

УДК 004.89

DOI: [10.26102/2310-6018/2025.50.3.009](https://doi.org/10.26102/2310-6018/2025.50.3.009)

Исследование и оценка качества аннотаций на естественном языке, сгенерированных мультиагентной системой

А.И. Кузнецова ✉, В.В. Носкин

*Волгоградский государственный технический университет, Волгоград,
Российская Федерация*

Резюме. Исследование посвящено оценке качества аннотаций на русском языке, сгенерированных мультиагентной системой для анализа временных рядов. Система состоит из четырех специализированных независимых агентов: аналитик дашборда, аналитик временного ряда, доменно-специфичный агент и агент для взаимодействия с пользователем. Аннотации генерируются на основе данных дашборда и временного ряда, анализируемых с использованием модели GPT-4o-mini и графа задач для агентов на базе LangGraph. Оценка качества аннотаций проводилась по метрикам понятности, читаемости, контекстуальной уместности и грамотности, а также с использованием адаптированной формулы индекса удобочитаемости Флеша для русского языка. Было разработано тестирование и проведено с участием 21 пользователя на 10 дашбордах – итого 210 оценок по десятибалльной шкале для каждого из показателей. Проведенная оценка и результаты показали эффективность аннотаций: понятность – 8,486, читаемость – 8,705, соответствие контексту – 8,890, грамотность – 8,724. Индекс удобочитаемости составил 33,6, что показывает среднюю сложность текста. Но такой показатель связан со спецификой области исследования и не учитывает расположение слов и их контекст, а только статические показатели длины. Взрослый человек и неспециалист в этой области способен воспринимать сложные слова в аннотации, что доказывают другие оценки. Все оставленные пользователями замечания будут учтены для улучшения формата и интерактивности системы в дальнейшем исследовании.

Ключевые слова: временной ряд, генерация аннотаций, LLM, мультиагентная система, дашборды.

Для цитирования: Кузнецова А.И., Носкин В.В. Исследование и оценка качества аннотаций на естественном языке, сгенерированных мультиагентной системой. *Моделирование, оптимизация и информационные технологии*. 2025;13(3). URL: <https://moitvvt.ru/ru/journal/pdf?id=1967> DOI: 10.26102/2310-6018/2025.50.3.009

Research and evaluation of the quality of natural language annotations generated by the multi-agent system

A.I. Kuznetsova ✉, V.V. Noskin

Volgograd State Technical University, Volgograd, the Russian Federation

Abstract. This study is devoted to assessing the quality of annotations in Russian generated by a multi-agent system for time series analysis. The system includes four specialized agents: a dashboard analyst, a time series analyst, a domain-specific agent, and an agent for user interaction. Annotations are generated by analyzing dashboard and time series data using the GPT-4o-mini model and a task graph implemented with LangGraph. The quality of the annotations was assessed using the metrics of clarity, readability, contextual relevance, and literacy, as well as using an adapted Flesch readability index formula for the Russian language. Testing was developed and conducted with the participation of 21 users on 10 dashboards – a total of 210 ratings on a ten-point scale for each of the metrics. The assessment and results showed the effectiveness of annotations: clarity - 8.486, readability - 8.705, contextual relevance – 8.890, literacy – 8.724. The readability index was 33.6, which shows the average

complexity of the text. This indicator is related to the specifics of the research area and does not take into account the arrangement of words and their context, but only static length indicators. An adult and a non-specialist in each field are able to perceive complex words in the annotation, which is proven by other ratings. All comments left by users will be taken into account to improve the format and interactivity of the system in further research.

Keywords: time series, annotation generation, LLM, multi-agent system, dashboards.

For citation: Kuznetsova A.I., Noskin V.V. Research and evaluation of the quality of natural language annotations generated by the multi-agent system. *Modeling, Optimization and Information Technology*. 2025;13(3). (In Russ.). URL: <https://moitvivr.ru/ru/journal/pdf?id=1967> DOI: 10.26102/2310-6018/2025.50.3.009

Введение

В условиях стремительного роста объема данных, их анализ и понимание играют ключевую роль для принятия эффективных бизнес-решений. Аналитики формируют отчеты, включающие в себя множество связанных графиков, метрик, диаграмм и таблиц для интерпретации сырых данных. Такие отчеты называются дашбордами – этот инструмент помогает аналитикам быстрее ориентироваться в приходящих данных, представляя их в более удобном визуальном формате. Однако анализ все еще требует значительных временных затрат и когнитивной нагрузки на аналитиков. Традиционные методы визуального анализа и ручные аннотации занимают много времени и не справляются с информационно насыщенными дашбордами.

Исследование [1] подтверждает, что использование естественно языковых описаний улучшает интерпретацию и понимание данных пользователями. Составление аннотаций (краткого содержания) для больших объемов данных, таким образом, помогает в восприятии и анализе аналитикам. С такой задачей хорошо справляются модели генерации текста NLG (Natural Language Generation), к которым относят набирающие популярность большие языковые модели (LLM). Модели, такие как Grok, DeepSeek, ChatGPT и другие, хорошо показывают себя во многих задачах, связанных не только с генерацией текста. Например, в обзоре [2] рассматривается применение больших языковых моделей для анализа временных рядов (последовательностей данных, изменяющихся во времени). Способность LLM к пониманию последовательностей данных и генерации текста делает такой инструмент перспективным для задач, связанных с временными рядами, таких как прогнозирование, классификация и аннотирование. Авторы подчеркивают потенциал больших языковых моделей в автоматизации задач, таких как медицина и финансы, где аннотации улучшают интерпретируемость получаемых временных рядов. Также в работе [3] рассматривается проблема понимания вопроса: могут ли большие языковые модели эффективно анализировать временные ряды, обрабатывая сложные данные и извлекая полезные инсайты? По результатам эксперимента LLM показывают хорошие результаты в обнаружении аномалий, но в задаче прогнозирования их преимущества менее очевидны – иногда простые модели работают не хуже или даже лучше. По представленным работам можно сделать вывод, что, большие языковые модели имеют хороший потенциал для поиска зависимостей и выявления аномалий во временных рядах.

Временные ряды являются одной из наиболее распространенных категорий данных в бизнесе. Для принятия более эффективных бизнес-решений, аналитики исследуют сезонность, тренды, аномалии временного ряда, но не всегда можно быстро и точно заметить скрытые инсайты в данных или на графике. Автоматическая генерация аннотаций на естественном языке для временных рядов может значительно ускорить

процесс их интерпретации, что, в свою очередь, позволит аналитикам быстрее выявлять ключевые закономерности и принимать обоснованные решения.

Целью данного исследования является оценка качества аннотаций на русском языке, генерируемых разработанной мультиагентной системой, предназначенной для анализа временных рядов.

Новизна подхода заключается в использовании мультиагентной архитектуры, где задачи анализа распределяются между специализированными агентами: аналитик дашборда, аналитик временного ряда, доменно-специфичный агент и агент для взаимодействия с пользователем. В дальнейшем исследовании предполагается итеративное взаимодействие аналитика с системой, которое позволит сократить время на анализ дашборда.

Так как оценка аннотации является по большей части субъективной, для оценки качества была разработана методология тестирования в виде опроса. Метрики оценки – полезность, ясность, контекстуальная уместность, грамотность. Также используется метрика индекса удобочитаемости Флеша (Flesch-Kincaid) для оценки читаемости сгенерированного текста.

Материалы и методы

Описание архитектуры. Мультиагентная архитектура предполагает, что система состоит из нескольких автономных агентов, каждый из которых выполняет специализированные функции и взаимодействует с другими агентами. Разработанное приложение представляет собой такую мультиагентную систему для анализа временных рядов и автоматической генерации аннотаций на естественном языке.

В статье [4] рассматривается применение LLM в качестве агента для анализа временных рядов. Авторы предлагают идею, которая подразумевает создание такой системы, которая будет не только выполнять задачи анализа временных рядов (таких как прогнозирование, обнаружение аномалий, классификация и аннотирование), но и отвечать на вопросы пользователей, помнить и понимать контекст, а также обрабатывать разные типы данных.

Авторы работы [5] разработали фреймворк TESSA – мультиагентную систему на базе LLM, предназначенную для генерации аннотаций временных рядов. Суть заключается в том, что общий агент генерирует обобщенные аннотации, а доменно-специфичный агент адаптирует аннотации под конкретную предметную область с учетом ее терминологии и специфики. Авторы объединяют признаки из текста и временных рядов, что позволяет эффективно интерпретировать данные.

Мультиагентные системы [6] состоят из автономных агентов, взаимодействующих друг с другом для решения сложных задач. К характеристикам мультиагентных систем относят: отсутствие глобального контроля, децентрализованные данные, асинхронные вычисления, каждый агент обладает неполной информацией или ограниченными возможностями. Разработанная архитектура соответствует данным принципам, в нее были заложены идеи описанных выше работ. Но в отличие от них, состоит из четырех агентов: агент аналитик дашборда, агент аналитик временного ряда, доменно-специфичный агент, агент для взаимодействия с пользователем.

Агент аналитик временного ряда проводит анализ закономерностей, наиболее часто выявляемых настоящими аналитиками в бизнесе. К ним относят следующие характеристики: тренд, сезонность, максимум и минимум в данных, наличие аномалий. Для дальнейшего аннотирования нужно понимать определения перечисленных характеристик. Сезонность бывает детерминированная и случайная [7]. Детерминированная сезонность описывает поведение временного ряда, при котором

условное среднее значение процесса изменяется в зависимости от времени года или определенного циклического паттерна. Случайная сезонность отличается от детерминированной тем, что сезонные паттерны не являются строго фиксированными. Тренд относится к любому систематическому изменению уровня ряда, то есть его долгосрочному направлению [8]. Тренд временного ряда может быть восходящим, нисходящим и стационарным. Как направление, так и наклон (скорость изменения) тренда могут оставаться постоянными или изменяться на протяжении всего ряда. Термин «аномалия» определяется как отклонение от общего распределения данных, например, к отдельному наблюдению или серии наблюдений [9]. Наиболее значимыми являются те отклонения, которые существенно отличаются от нормы. К примерам можно отнести сбои в оборудовании, финансовые кризисы, медицинские аномалии. Все перечисленные характеристики анализирует разработанная система.

Приложение состоит из нескольких модулей, взаимодействующих через граф задач, реализованный с помощью библиотеки LangGraph. Пользователь загружает изображение дашборда и данные временного ряда, для которых хочет получить аннотацию. Пример интерфейса представлен на Рисунке 1. Интерфейс на базе Streamlit обеспечивает реактивное обновление интерфейса, а функции обратного вызова синхронизируют состояние интерфейса с графом задач. Например, загрузка нового изображения очищает директорию и историю чата, иницилируя новый цикл анализа.

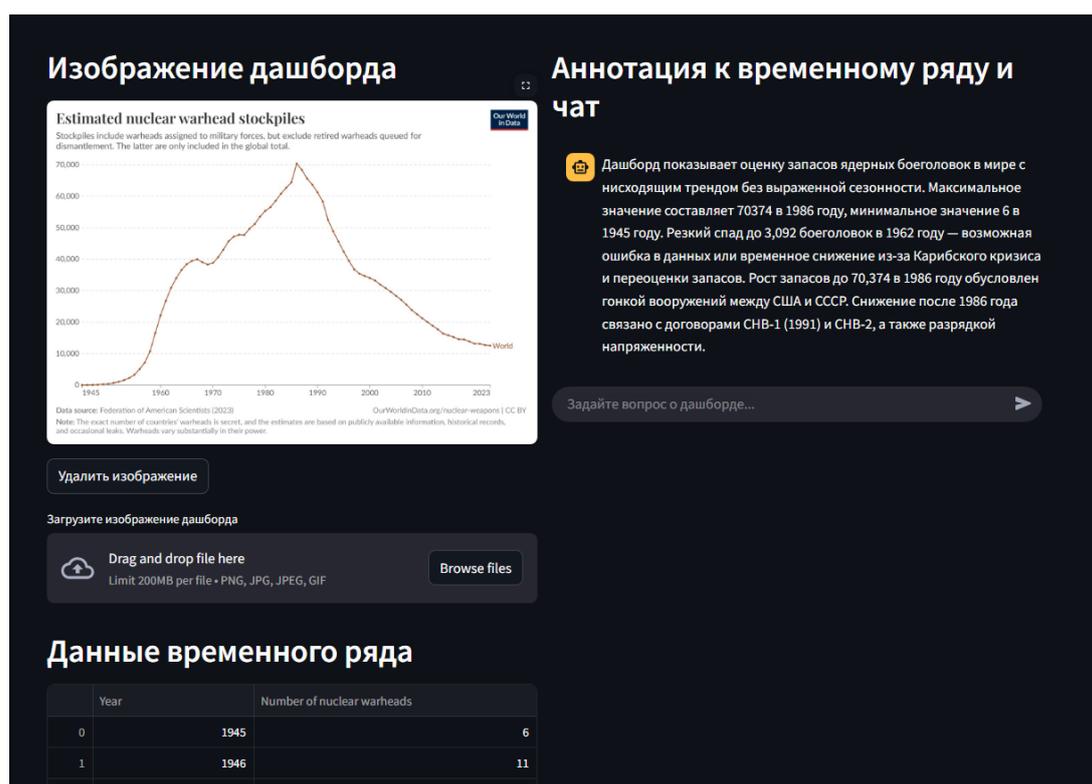


Рисунок 1 – Интерфейс для взаимодействия с системой
 Figure 1 – Interface for interaction with the system

Агент-аналитик дашборда принимает на вход изображение с дашбордом и отвечает за анализ этого изображения с использованием модели GPT-4o-mini. Для определения признаков был разработан промпт, в ответ на который модель выдает ответ в формате JSON с выявленными визуальными метриками на графике.

Адаптация к домену происходит также на основе LLM модели GPT-4o-mini, с помощью которой определяется область задачи на основе изображения дашборда и первых 100 строк данных.

Затем агент-аналитик временных рядов получает на вход изображение, полные данные, метрику и область дашборда. Агент выполняет анализ данных на основе модели GPT-4o-mini с помощью цепочки промптов для определения тренда, сезонности, аномалий, минимальных и максимальных значений из данных, а также предположения гипотез на основе всех полученных данных. Ответ возвращается в формате JSON и передается агенту чата.

Агент чата и генерации аннотаций отвечает за обработку пользовательских запросов и «очеловечивание» информации, присланной в формате JSON от агента-аналитика временных рядов. Он собирает всю информацию по плану в промпте, поэтому структура аннотаций получается примерно схожей – метрика, область, тренд, сезонность, максимум, минимум, аномалии и гипотезы. Затем описывает все на естественном русском языке и выдает аннотацию для пользователя.

Компоненты системы взаимодействуют через асинхронный граф задач, что позволяет эффективно обрабатывать как автоматический анализ (при загрузке данных), так и пользовательские запросы к сформированной аннотации. На Рисунке 2 представлена схема узлов графа задач системы.

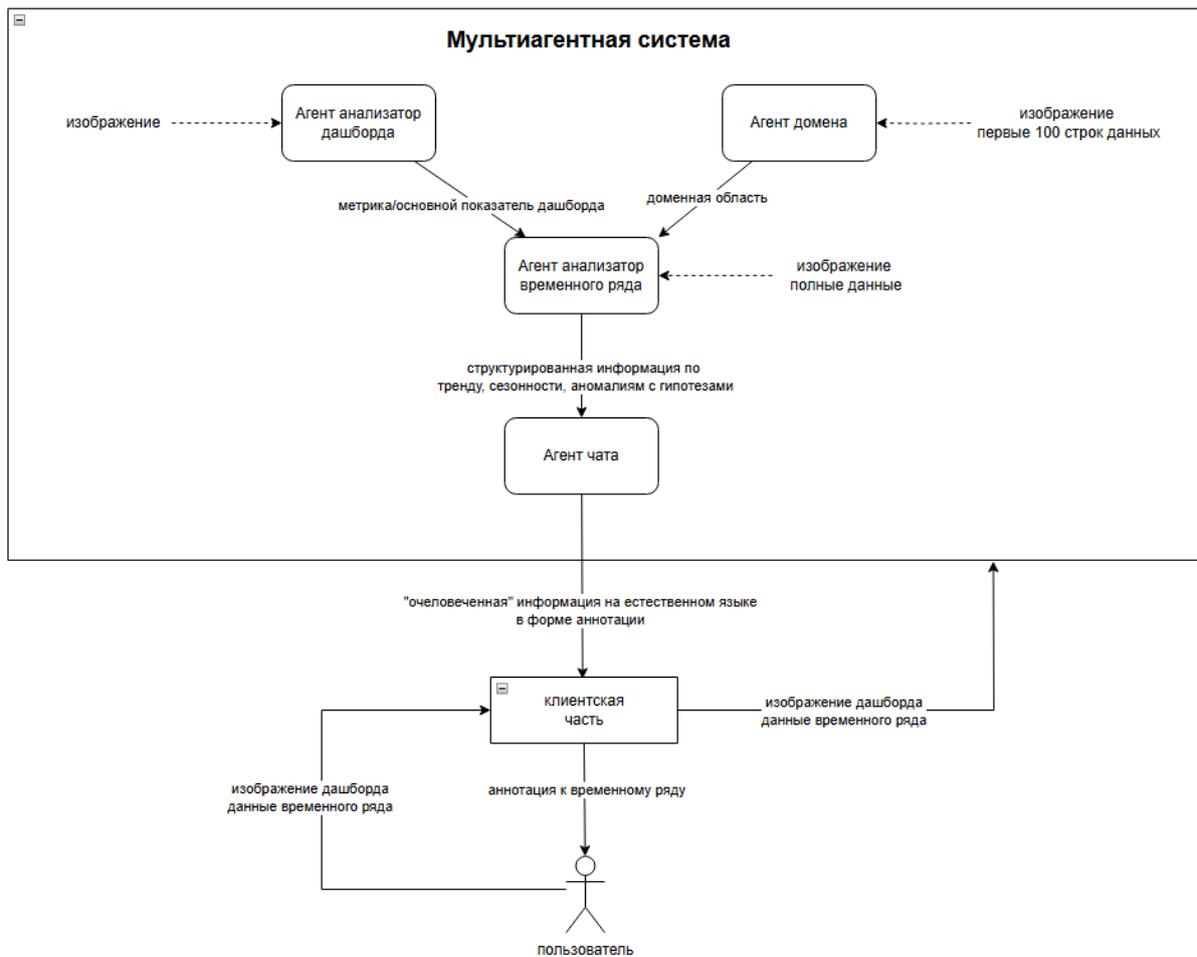


Рисунок 2 – Схема узлов графа задач системы
 Figure 2 – Scheme of nodes of the system task graph

Далее предполагается взаимодействие пользователей с системой. Тогда агент чата присылает запрос пользователя всем агентам, а они уже решают, в их ли области «компетенций» находится запрос и отвечают на него. Агент-аналитик дашборда отвечает за метрики, прочие графики, подписи и за изображение в целом; агент аналитик временного ряда отвечает за информацию по временному ряду и гипотезы; доменный агент может ответить на вопрос по специфике области дашборда. Если ответов несколько, агент чата анализирует их и собирает в один неповторяющийся ответ для пользователя, который позволит расширить понимание дашборда и данных и ускорить процесс принятия решений аналитиком.

Система является гибкой для изменений, ее можно адаптировать и для других задач, например, прогнозирования. Также можно гибко изменять существующие модули, настраивать промпты и заменять компоненты, что представляет большие возможности для разработчиков. Однако, стоит понимать, что у использования больших языковых моделей существуют ограничения. К ним можно отнести зависимость от качества данных и галлюцинации. Галлюцинации – это ошибки, которые модель допускает в своих ответах, включая выдуманные факты, предвзятость и логические расхождения [10]. Таким образом, несмотря на адаптивность и гибкость системы, ее эффективность зависит от стандартных ограничений больших языковых моделей.

Тестирование и метрики. Тестирование было проведено в формате опроса пользователей, в ходе которого был опрошен 21 человек. Предполагалась следующая работа с опросом: было предложено десять сгенерированных системой аннотаций к десяти собранным дашбордам с временными рядами (ссылки на источники можно найти в результатах и в списке литературы). Для оценки использовались метрики: понятность, читаемость, соответствие контексту, грамотность. Также предлагалось дополнить личными комментариями каждую аннотацию для улучшения качества текста и дальнейшего тестирования.

Принцип работы с опросом:

1. Посмотреть на картинку: оценить показатель, сезонность, тренд, максимальные и минимальные значения.
2. Предположить, с чем могут быть связаны такие закономерности.
3. Прочитать аннотацию, сравнить с тем, как интерпретировали для себя дашборд.
4. Оценить представленную аннотацию.
5. Описать расхождения или другие комментарии в последнем пункте.

Вопросы предполагались следующие:

– Насколько аннотация помогает понять содержание дашборда или предоставляет ценную информацию? Выберите ответ по шкале от 1 до 10 (1 – Аннотация НЕ содержит полезной информации для понимания графика; 10 – Аннотация полностью раскрывает суть дашборда, помогая быстрее разобраться).

– Насколько аннотация понятна и легко читаема? Выберите ответ по шкале от 1 до 10 (1 – Аннотация очень запутанная, тяжело читать; 10 – Аннотация ясная, легко читается).

– Насколько аннотация соответствует контексту представленного дашборда? Выберите ответ по шкале от 1 до 10 (1 – Аннотация НЕ соответствует контексту; 10 – Аннотация полностью соответствует контексту).

– Насколько аннотация соответствует нормам языка (грамматика, орфография, пунктуация)? Выберите ответ по шкале от 1 до 10 (1 – Аннотация содержит ошибки, текст неграмотный; 10 – Аннотация без ошибок, грамотно составлена).

– Есть какие-то дополнительные комментарии? Что бы вы добавили в аннотацию? Каких показателей не хватает? Если все устраивает, напишите об этом.

Также аннотации оценивались по формуле индекса удобочитаемости Флеша [11]. Такая метрика оценивает сложность текста на английском языке по следующей формуле:

$$FRE = 206,835 - 1,015 \cdot ASL - 84,6 \cdot ASW,$$

где ASL – средняя длина предложения в словах (англ. average sentence length), ASW – средняя длина слова в слогах (англ. average number of syllables per word).

Индекс по шкале FRES распределяется следующим образом: 100 – очень легко читается, 65 – достаточно простой язык, который несложно читать, 30 – немного трудно читать, 0 – очень трудно читать.

Оценка аннотаций на русском языке была проведена¹ с помощью скорректированной формулы, так как в русском языке средняя длина предложения меньше, а слова длиннее. Скорректированная формула представлена ниже:

$$FRE = 206,835 - 1,52 \cdot ASL - 65,14 \cdot ASW$$

Результаты и обсуждение

Приведена ссылка на открытые результаты опроса². Далее представлены сгенерированные аннотации и ссылки на источники дашбордов и данных:

1. Детская смертность³.

Дашборд показывает показатель детской смертности в Бразилии, демонстрируя нисходящий тренд без выраженной сезонности. Максимальное значение составляет 26,4% в 1934 году, минимальное – 1,4% в 2023 году, что показывает заметное снижение показателя за 90 лет. Аномалии не обнаружены.

2. Курс золота в рублях⁴.

Дашборд посвящен финансам и отображает данные по золоту (Банк России) с восходящим трендом и выраженной сезонностью. Минимальное значение составляет 7851,85 на 2025-03-19, максимальное значение 9303,76 на 2024-11-29. К аномалиям можно отнести повторяющееся значение 8551,74 с 2024-12-29 по 2025-01-09, что может быть связано с отсутствием обновления данных в течение новогодних праздников.

3. Уровень безработицы среди женщин⁵.

Дашборд в области экономики показывает уровень безработицы среди женщин с нисходящим трендом с неявно выраженной сезонностью. Максимальное значение составляет 16,2% на 1 апреля 2020 года, минимальное – 2,7% на 1 мая 1953. С мая по апрель 2020 года обнаружен резкий скачок с 4,4% до 16,2%. Скорее всего, это связано с пандемией COVID-19, когда массовые локдауны и закрытие предприятий привели к потере рабочих мест. Женщины, часто занятые в сферах услуг, розничной торговли и образования, пострадали особенно сильно.

4. Индекс трат⁶.

Дашборд в финансовой области показывает индекс Visa Spending Momentum Index: Non-discretionary с нисходящим трендом и выраженной сезонностью.

¹ Проверка читаемости текста. INtexty. URL: <https://intexty.com/chitabelnost-teksta/> (дата обращения: 20.05.2025).

² Результаты опроса. Google Таблицы. URL: <https://clck.ru/3MRKUa> (дата обращения: 20.05.2025).

³ Child mortality rate, 1934 to 2023. Our World in Data. URL: <https://ourworldindata.org/grapher/child-mortality?country=~BRA> (дата обращения: 20.05.2025).

⁴ Золото (Банк России) (RUB). InvestFunds. URL: <https://investfunds.ru/indexes/224/> (дата обращения: 20.05.2025).

⁵ Unemployment Rate – Women. Federal Reserve Economic Data | FRED | St. Louis Fed. URL: <https://fred.stlouisfed.org/series/LNS14000002> (дата обращения: 20.05.2025).

⁶ Visa Spending Momentum Index: Non-discretionary: United States (VISASMINNSA). Federal Reserve Economic Data | FRED | St. Louis Fed. URL: <https://fred.stlouisfed.org/series/VISASMINNSA#> (дата обращения: 20.05.2025).

Максимальное значение составляет 108, зафиксированное 1 апреля 2021 года, минимальное – 94, зарегистрированное 1 апреля 2020 года. Отмечен резкий спад показателя в апреле 2020 года, связанный с пандемией COVID-19 и локдаунами, которые сократили потребительские расходы. Пик в январе 2021, возможно, связан с постпандемийным восстановлением и сезонными факторами.

5. Импорт товаров в США⁷.

Дашборд отображает импорт товаров в США с восходящим трендом и выраженной сезонностью. Максимальное значение составляет 360,000 на 2023-01-01, минимальное – 40,000 на 1992-01-01. В 2008–2009 годы наблюдается резкий спад – результат глобального финансового кризиса, сократившего спрос. В 2021–2025 годы видно быстрое восстановление и рост, особенно в 2025 году (344,272 млн долларов в марте), что может быть связано с восстановлением экономики и ростом цен на товары.

6. Индекс экономической активности⁸.

Дашборд отображает Индекс глобальной реальной экономической активности с 1968 по 2025 годы с восходящим трендом и неявной сезонностью. Максимальное значение индекса это 189,8 на 2008-05-01, минимальное значение составляет –160,7 на 2016-02-01. Резкий спад с 188,1 в ноябре 2007 до –75,3 в декабре 2008 – результат глобального финансового кризиса. В феврале 2020 наблюдается спад до –120,8 из-за пандемии COVID-19, нарушившей глобальные цепочки поставок. Пик в октябре 2021 отражает восстановление после пандемии.

7. Потребление калорий⁹.

Дашборд показывает ежедневное потребление калорий на душу населения в России с восходящим трендом и без выраженной сезонности. Максимальное значение составляет 3433 ккал в 2022 году, минимальное – 2792 ккал в 1992 году. Спад в 1996 связан с экономическим кризисом и дефолтом в России, который сказывался на продовольственном рынке. Максимальное значение может быть связано с ростом внутреннего производства и адаптацией к санкциям.

8. Смертность от сердечно-сосудистых заболеваний¹⁰.

Дашборд посвящен анализу смертности от сердечно-сосудистых заболеваний в России с 1980 по 2019 годы, демонстрируя нисходящий тренд без выраженной сезонности. Максимальное значение составило 655 в 2003 году, минимальное – 314 в 2019 году. Пик в 1990-х связан с экономическим кризисом, снижением доступности медицинских услуг и ростом стрессовых факторов. При этом абсолютный максимум наблюдается в 2003 году, что может быть связано с накопленным эффектом плохого питания, алкоголизма и недостаточного медицинского обслуживания. Ускоренное снижение с 464,5 в 2011 году до 314,4 в 2019 году может быть связано с государственными программами по борьбе с сердечно-сосудистыми заболеваниями и общим улучшением жизни населения.

9. Оценка запасов ядерных боеголовок¹¹.

Дашборд показывает оценку запасов ядерных боеголовок в мире с нисходящим трендом без выраженной сезонности. Максимальное значение составляет 70374 в 1986

⁷ U.S. Imports of Goods by Customs Basis from World (IMP0004). Federal Reserve Economic Data | FRED | St. Louis Fed. URL: <https://fred.stlouisfed.org/series/IMP0004#> (дата обращения: 20.05.2025).

⁸ Index of Global Real Economic Activity (IGREA). Federal Reserve Economic Data | FRED | St. Louis Fed. URL: <https://fred.stlouisfed.org/series/IGREA#> (дата обращения: 20.05.2025).

⁹ Daily supply of calories per person, 1992 to 2022. Our World in Data. URL: <https://ourworldindata.org/grapher/daily-per-capita-caloric-supply?time=1910..latest&country=~RUS> (дата обращения: 20.05.2025).

¹⁰ Death rate from cardiovascular diseases, 1980 to 2019. Our World in Data. URL: <https://ourworldindata.org/grapher/cardiovascular-disease-death-rate-who-mdb?country=~RUS> (дата обращения: 20.05.2025).

¹¹ Estimated nuclear warhead stockpiles. Our World in Data. URL: https://ourworldindata.org/grapher/nuclear-warhead-stockpiles-lines?country=~OWID_WRL (дата обращения: 20.05.2025).

году, минимальное значение 6 в 1945 году. Резкий спад до 3,092 боеголовок в 1962 году – возможная ошибка в данных или временное снижение из-за Карибского кризиса и переоценки запасов. Рост запасов до 70,374 в 1986 году обусловлен гонкой вооружений между США и СССР. Снижение после 1986 года связано с договорами СНВ-1 (1991) и СНВ-2, а также разрядкой напряженности.

10. Рейтинг Эло, достигнутый компьютером в шахматах¹².

Дашборд отображает Ело рейтинг, достигнутый компьютерами в шахматах с 1985 по 2023 годы с восходящим трендом и без выраженной сезонности. Максимальное значение составляет 3591 в 2022 и 2023 годах, а минимальное – 2016 в 1985 году. Рост рейтинга обусловлен улучшением алгоритмов поиска и увеличением вычислительной мощности.

В Таблице 1 отображены результаты оценки аннотаций по четырем метрикам в ходе опроса: понятность, читаемость, соответствие контексту, грамотность, а также метрика индекса удобочитаемости Флеша (Flesch-Kincaid).

Таблица 1 – Результаты оценки аннотаций после опроса

Table 1 – Results of the evaluation of annotations after the survey

Дашборд	Понятность	Читаемость	Контекст	Грамотность	Индекс Флеша
1. Детская смертность	8,238	9,429	9,000	8,952	18
2. Курс золота в рублях	7,619	7,762	8,238	8,619	27
3. Уровень безработицы	8,524	8,619	8,667	8,762	38
4. Индекс трат	8,476	8,238	9,000	8,381	25
5. Импорт товаров в США	8,571	7,857	8,905	7,286	43
6. Экономическая активность	8,286	8,667	9,095	8,952	42
7. Потребление калорий	8,762	9,238	8,571	9,286	37
8. Смертность от сердечных заболеваний	8,714	9,048	9,048	8,857	29
9. Оценка запасов ядерных боеголовок	9,000	9,143	9,190	9,238	44
10. Рейтинг Эло у компьютера	8,667	9,048	9,190	8,905	33
Итого	8,486	8,705	8,890	8,724	33,6

По результатам видно, что пользователей устраивает уровень понятности, читаемости, соответствия контексту и грамотности. Однако было много замечаний, связанных с нечитаемым форматом дат или с некоторым недостатком информации, которую пользователи хотели бы видеть. В дальнейшем исследовании такие замечания будут учтены, также будет добавлено итеративное взаимодействие с системой, в ходе которого будет возможно получить недостающую информацию путем написания вопросов системе в чат. Метрика индекса удобочитаемости Флеша получилась достаточно низкой – пользователям будет трудно читать текст. Скорее, этот показатель больше связан со спецификой областей дашбордов и данных в целом. Так как в формуле используются однообразные параметры текста (средняя длина слова, средняя длина предложения), такой индекс не адаптирован к текстам для взрослых людей – у начитанного человека и профессионалов не должно возникать затруднений с пониманием многосложных слов.

¹² Highest chess rating ever achieved by computers. Our World in Data. URL: <https://ourworldindata.org/grapher/computer-chess-ability> (дата обращения: 20.05.2025).

Заключение

Проведенное исследование продемонстрировало эффективность разработанной мультиагентной системы для автоматической генерации аннотаций на естественном языке к дашбордам с временными рядами. Средние оценки по метрикам понятности (8,486), читаемости (8,705), соответствия контексту (8,890) и грамотности (8,724) свидетельствуют о том, что сгенерированные аннотации являются полезным инструментом для упрощения анализа данных и выявления ключевых закономерностей.

В дальнейшем развитии системы планируется оптимизация формата представления данных, улучшение читаемости текстов с учетом специфики аудитории и расширение возможностей взаимодействия с пользователем. Такой подход позволит повысить практическую ценность системы для аналитиков, ускорить процесс принятия бизнес-решений и обеспечить более точную интерпретацию сложных временных рядов.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Gkatzia D., Lemon O., Rieser V. Natural Language Generation Enhances Human Decision-Making with Uncertain Information. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016: Volume 2: Short Papers, 07–12 August 2016, Berlin, Germany*. Association for Computational Linguistics; 2016. P. 264–268. <https://doi.org/10.18653/v1/P16-2043>
2. Jiang Yu., Pan Z., Zhang X., et al. Empowering Time Series Analysis with Large Language Models: A Survey. In: *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence Survey Track, IJCAI 2024, 03–09 August 2024, Jeju, South Korea*. 2024. P. 8095–8103. <https://doi.org/10.24963/ijcai.2024/895>
3. Tang F., Ding Yi. Are Large Language Models Useful for Time Series Data Analysis? arXiv. URL: <https://doi.org/10.48550/arXiv.2412.12219> [Accessed 15th May 2025].
4. Jin M., Zhang Yi., Chen W., et al. Position: What Can Large Language Models Tell Us About Time Series Analysis. In: *Proceedings of the 41st International Conference on Machine Learning, ICML 2024, 21–27 July 2024, Vienna, Austria*. 2024. <https://doi.org/10.48550/arXiv.2402.02713>
5. Lin M., Chen Zh., Liu Ya., et al. Decoding Time Series with LLMs: A Multi-Agent Framework for Cross-Domain Annotation. arXiv. URL: <https://doi.org/10.48550/arXiv.2410.17462> [Accessed 15th May 2025].
6. Sycara K.P. Multiagent Systems. *AI Magazine*. 1998;19(2):79. <https://doi.org/10.1609/aimag.v19i2.1370>
7. Ghysels E., Osborn D.R. *The Econometric Analysis of Seasonal Time Series*. New York: Cambridge University Press; 2001. 228 p.
8. Jebb A.T., Tay L., Wang W., Huang Q. Time Series Analysis for Psychological Research: Examining and Forecasting Change. *Frontiers in Psychology*. 2015;6. <https://doi.org/10.3389/fpsyg.2015.00727>
9. Darban Z.Z., Webb G.I., Pan Sh., Aggarwal Ch., Salehi M. Deep Learning for Time Series Anomaly Detection: A Survey. *ACM Computing Surveys*. 2024;57(1). <https://doi.org/10.1145/3691338>
10. Orgad H., Toker M., Gekhman Z., et al. LLMs Know More Than They Show: On the Intrinsic Representation of LLM Hallucinations. In: *The Thirteenth International Conference on Learning Representations, ICLR 2025, 24–28 April 2025, Singapore*. 2025. <https://doi.org/10.48550/arXiv.2410.02707>
11. Štajner S., Evans R., Orasan C., Mitkov R. What Can Readability Measures Really Tell Us About Text Complexity? In: *Proceedings of the Eighth International Conference on*

Language Resources and Evaluation, LREC 2012, 23–25 May 2012, Istanbul, Turkey.
European Language Resources Association (ELRA); 2012. P. 14–21.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Кузнецова Анна Игоревна, студентка, **Anna I. Kuznetsova**, Student, Volgograd State
Волгоградский государственный технический университет, Волгоград, Российская Федерация. Technical University, Volgograd, the Russian
Federation.
e-mail: anniekuznec@mail.ru
ORCID: [0009-0003-5767-6810](https://orcid.org/0009-0003-5767-6810)

Носкин Виктор Викторович, аспирант, **Victor V. Noskin**, Postgraduate, Volgograd
Волгоградский государственный технический университет, Волгоград, Российская Федерация. State Technical University, Volgograd, the
Russian Federation.
e-mail: vitek2012rus@gmail.com
ORCID: [0000-0002-0795-2675](https://orcid.org/0000-0002-0795-2675)

*Статья поступила в редакцию 28.05.2025; одобрена после рецензирования 26.06.2025;
принята к публикации 01.07.2025.*

*The article was submitted 28.05.2025; approved after reviewing 26.06.2025;
accepted for publication 01.07.2025.*