих методов способствовали работы Айвазян С.А., Мхитарян В.С., Лукашин Ю.П. Пономарев Д.С., Vox G., Jenkins G., Chen T., Guestrin C. Однако такие методы либо требуют использования большой обучающей выборки, либо не способны улавливать нелинейные зависимости.

Поскольку решаемая задача заключается в прогнозировании коротких, но в то же время сложных последовательностей, описывающих динамику социальных показателей, необходим метод, который может сочетать в себе преимущества экспертного эстраполяционного и нейросетевого прогнозирования. Наиболее значительный вклад в развитие разработки гибридных моделей прогнозирования социальных показателей внесли Китова О.В., Дьяконова Л.П. и Савинова В.М., создавшие интеллектуальную информационную систему прогнозирования «СГМ Горизонт», которая сочетает в себе экспертные методы и нейронную сеть.

Объектом исследования являются короткие временные ряды различной структуры и характеристик, описывающие социальные процессы и требующие прогнозирования и оценки для принятия управленческих решений в социкиберфизических системах.

Предметом исследования являются методы и алгоритмы организации взаимодействия программ, решающие задачу автоматизированного прогнозирования коротких временных рядов и применимые для социальных показателей.

Целью исследования является повышение точности прогнозов временных рядов на основе разработанных гибридных модели и алгоритма прогнозирования.

Для её достижения поставлены и решены следующие задачи:

- 1) Провести анализ существующих моделей и методов прогнозирования коротких временных рядов, использующихся для прогнозирования социальных показателей;
- 2) Разработать подход для гибридизации метода прогнозирования коротких временных рядов;
- 3) Разработать программное обеспечение интеллектуальной системы прогнозирования коротких временных рядов;
- 4) Провести оценку точности гибридного алгоритма в задаче прогнозирования коротких временных рядов;

Научная новизна диссертационной работы состоит в следующем:

- 1) Разработаны гибридная модель и алгоритм прогнозирования коротких временных рядов, объединяющая статистическую модель, модель машинного обучения и большую языковую модель на основе динамического подбора весовых коэффициентов влияния отдельных моделей (п. 3 паспорта специальности);
- 2) Разработан алгоритм прогнозирования на основе TimeVLM (Временная зрительно-языковая модель), отличающийся меньшим числом активных параметров, что позволяет применять данный алгоритм для коротких временных рядов (п. 4 паспорта специальности);
- 3) Реализован алгоритм автоматической генерации структурированного текстового запроса на основе исторических данных и веб-контекста для настройки большой языковой модели, имитирующей работу эксперта (п. 3 паспорта специальности).

Теоретическая значимость исследования заключается в разработке нового гибридного метода прогнозирования коротких временных рядов (10–30 наблюдений), что критически важно для анализа динамичных социальных показателей в условиях дефицита данных.

Практическая значимость исследования заключается в создании готового цифрового инструмента для аналитиков и государственных служащих, позволяющего получать высокоточные краткосрочные прогнозы социальных процессов при дефиците исторической информации, а также в предложенной архитектуре программного обеспечения, позволяющей реализовывать гибридные модели и алгоритмы прогнозирования коротких временных рядов.

Методы исследования. Для решения задач, поставленных в диссертационной работе, использовались методы машинного обучения, глубокого обучения и программной инженерии.

Положения, выносимые на защиту:

1. Гибридная модель и алгоритм для прогнозирования коротких временных рядов
2. Адаптированный алгоритм TimeVLM для прогнозирования коротких временных рядов
3. Алгоритм генерации структурированного текстового запроса к большой языковой модели

Соответствие паспорту научной специальности. Содержание работы соответствует паспорту научной специальности 2.3.5 Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей (технические науки): п.4 «Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов.», п.3 «Модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем».

Степень достоверности и апробации результатов.

Обоснованность выдвинутых научных положений, теоретических выводов и полученных результатов подтверждается всесторонним анализом современного состояния проблематики в исследуемой области, строгим соответствием теоретических построений данным экспериментальных исследований, а также широкой апробацией ключевых результатов в рецензируемых научных изданиях и на международных научных форумах.

Исследования, отражённые в диссертации, проведены в рамках научно-исследовательской работы «Создание интеллектуальных унифицированных систем поддержки принятия стратегических решений на базе глубокой и прогнозной аналитики больших данных» в рамках проекта BaumanDeerAnalytics.

На предложенные решения получены 3 свидетельства о государственной регистрации программы для ЭВМ: «Гибридное прогнозирование временных рядов», № 2025611275 от 17.01.2025; «Программный модуль TimeVLM, адаптированный для прогнозирования коротких временных рядов» №

2026616858 от 12.03.2026; «Программный модуль LLM-эксперта для прогнозирования коротких временных рядов» № 2026664210 от 14.05.2026.

Личный вклад соискателя. Основные научные положения, теоретические выводы и практические решения, результаты тестирования, реализация адаптированных алгоритмов и гибридной модели предложены и реализованы автором самостоятельно.

Публикации. По материалам диссертации опубликовано 9 работ, включая 3 публикации в рецензируемых научных изданиях по специальности 2.3.5 из списка высшей аттестационной комиссии, 6 публикаций в РИНЦ; получено 3 свидетельство о регистрации программы для ЭВМ.

Структура и объем диссертации.

Диссертация состоит из введения, четырех глав, заключения, списка литературы, двух приложений. Общий объем диссертации составляет 157 страниц. Диссертация содержит 43 рисунка, 32 таблицы. Список литературы включает 94 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы диссертационного исследования, сформулирована цель диссертации и научная новизна, дана общая характеристика работы, раскрыта практическая значимость полученных результатов, перечислены положения, выносимые на защиту.

В **первой главе** проведен комплексный анализ существующих методов к прогнозированию временных рядов социальных показателей, особое внимание уделено анализу методов работы с короткими временными рядами (менее 30 точек отсчета), что является критически важным для современных социальных явлений, не имеющих достаточной исторической базы. Проведена систематизация методов прогнозирования по степени формализации и специфике применения. Установлено, что все существующие методы могут быть классифицированы по шести основным категориям: экспертные и эвристические методы, методы моделирования, нормативные методы, методы экстраполяции, методы машинного обучения и нейросетевые методы. Проведен комплексный сравнительный анализ всех категорий методов по семи ключевым параметрам: требования к данным, сложность реализации, интерпретируемость результатов, учет взаимосвязей, гибкость/адаптивность, использование экзогенных переменных и эффективность для коротких временных рядов.

Анализ выявил критическое противоречие: методы с высокой эффективностью для коротких временных рядов (экспертные, некоторые нормативные и экстраполяционные) имеют ограниченную способность учитывать экзогенные переменные и сложные взаимосвязи. В то же время методы с высокой способностью учитывать экзогенные переменные и моделировать сложные зависимости (машинное обучение, нейронные сети) неэффективны для коротких временных рядов. Рассмотрены существующие гибридные методы: ARIMA-LSTM, Temporal Fusion Transformer (TFT), декомпозиционные модели, комбинации генетических алгоритмов с нейронными сетями, ансамбли нейронных сетей.

Существующие гибридные методы не решают проблему коротких временных рядов и требуют значительных вычислительных ресурсов. Выявлена необходимость разработки специализированной гибридной модели, адаптированной для коротких временных рядов социальных показателей.

Прогнозирование коротких временных рядов предъявляет особые требования к методам и программному обеспечению: необходимо избегать переобучения, минимизировать накладные расходы на настройку и обеспечить максимальную интерпретируемость. В связи с этим возникает потребность в создании специализированного программного обеспечения, которое позволит автоматизировано создавать гибридные модели прогнозирования на основе параметров входного временного ряда.

Выявленные ограничения существующих методов обосновывают необходимость разработки гибридного адаптивного метода, который сочетал бы преимущества различных методов прогнозирования.

Вторая глава посвящена разработке гибридной модели прогнозирования коротких временных рядов. Математическая постановка задачи: пусть имеется временной ряд:

$$X = \{x_1, x_2, \dots, x_n\}, \text{ где } n \leq 30.$$

Требуется построить прогноз:

$$\hat{Y} = \{\hat{x}_{n+1}, \hat{x}_{n+2}, \dots, \hat{x}_{n+h}\}, h \in \{1, 2, 3, 4, 5\}$$

при ограничениях: 1. $\min(n) \geq 10$ (минимальная длина для статистической значимости) 2. $\max(n) \leq 30$ (короткий ряд) 3. Отсутствие достоверной информации о сезонности 4. Возможность структурных сдвигов.

За основу взят гибридный метод, заключающийся в взвешенной сумме предсказаний, полученных на различных этапах работы алгоритма. Основные метрики качества прогнозирования: $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ $MAPE =$

$$\frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\%, RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}.$$

Результирующий прогноз складывается из прогнозов базовых моделей с учетом ошибки, которые эти модели показали при обучении. Затем полученный прогноз корректируется с помощью «эксперта», в роли которого выступает большая языковая модель (LLM) со специальным контекстом. В качестве базовых моделей служат модели из 3 классов, по одной из каждого класса: экстарполяционные, модели машинного обучения, нейросетевые модели.

Для обеспечения интеграции прогнозов, полученных от базовых моделей SARIMAX, XGBoost и TimeVLM (Временная зрительно-языковая модель) была предложена модель экспоненциальных весов (1), которая учитывает длину входного ряда. Таким образом учитывается необходимость обращать большее внимание на недавние ошибки моделей в связи с короткой исторической выборкой.

$$w_i(t) = \frac{\exp(-\beta \times ER_{i(t)} \times \alpha(n))}{\sum_{j=1}^N \exp(-\beta \times ER_{j(t)} \times \alpha(n))}, \quad (1)$$

где $w_i(t)$ — вес модели i в момент t , β — коэффициент чувствительности ($\beta = 1.2$), $ER_{i(t)}$ — экспоненциально взвешенная ошибка модели i , $\alpha(n)$ — корректирующий коэффициент для коротких рядов. Корректирующий коэффициент $\alpha(n)$ вычисляется по формуле:

$$\alpha(n) = 1 + \ln\left(\frac{30}{n}\right).$$

Логарифмическая коррекция обоснована следующими принципами:

Нелинейность: при уменьшении длины ряда с 30 до 15 точек коэффициент α возрастает с 1.0 до ~ 1.69 , то есть на 69% (≈ 0.05 на каждую исключённую точку). При дальнейшем уменьшении с 15 до 10 точек рост составляет 24% (от 1.69 до 2.10), то есть уже ≈ 0.08 на каждую исключённую точку. Это отражает потерю информативности в диапазоне 10-20 наблюдений, где параметрические модели теряют способность адекватно оценивать тренды и сезонность.

Асимптотическое поведение: Логарифмическая зависимость $\alpha(n)$ обеспечивает максимальное усиление штрафа за ошибку при минимальной длине ряда $n = 10$ ($\alpha \approx 2.10$), что соответствует максимальной неопределённости. При $n = 30$ коэффициент $\alpha(n)$ равен 1, и модель переходит к стандартному EWA-взвешиванию.

Предложенная формула адаптивных весов представляет собой модификацию классического алгоритма экспоненциальных весов (Littlestone & Warmuth, 1994). Благодаря им средний регрет ансамбля убывает к нулю с ростом числа шагов, как $O(\sqrt{\ln(N/T)})$ (Zhang et al., 2024; Flaspohler et al., 2021).

Оригинальная временная зрительно-языковая модель проектировался исключительно для длинных временных рядов с длиной $L \geq 96$, используя фиксированные параметры длины патча $p = 16$, и сдвига $s = 8$, обеспечивающие число патчей $N_p \geq 11$. Патч — это скользящее окно, вырезающее непрерывный фрагмент ряда. Для адаптации данной модели для коротких временных рядов с $L \leq 30$ предложено перевести параметры p и s в гиперпараметры, подбираемые во время обучения модели. Для гарантии невырожденности матрицы весов внимания вводится пороговая функция адаптации для коротких временных рядов:

$$N_p = \left\lfloor \frac{L - p(L)}{s(L)} \right\rfloor + 1; p(L) \in \{3, 4, 5\}, s(L) \in \{1, 2\},$$

где N_p — число патчей, $p(L)$ — длина одного патча, $s(L)$ — сдвиг между патчами.

Прогнозы, полученные на малых выборках, часто обладают систематическим смещением, вызванным недостаточной репрезентативностью данных. Для его компенсации вводится поправка $\Delta_{bias}(n, h)$. Скорректированный прогноз вычисляется по формуле:

$$\hat{Y}_{corrected}(t + h) = \hat{Y}_{basic}(t + h) + \Delta_{bias}(n, h).$$

Смещение малой выборки $\Delta_{bias}(n, h)$ рассчитывается следующим образом:

$$\Delta_{bias}(n, h) = \gamma \times h \times \hat{\sigma} \times \sqrt{\frac{1}{n} + \frac{1}{T}},$$

где γ — коэффициент компенсации ($\gamma = 0.15$),

T — теоретическая длина среднего ряда ($T = 30$),

$\hat{\sigma}$ — оценка стандартного отклонения.

Данная формула теоретически обоснована результатами Pesaran & Timmermann (2004).

LLM-эксперт преобразует качественную информацию в количественную коррекцию через контекстный вектор $C = [Y_{hist}, \hat{y}_{ens}, X_{meta}, W_{web}]$, где $Y_{hist} = \{y_1, \dots, y_n\}$ — исторический временной ряд, $\hat{y}_{ens} = \{\hat{y}_1, \hat{y}_2, \hat{y}_3\}$ — прогнозы базовых моделей, X_{meta} — метаданные (например экзогенные переменные), W_{web} — веб-контекст.

Текстовый запрос формируется как структурированный запрос:

$$Prompt = f_{struct}(C, Role, Task)$$

Для обеспечения регулировки уровня доверия к эксперту предложен коэффициент доверия:

$$w_{LLM(n)} = \begin{cases} 0.15 + 0.35 \times \frac{n-5}{15}, & \text{если } 5 \leq n \leq 30; \\ 0.5, & \text{если } n > 30; \\ 0.1, & \text{если } n < 5. \end{cases}$$

Линейная зависимость веса $w_{LLM(n)}$ от n в диапазоне $5 \leq n \leq 30$ обеспечивает плавное увеличение доверия к экспертной коррекции по мере роста статистической базы. Форма вычисления коэффициента обоснована результатами экспериментов на реальных и синтетических данных.

Итоговый прогноз вычисляется по формуле:

$$\hat{Y}_{final}(t+h) = \hat{Y}_{ensemble}(t+h) + \Delta_{bias}(n,h) + w_{LLM}(n) \times \Delta_{LLM}(t+h),$$

где $\hat{Y}_{ensemble}$ — ансамблевый прогноз, $\Delta_{bias}(n,h)$ — корректировка смещения для короткого временного ряда, $w_{LLM}(n)$ — доверие к LLM-эксперту, Δ_{LLM} — поправка LLM-эксперта.

В третьей главе разработаны алгоритмы, реализующие созданную модель. Основными алгоритмами, разработанными для реализации предложенной модели, являются алгоритмы гибридного прогнозирования, генерации запросов и алгоритм коррекции с помощью LLM-эксперта. В данном случае LLM-эксперт — это большая языковая модель, обогащенная внешним контекстом, релевантным временному ряду, поступившему на вход. Система использует клиент-серверную архитектуру с разделением на 3 уровня: презентационный уровень (веб-интерфейс), бизнес-логику (API и модели) и уровень данных (обработка временных рядов). Архитектура системы представлена на Рисунке 1.

Процесс прогнозирования включает следующие этапы:

1. Загрузка файла с расширением «.csv» или «.xlsx», который содержит в себе 2 колонки: дата, значение параметра.
2. От интерфейсной части данные и параметры прогноза поступают к серверной части.
3. Начинается этап обучения выбранной модели.
4. С помощью специальной библиотеки подбираются параметры моделей, которые дают наилучшую точность.

5. Модели обучаются на подобранных параметрах и генерируют прогнозы.
6. LLM-эксперт анализирует прогнозы и вносит коррекции.
7. Серверная часть возвращает прогноз, который можно скачать в файлы формата «.csv» или «.xlsx».

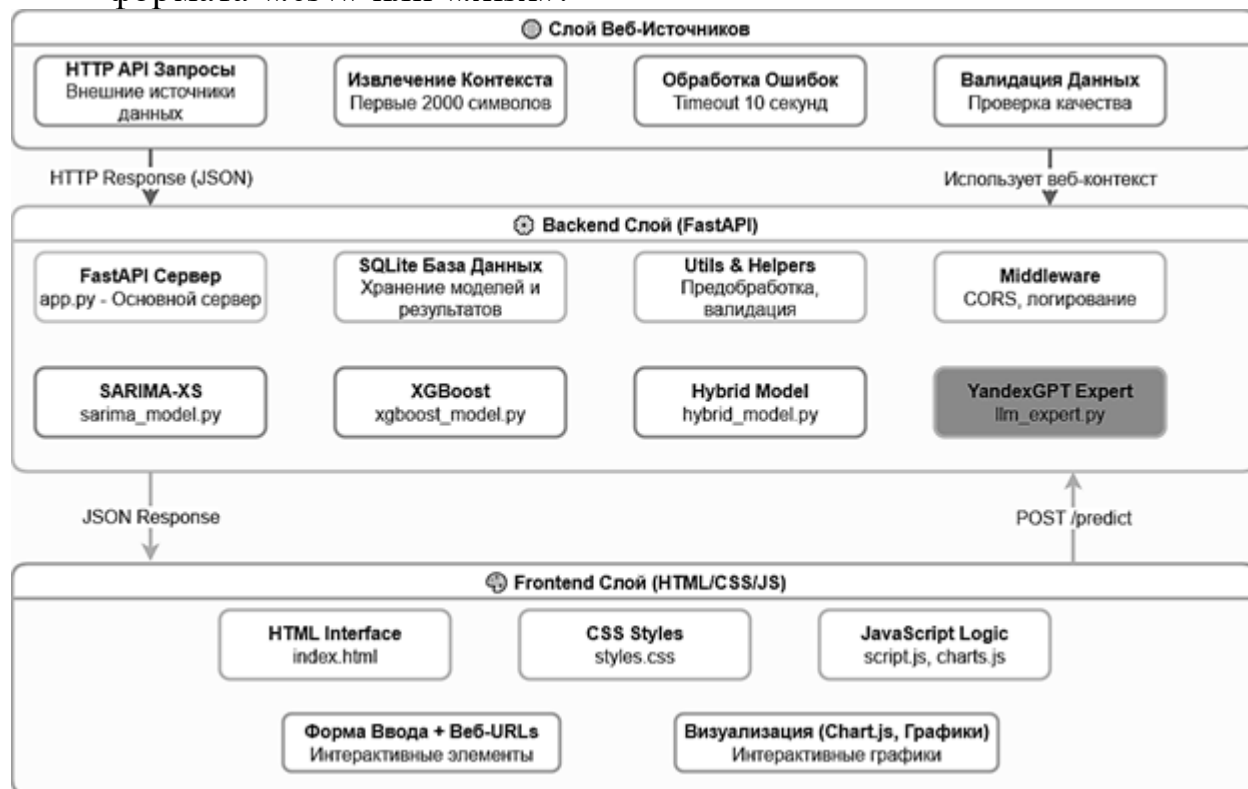


Рисунок 1. Архитектура разработанной системы

Одним из главных компонентов реализованной программной системы является алгоритм по автоматизированной экспертной коррекции полученных базовых прогнозов. Данный алгоритм включает в себя обработку и анализ источников, предоставленных пользователем при формировании запроса и автоматическое формирование текстового запроса. Псевдокод алгоритма представлен в листинге 1.

Листинг 1. Псевдокод алгоритма формирования текстового запроса

Вход: задача $\mathcal{T}$, контекст $\mathcal{C} = \{\text{facts, audience, goal}\}$, ограничения $\Phi = \{\varphi_i\}$, примеры $\mathcal{E}$, бюджет токенов τ_{max}

Выход: структурированный промпт p^*

```

1  $p \leftarrow \emptyset$ 
2  $p.\text{role} \leftarrow \text{AssignExpertRole}(\mathcal{T}.\text{domain})$     ▷ Роль
3  $p.\text{ctx} \leftarrow \langle \mathcal{C}.\text{facts}, \mathcal{C}.\text{audience}, \mathcal{C}.\text{goal} \rangle$     ▷ Контекст
4  $p.\text{task} \leftarrow \mathcal{T}.\text{description}$     ▷ Задача
5 if  $\mathcal{E} \neq \emptyset$  then  $p.\text{task} \leftarrow p.\text{task} \parallel \text{FormatShots}(\mathcal{E})$     ▷ Few-Shot
6  $p.\text{constr} \leftarrow \parallel_{\{\varphi \in \Phi\}} \varphi$     ▷ Ограничения (конкатенация)
7  $p^* \leftarrow p.\text{role} \parallel p.\text{ctx} \parallel p.\text{task} \parallel p.\text{constr} \parallel \text{"Answer:"}$ 
8 if  $|p^*| > \tau_{\text{max}}$  then    ▷ Валидация по бюджету
9    $p.\text{ctx} \leftarrow \text{Truncate}(p.\text{ctx}, |p.\text{ctx}| - (|p^*| - \tau_{\text{max}}))$ 
10  $p^* \leftarrow \text{Reassemble}(p)$ 

```

11 return p^*

При анализе источников обработке источников также используется языковая модель, так как входные данные могут различную структуру, и реализация универсального парсера в данном случае не представляется возможным. В итоговом варианте исходные данные преобразуются в выжимки по каждому источнику. Эти выжимки должны содержать наиболее релевантную по отношению к предметной области информацию. На этапе формирования запроса осуществляется интеграция прогнозов, полученных на предыдущих этапах, а также формирование дополнительных признаков: разности между прогнозами, статистические данные о временном ряде. Также важным фактором является наделение эксперта «ролью», которая формируется на основе контекста, полученного из метаданных, предоставляемых пользователем.

Одной из базовых моделей, прогноз которой используется в гибридной модели, является адаптированная модель TimeVLM. Адаптация заключается в использовании замороженного визуального энкодера, а также динамических патчах. Псевдокод алгоритма представлен в листинге 2.

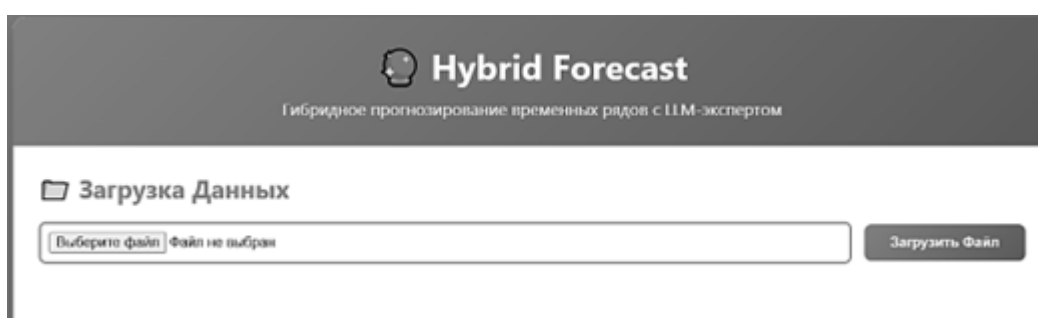
Листинг 2. Псевдокод адаптированного алгоритма TimeVLM

Вход: временной ряд $x_{en} \in \mathbb{R}^{I \times L \times D}$, длина патча p_l , страйд s_t , память $M \in \mathbb{R}^{M \times D M}_{ex}$, горизонт прогноза T_{ph}

Выход: прогноз $\hat{y} \in \mathbb{R}^{I \times T_{ph} \times D}$, обновлённая память M'

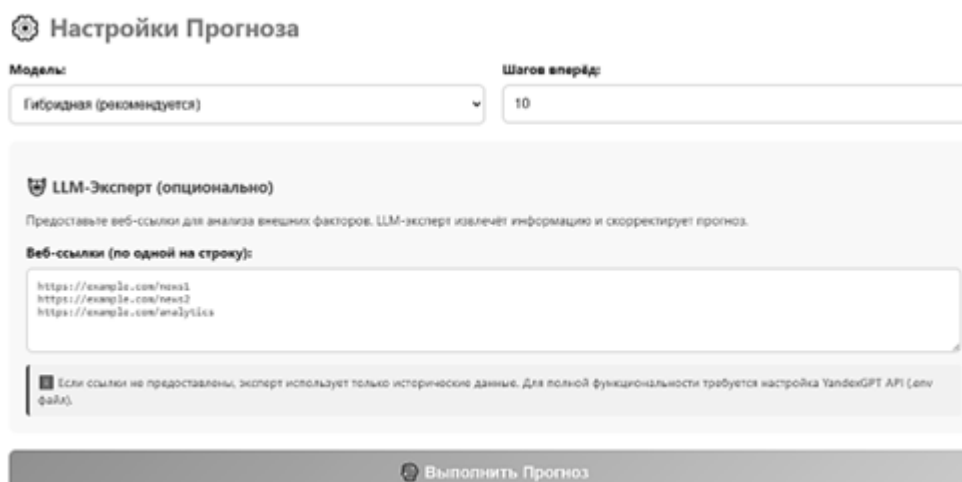
```
1  E_p ← PatchEmbed( $x_{\{en\}}$ ,  $p_l$ ,  $s_t$ )           ∈  $\mathbb{R}^{\wedge\{I \times N_p \times D_{\{max\}}\}}$ ,  $N_p = \lfloor (L - p_l) / s_t \rfloor + 1$ 
2  M ← CircularUpdate( M, mean( $E_p$ , dim=2) )    ▷ обновление банка памяти
3  A ← CosineSim( $E_p$ , M)                        ▷ сходство: локальные признаки
4  F_local ← TopK(A, k) || SelfAttn(TopK(A,k))  ▷ top-k + глобальные зависимости
5  F_global ← mean(A, dim=2)                    ▷ агрегированный контекст
6  F_tem ← F_local || F_global                  ∈  $\mathbb{R}^{\wedge\{I \times D_{\{max\}}\}}$  ▷ темпоральное представление
7  P ← PosEnc( $x_{\{en\}}$ ) || FreqEnc( $x_{\{en\}}$ )    ▷ фаза VAL: позиционное + частотное кодирование
8  I_raw ← MultiScaleConv(P, scales={7,14,30}) → 3-канальное изображение
9  I ← BicubicResize(I_raw,  $H_{\{VLM\}} \times W_{\{VLM\}}$ ) ▷ нормировка в [0,255]
10 T ←  $\langle \mu(x_{\{en\}}), \sigma^2(x_{\{en\}}), Trend(x_{\{en\}}) \rangle$  ▷ фаза TAL: статистики + тренд
11 Prompt ← Template(T, F_tem)                 ▷ адаптированный текст (Алгоритм 1)
12 F_vlm ← FrozenVLM(I, Prompt)                ∈  $\mathbb{R}^{\wedge\{I \times L_f \times D_h\}}$  ▷ фаза 4, параметры заморожены
13 G ←  $\sigma$ ( CrossAttn(F_tem, F_vlm) )         ∈  $(0,1)^{\wedge\{I \times D_{\{max\}}\}}$  ▷ мультимодальные ворота
14 F_fused ←  $G \odot F_{tem} + (1 - G) \odot Proj(F_{vlm})$  ▷ взвешенное слияние
15  $\hat{y} \leftarrow Linear(F_{fused})$              ∈  $\mathbb{R}^{\wedge\{I \times T_{\{ph\}} \times D\}}$  ▷ прогнозирование
16 return  $\hat{y}$ , M
```

Формы интерфейса представлены на рисунках 2-4. Пользователь может загрузить файл формата «.csv» или «.xlsx», а также самостоятельно выбрать необходимую модель, которая будет использована для прогноза и сам горизонт прогноза. На рисунке 3 представлена форма загрузки файла. На рисунке 4 представлена форма настройки интерфейса. Пользователь может выбрать необходимую ему модель (Гибридная, SARIMAX, XGBoost, TimeVLM), а также указать горизонт прогноза и релевантные веб-ресурсы, которые будет обрабатывать LLM-эксперт. На рисунке 4 представлена форма получения прогноза. В результате прогнозирования пользователь получает интерактивный график с указанием прогнозных значений. Указываются ошибки, которые были получены выбранной моделью на этапе обучения. Также выводится комментарий LLM-эксперта и веса моделей в случае выбора гибридной модели для прогноза.



The screenshot shows the 'Hybrid Forecast' application interface. At the top, there is a header with the title 'Hybrid Forecast' and a subtitle 'Гибридное прогнозирование временных рядов с LLM-экспертом'. Below the header, there is a section titled 'Загрузка Данных' (Data Upload). It features a file selection button labeled 'Выберите файл' (Select file) and a status indicator 'Файл не выбран' (File not selected). To the right of the file selection area is a button labeled 'Загрузить Файл' (Upload File).

Рисунок 2. Форма загрузки



The screenshot shows the 'Настройки Прогноза' (Forecast Settings) interface. It includes a 'Модели:' (Models) dropdown menu with 'Гибридная (рекомендуется)' (Hybrid (recommended)) selected. To the right is a 'Шагов вперёд:' (Steps forward) input field with the value '10'. Below these is a section for 'LLM-Эксперт (опционально)' (LLM Expert (optional)). It contains a text area for 'Веб-ссылки (по одной на строку):' (Web links (one per line):) with example URLs. At the bottom, there is a button labeled 'Выполнить Прогноз' (Execute Forecast).

Рисунок 3. Форма настройки прогноза



Рисунок 4. Форма с результатами прогноза

В четвертой главе описано практическое применение разработанной гибридной модели на различных наборах данных.

Для проведения экспериментального исследования использованы шесть наборов реальных социальных данных Российской Федерации, охватывающих период с 1991 года по 2025 год. Выбор данного периода обусловлен наличием существенных структурных изменений в российском обществе, что создаёт дополнительные сложности для прогнозирования и позволяет оценить устойчивость моделей в условиях нестационарности. Характеристики наборов данных представлены в таблице 1.

Таблица 1. Характеристики экспериментальных наборов данных

№	Наименование показателя	Источник	Период	n_train
1	Число музеев в РФ	Росстат	2015–2025 (полугодовая частота)	15
2	Коэффициент смертности	Росстат	1991-2023	29
3	Ожидаемая продолжительность жизни	Росстат	2005–2024	15
4	Уровень безработицы среди молодежи	Росстат	1991–2024	29
5	Работники туриндустрии	Каф. туризма МГУ	2013–2024	9

6	Индекс человеческого развития	ООН	2000-2023	21
---	-------------------------------	-----	-----------	----

Результаты тестирования моделей представлены в таблицах 2-7.

Таблица 2. Результаты тестирования для первого набора

Модель	MAE	MAPE	RMSE
SARIMAX	53.8965	1.7295%	62.3285
XGBoost	83.8423	2.6948%	91.5302
TimeVLM	103.4427	3.3585%	138.0144
Гибридная модель	28.3492	0.9099%	32.4647

Таблица 3. Результаты тестирования для второго набора

Модель	MAE	MAPE	RMSE
SARIMAX	1.9265	12.4061%	2.5854
XGBoost	1.8461	11.9378%	2.4522
TimeVLM	2.4475	15.9239%	3.4410
Гибридная модель	1.7328	11.2262%	2.3081

Таблица 4. Результаты тестирования для третьего набора

Модель	MAE	MAPE	RMSE
SARIMAX	2.9103	4.0515%	3.0307
XGBoost	2.8943	4.0286%	3.0145
TimeVLM	3.2274	4.4847%	3.3578
Гибридная модель	1.7557	2.4569%	2.0089

Таблица 5. Результаты тестирования для четвертого набора

Модель	MAE	MAPE	RMSE
SARIMAX	2.5741	24.2272%	3.1569
XGBoost	3.1139	31.7629%	4.3357
TimeVLM	4.3436	41.5386%	5.3153
Гибридная модель	2.4980	24.2199%	3.2224

Таблица 6. Результаты тестирования для пятого набора

Модель	MAE	MAPE	RMSE
SARIMAX	16436.0000	1.5745%	16559.1497
XGBoost	18556.3211	1.8059%	24788.6900
TimeVLM	77507.5625	7.4233%	78204.0097
Гибридная модель	16436.0000	1.5558%	21801.2798

Таблица 7. Результаты тестирования для шестого набора

Модель	MAE	MAPE	RMSE
SARIMAX	0.0118	1.4355%	0.0121
XGBoost	0.0063	0.7701%	0.0065
TimeVLM	0.0198	2.3957%	0.0223
Гибридная модель	0.0049	0.5932%	0.0056

Из результатов эксперимента видно, что предложенная модель показала наименьшую ошибку MAPE на всех наборах данных.

ОСНОВНЫЕ ВЫВОДЫ И ЗАКЛЮЧЕНИЕ

В **заклучении** сформулированы основные результаты исследования:

- 1) Проведен анализ существующих моделей и методов прогнозирования временных рядов;
- 2) Разработана гибридная модель прогнозирования и архитектура программного обеспечения интеллектуальной системы на основе большой языковой модели;
- 3) Разработан гибридный алгоритм прогнозирования коротких временных рядов;
- 4) Проведено тестирование разработанного программного обеспечения на реальных данных.

В диссертации решена актуальная **научная задача**, заключающаяся в разработке нового гибридного метода прогнозирования коротких временных рядов и оригинального программного обеспечения.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Цаплин С.Т., Березкин Д.В. Применение методов бутстрэппинга для коротких временных рядов с целью повышения точности предиктивных моделей / Цаплин С. Т., Берёзкин Д. В. // Управление качеством инженерного образования: перспективы искусственного интеллекта : материалы 3-ей Международной конференции, посвящённой 70-летию кафедры ИУ6, Москва, 30 октября-2 ноября 2023 года / МГТУ им. Н. Э.

- Баумана (национальный исследовательский университет). - М., 2024. - С. 99-102.
2. Цаплин С. Т., Березкин Д. В. Применение гибридных моделей для прогнозирования временных рядов, описывающих динамику вакансий // Конференция «Ломоносов-2025». Секция «Искусственный интеллект и цифровая трансформация в бизнесе и государственном управлении». Москва, 2025. URL: https://confhub.ru/archive/Lomonosov_2025/data/37122/182518_uid1042084_report.pdf
 3. Цаплин С. Т., Березкин Д. В., Применение коммерческих LLM для прогнозирования временных рядов // Материалы Всероссийской студенческой конференции «Студенческая научная весна», посвящённой 85-летию со дня рождения академика И.Б. Фёдорова. — Москва, 2025. — С. 208-209.
 4. Цаплин С. Т., Березкин Д. В., Можаров Г. П., Сравнительный анализ методов аугментации коротких временных рядов // «Динамика сложных систем - XXI век», № 2, 2026 г.
 5. Цаплин С. Т. Возможности и проблемы прогнозирования коротких временных рядов // Сборник материалов Международной научно-практической конференции "Инновационные, информационные и коммуникационные технологии"
 6. Цаплин С. Т., Гибридное моделирование для долгосрочных прогнозов // Машиностроение: сетевой электронный научный журнал
 7. Цаплин С. Т., Анализ возможностей применения VLM для прогнозирования временных рядов // Информационно-измерительные и управляющие системы № 3, 2026 г.
 8. Цаплин С. Т., Прогнозирование социально-экономических показателей российской федерации на основе гибридного подхода, Материалы XI Международного научного конгресса по информатике (CSIST-2025) Республика Беларусь, Минск, 29–31 октября 2025 г.
 9. Пролетарский А. В., Березкин Д. В., Цаплин С. Т., Применение генеративных нейронных сетей для предсказания экстремально коротких временных рядов // Нейрокомпьютеры: разработка, применение, № 1, 2026 г.
 10. Свидетельство о регистрации программы для ЭВМ «Гибридное прогнозирование временных рядов», № 2026511275 от 17.01.2024
 11. Свидетельство о регистрации программы для ЭВМ «Программный модуль TimeVLM, адаптированный для прогнозирования коротких временных рядов» № 2026616858 от 12.03.2026
 12. Свидетельство о регистрации программы для ЭВМ «Программный модуль LLM-эксперта для прогнозирования коротких временных рядов» № 2026664210 от 14.05.2026