



## АВТОМАТИЗАЦИЯ ТЕСТИРОВАНИЯ ОБУЧАЮЩИХСЯ СРЕДСТВАМИ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И СОВРЕМЕННЫХ ИНТЕГРАЦИОННЫХ ПЛАТФОРМ

**Д. В. Власов**

*Российский государственный педагогический университет имени А.И. Герцена, Россия*

**Аннотация.** Исследование посвящено разработке интеллектуальной системы автоматизации тестирования на основе генеративных языковых моделей и интеграционных платформ. Проблема масштабируемости оценки знаний в массовом образовании решается через комбинацию автоматизированной проверки и многоуровневого анализа аномалий. Система реализована на платформе Albato с использованием Yandex.Forms, Google Sheets и API больших языковых моделей. Новизна заключается в двухступенчатом анализе аномалий: жёстком критериальном и интеллектуальном LLM анализе. Практическая значимость подтверждается готовым к внедрению прототипом.

**Ключевые слова.** Автоматизация тестирования; генеративные модели; анализ аномалий; интеграционные платформы; образовательная аналитика; искусственный интеллект; оценка знаний.

## AUTOMATION OF STUDENT TESTING USING GENERATIVE ARTIFICIAL INTELLIGENCE AND MODERN INTEGRATION PLATFORMS

**D. V. Vlasov**

*Herzen State Pedagogical University of Russia, Russia*

**Annotation.** The study focuses on developing an intelligent testing automation system based on generative language models and integration platforms. The scalability problem of knowledge assessment in mass education is solved through a combination of automated verification and multilevel anomaly analysis. The system is implemented on the Albato platform using Yandex.Forms, Google Sheets, and large language model APIs. The novelty

lies in two stage anomaly analysis: strict criteria based and intelligent LLM analysis. Practical significance is confirmed by a ready to implement prototype.

**Keywords.** Testing automation; generative models; anomaly analysis; integration platforms; educational analytics; artificial intelligence; knowledge assessment.

## Введение

Генеративные языковые модели трансформируют образовательные процессы через автоматизацию создания учебных материалов и мониторинг успеваемости студентов. Это позволяет сокращать время преподавателей на разнообразные рутинные задачи, а также обеспечивает студентам обратную связь с минимальной задержкой. Практика такого рода способна повышать эффективность обучения по сравнению с традиционными методами.

Оценка учебных достижений в массовом образовании сталкивается с труднопреодолимой проблемой масштабируемости [1]. К примеру, ручная проверка тестов требует от преподавателя значительных временных затрат — при нагрузке в десятки и сотни студентов в сочетании с их массовыми тестированиями создается критическая, а иногда и невыполнимая нагрузка. Активно развивающееся массовое онлайн-обучение еще больше усугубляет проблему, поскольку преподаватели обрабатывают результаты сотен, а то и тысяч студентов без возможности обеспечить единообразие критериев оценки [1]. При этом неизбежные субъективные искажения влияют на итоговые баллы, создавая разброс оценок для идентичных или похожих работ.

Одним из выходов в данной ситуации являются интеграционные платформы, которые автоматизируют бизнес-процессы, связывая разнородные сервисы — системы управления обучением, онлайн-формы, генеративные ИИ-модели, облачные хранилища, почтовые сервисы, мессенджеры — в единые автоматизированные сценарии без необходимости профессионального программирования таких систем [2]. В образовании их потенциал недооценен, хотя комбинация генеративного искусственного интеллекта (ИИ) и интеграционных платформ создает новое поколение систем тестирования, которые под силу создать и настроить чуть ли не каждому отдельному преподавателю. Эти системы могут автоматизировать рутинные операции и проводить глубокий анализ результатов их выполнения.

Нынешнее «цифровое поколение» студентов ожидает быстрой объективной обратной связи [3] по результатам своей ежедневной учебной деятельности. Любые

задержки в несколько дней снижают педагогическую ценность каких-либо комментариев преподавателей — их обратной связи, поскольку контекст выполненного задания теряет актуальность в памяти студента. Автоматизация такой обратной связи становится критически важной.

В то же время, одна из ключевых проблем автоматизированного тестирования — это выявление аномальных результатов. Стандартные системы автопроверки часто ограничиваются подсчётом правильных и неправильных ответов, игнорируя контекстуальные аспекты [4]. Подозрительно высокий результат может указывать на исключительные способности студента или на использование недозволенных методов. Аномально низкий результат, в свою очередь, может сигнализировать о пробелах в знаниях, системных проблемах в понимании материала или об отсутствии мотивации.

**Цель исследования** — разработать концептуальную модель и практическую реализацию интеллектуальной системы автоматизации тестирования, основанной на комбинации генеративного ИИ, интеграционных платформ и многоуровневого анализа аномалий. Указанная система должна обеспечивать автоматическую оценку с высокой точностью и объективностью по сравнению с ручной экспертной оценкой, а также предоставлять преподавателю (эксперту) аналитические инструменты для выявления нетипичных результатов.

### **Методология исследования / теоретические основы**

Исследование использует междисциплинарный подход, объединяющий методы простой программной инженерии, базового анализа данных и педагогического проектирования. Выбранная ключевая методологическая установка — это концепция «сценарной демонстрации», при которой теоретические положения подкрепляются функционирующим технологическим прототипом.

Центральный метод исследования — прототипирование на базе российской интеграционной платформы Albato. данный выбор обусловлен наличием русскоязычного интерфейса, обширной библиотеки «коннекторов» к российским сервисам и довольно гибкого визуального конструктора сценариев без программирования или с базовым программированием небольших фрагментов кода на языке JavaScript. Предлагается спроектированная и реализованная интеграционная связка, обеспечивающая полный цикл обработки ответов на тестовые задания.

Для оценки ответов студентов применяется интеллектуальный анализ с использованием больших языковых моделей (LLM — large language models).

Апробированы модели AI-ассистентов от OpenAI (ChatGPT) и Sber GigaChat. Данный выбор основан как на фактической эффективности в задачах обработки естественного языка для образовательного контента [5; 6], так и на необходимости использования современных российских программных продуктов. Используются официальные API (OpenAI и Gigachat) с запросами, специально сконструированными для педагогической оценки в JSON-формате.

**Архитектурный принцип системы:** одна специализированная под рассматриваемую задачу интеграция/связка в Albato, записывающая данные ответов студентов на тест в таблицу Google Sheets как единственную "шину данных" (рис. 1) для возможного последующего анализа и обработки преподавателем.



*Рис. 1. Общая схема функционирования системы автоматизации тестирования*

*Fig. 1. General scheme of the testing automation system operation*

1. **Критериальный анализ:** оценка от LLM-ассистента (ИИ-ассистента) сравнивается с пороговыми значениями (ниже 20 и выше 85 баллов по 100-балльной шкале). Такой метод выявляет явные случаи крайне низких и высоких результатов.
2. **Интеллектуальный LLM-анализ:** генеративная модель получает ответы студента и результат предыдущей пороговой оценки аномалий, принимает финальное

решение о наличии аномалии и формирует еще один отдельный текстовый комментарий с объяснением по поводу возможных нетипичных результатов студента.

Для сбора данных в представляемом решении используются веб-форма, созданная в сервисе Yandex.Forms, а для хранения и обработки — электронные таблицы Google Sheets (на момент написания статьи в Albato нет поддержки Yandex-таблиц). Такой достаточно простой подход обеспечивает гибкость создания тестовых заданий и удобный доступ к данным для автоматизированной системы и преподавателя.

## Результаты и их обсуждение

**Архитектура и реализация системы.** Разработана гибкая настраиваемая связка независимых облачных сервисов в рамках единого небольшого технологического контура. Жизненный цикл обработки данных при этом представляет из себя определенную последовательность шагов такой интеграционной связки, которая представляет из себя отдельный и наиболее важный автоматизированный подпроцесс в показанной на рис. 1 общей схеме сбора и проверки ответов.

**Подпроцесс "Автоматическая проверка и анализ аномалий"** запускается автоматически каждый раз при отправке студентом некоторой заполненной Yandex-формы с вопросами тесты (рис. 2). Система на базе платформы Albato получает данные от каждого студента через специальный настраиваемый в Yandex-форме JSON-RPC webhook (случайный URL-адрес), который генерируется сервисом Albato. Созданная связка Albato отправляет полученные данные Yandex-формы на проверку LLM-ассистенту, который проводит первичное оценивание с выставлением отметки по 100-балльной шкале и с генерацией текстового комментария. Затем, как видно из диаграммы на рис. 2, запускается специальный подпроцесс обнаружения аномалий, где, помимо тривиального обнаружения низких и высоких баллов по простым критериям, уже совершенной другой специализированный LLM-ассистент проводит анализ аномальности ответов, с последующей дополнительной записью результатов «единий JSON-документ», который выполняет роль единого контекста в рассматриваемой интеграционной связке.

На рис. 3 представлен скриншот практической реализации всей автоматизации. Описание шагов связки:

1. *Шаг 1:* получение данных из веб-формы через триггер «Яндекс.Форма – Тест».

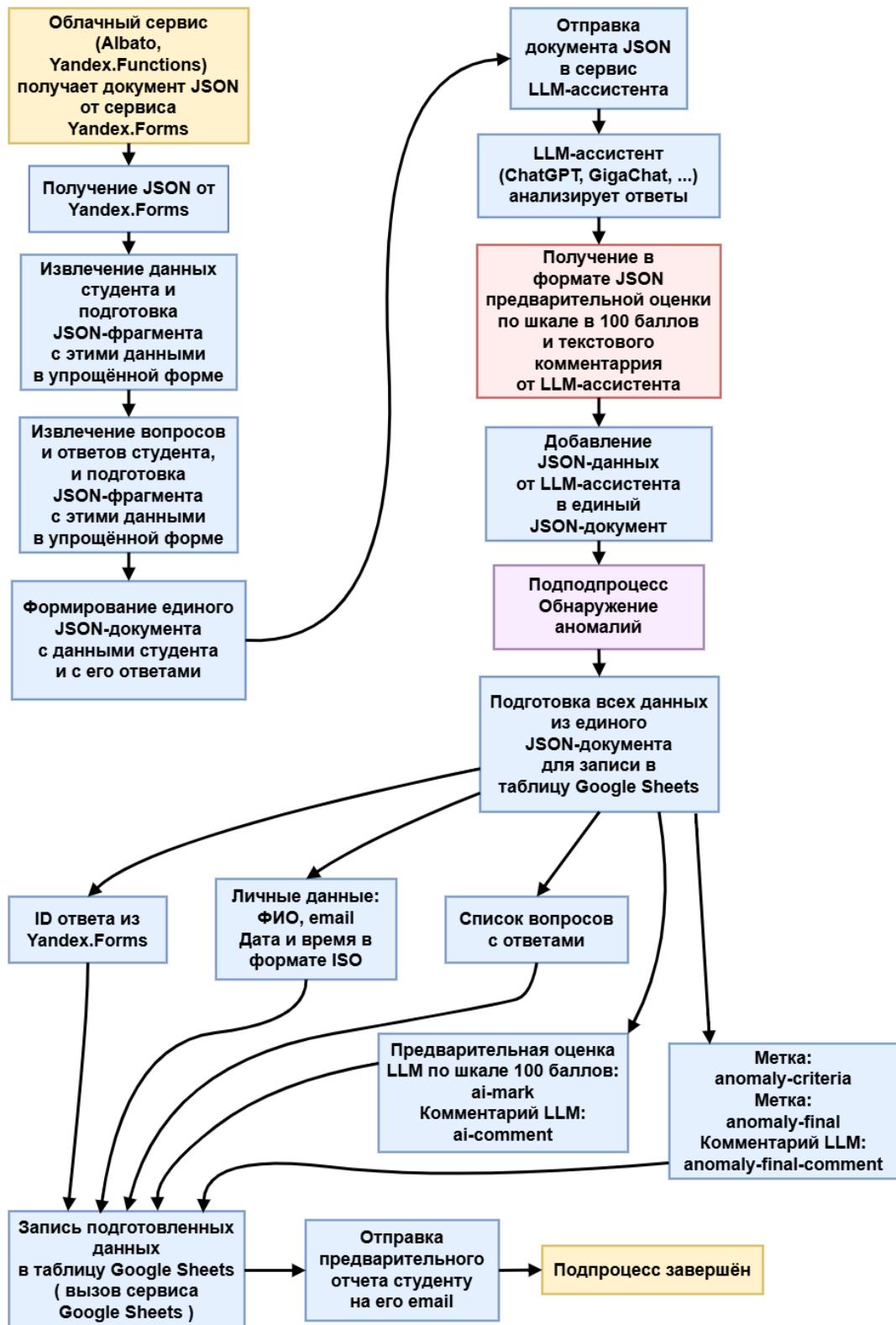
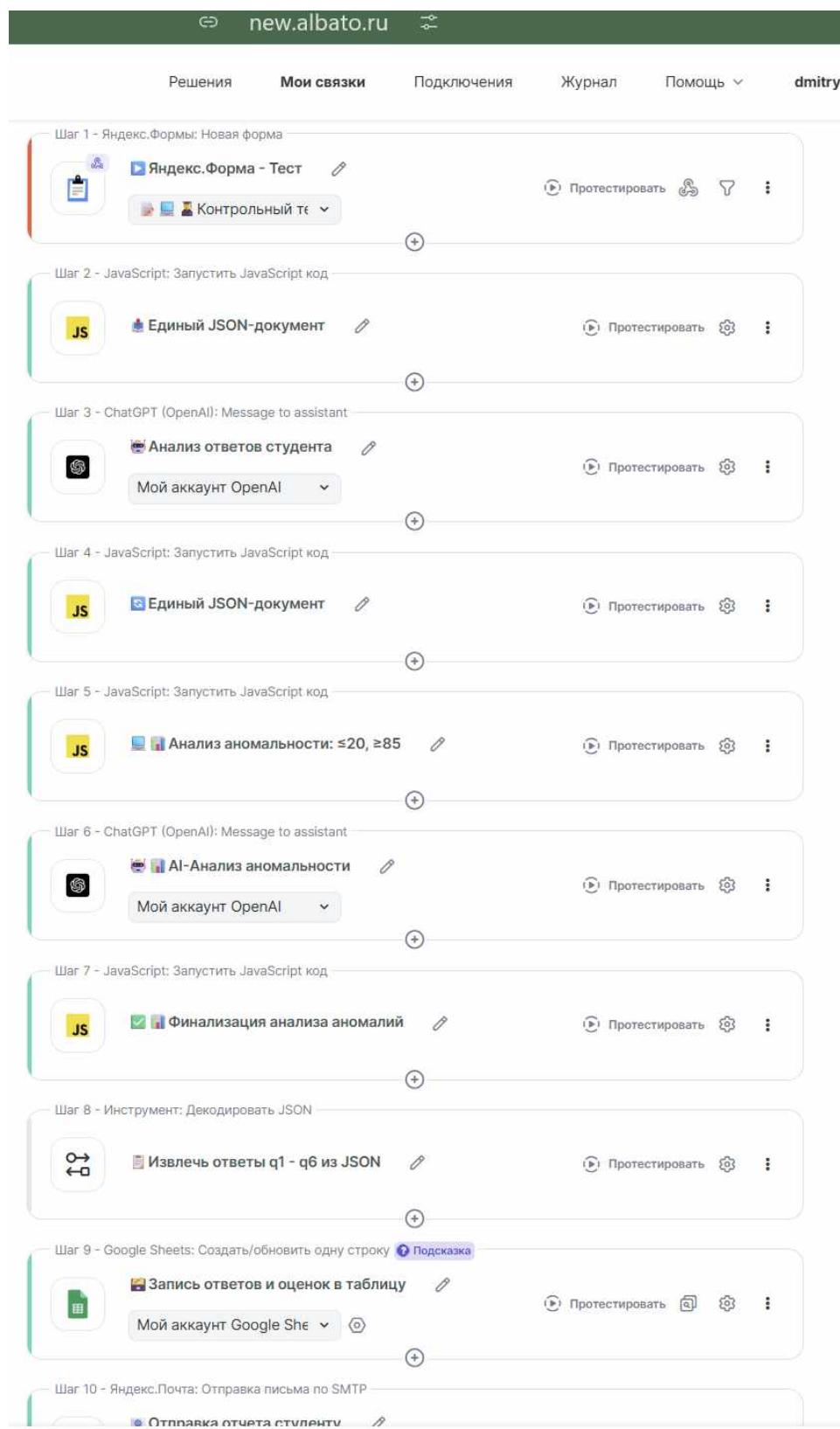


Рис. 2. Подпроцесс автоматической проверки ответов и анализа аномалий  
 Fig. 2. Subprocess of automatic answer verification and anomaly analysis

2. Шаг 2: первичная обработка отдельным JavaScript-модулем, в котором JSON-данные от Yandex.Forms преобразуются в тот самый более простой по структуре и более компактный «единый JSON-документ».



*Rus. 3. Скриншот связки Albato: интеграция автоматической проверки (шаги 1–10)*  
*Fig. 3. A screenshot of an Albato integration: automatic checking integration (steps 1–10)*

3. *Шаг 3*: оценка ответов в шаге «Анализ ответов студента» через API выбранного LLM-ассистента (OpenAI/ChatGPT или Gigachat) с промптом для педагогической оценки (см. примеры настройки этого типа узла на рис. 4 и рис. 5). На рис. 5 и рис. 6 (Gigachat) показан пример формирования по промпту (запросу) и контексту обычного текстового ответа. Такой промпт, разумеется, можно самостоятельно переделать и "попросить" Gigachat прислать именно JSON-документ нужного формата. Важным отличием OpenAI-ассистента от Gigachat является то, что контекст (системный "настроенный" промпт LLM-ассистента) можно записать в панели управления на сайте OpenAI (см. рис. 7), а затем дополнительно приложить необходимы проверочные материалы теста в векторную базу данных. В случае с Gigachat материалы теста приходится складывать в поле контекста (поле системного промпта) прямо в Albato.

4. *Шаг 4*: вторичная обработка JavaScript-модулем «единого JSON-документа» (контекста связки), в который добавляется отметка (до 100-ти баллов) и комментарий от LLM-ассистента об ответах студента.

5. *Шаг 5*: простой критериальный анализ аномалий JavaScript-модулем «Анализ аномальности:  $\leq 20$ ,  $\geq 85$ » с добавлением булевой метки `anomaly_criteria` в «единый JSON-документ». Более подробно процесс анализа аномалий рассмотрен ниже в разделе «Двухуровневый анализ аномалий».

6. *Шаг 6*: LLM-анализ аномалий с помощью второго специализированного LLM-ассистента от OpenAI (также можно заменить на Gigachat). LLM-ассистент принимает во внимание значение метки `anomaly_criteria` из шага 5 благодаря правильно написанному для него промпту и отдает свой ответ в виде небольшого JSON-документа с другой булевой меткой `anomaly_final` и своим комментарием `anomaly_final_comment`.

7. *Шаг 7*: финализация анализа аномалий отдельным JavaScript-модулем с еще одним дозаполнением «единого JSON-документа» меткой `anomaly_final` и комментарием `anomaly_final_comment`.

8. *Шаг 8*: извлечение ответов на вопросы инструментом типа «Декодировать JSON» для записи в таблицу Google Sheets.

9. *Шаг 9*: запись личных данных студента, ответов на тест и всех результатов проверки на аномалии в таблицу Google Sheets для проверки преподавателем.

10. *Шаг 10*: отправка отчета студенту с его отметкой и комментариями от первого LLM-ассистента через Yandex.Почту.

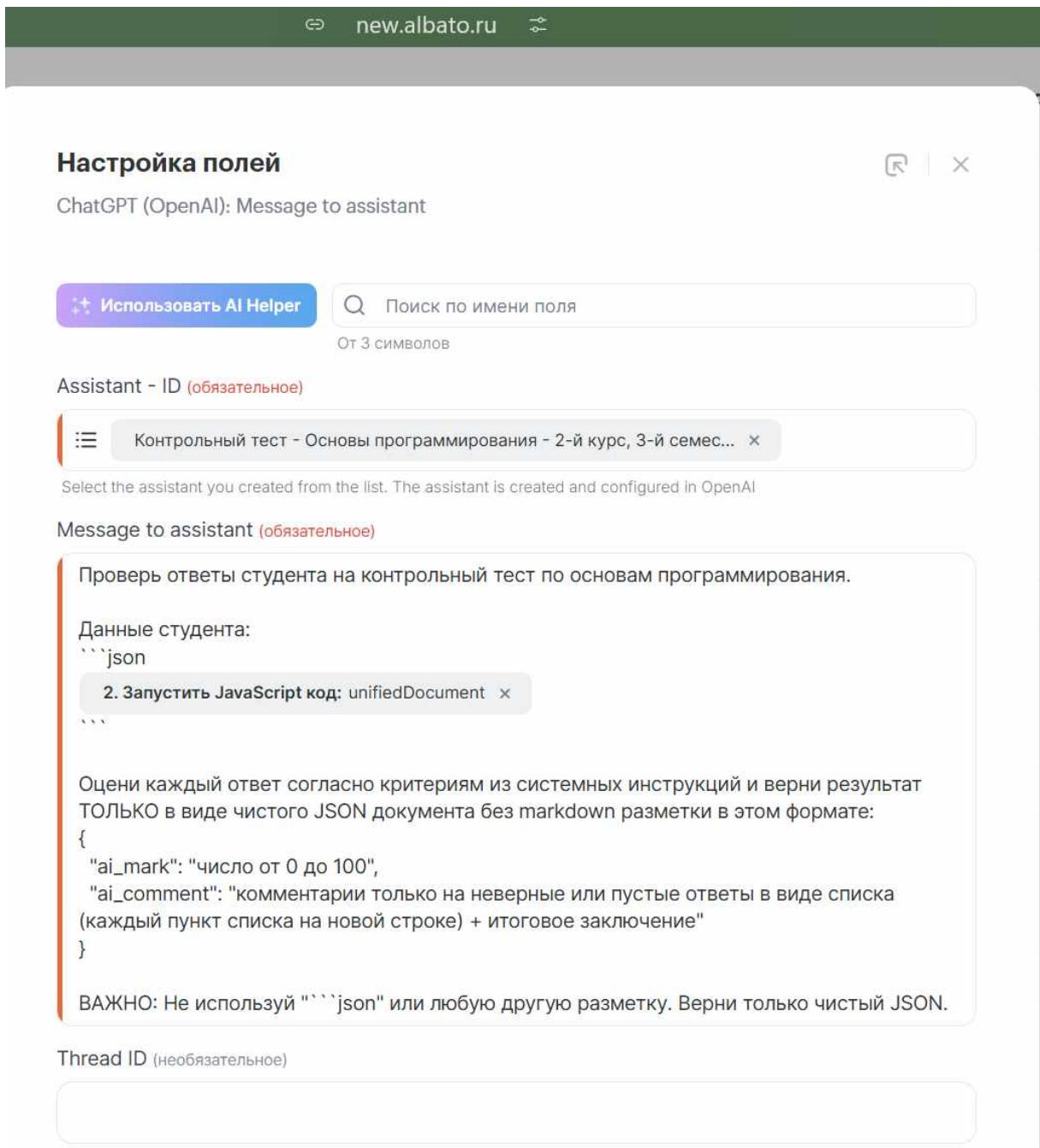


Рис. 4. Узел OpenAI в Albato: модель и параметры JSON-ответа

Fig. 4. OpenAI node in Albato: model and JSON response settings

Двухуровневый анализ аномалий. Механизм анализа аномалий переводит систему от примитивной дихотомической оценки «верно/неверно» к более глубокому контекстуальному пониманию результатов каждого студента (рис. 8). Система не заменяет преподавателя, а дополняет его, выступая в роли интеллектуального ассистента.

new.albato.ru

## Настройка полей

ГигаЧат: Создать запрос к модели с контекстом

Использовать AI Helper

Поиск по имени поля

От 3 символов

Название модели - ID (обязательное)

GigaChat-Pro

Для выбора последней актуальной модели используйте GigaChat:latest

Контекст (обязательное)

Контекст:

Роль: преподаватель математики, проводивший обзорную лекцию по фракталам.  
Внимание: комментируй только неверные ответы вопросы! Ответ студенту должен быть в формате обычного текста! Не вставляй в ответ данные в формате JSON! Если ответ на вопрос верный - не пиши на него комментарий. Верными ответами считай только те, которые тебе даны в тексте теста из контекста.

```
```markdown
# Тест. Фракталы в природе и их применение

## Q1 - один вариант ответа: Какие природные объекты являются примерами фракталов?

Варианты:
A. Листья деревьев
B. Молнии
C. Береговые линии
D. Все вышеперечисленное

Правильный ответ: D. Все вышеперечисленные объекты демонстрируют свойства самоподобия, характерные для фрактальных структур.

```

Рис. 5. Пример узла Gigachat в Albato: модель и параметры ответа

Fig. 5. An example of a Gigachat node in Albato: model and response settings

вопросы! Ответ студенту должен быть в формате обычного текста! Не вставляй в ответ данные в формате JSON! Если ответ на вопрос верный - не пиши на него комментарий. Верными ответами считай только те, которые тебе даны в тексте теста из контекста.

Указать контекст, например - "Отвечай как научный сотрудник"

Запрос пользователя (обязательное)

Инструкция: проанализировать ответы студента по тесту «Фракталы в природе и их применение». Определить, какие из ответов неверны, и составить перечень с краткими комментариями ошибок. Итоговый комментарий от преподавателя должен содержать 2-3 предложения, обобщающих результат и предлагающих рекомендации по доработке материала. Внимание: комментируй только неверные ответы вопросы! Если ответ на вопрос верный - не пиши на него комментарий. Верными ответами считай только те, которые тебе даны в тексте теста из контекста.

Ответ студента на тест:

```json

4. Запустить JavaScript код: jsonParsedAnswersAsString

Образец твоего ответа с результатами проверки теста:

```

Список неверных ответов (если есть):

- Вопрос QN - неверно: причина неверного ответа;
- Вопрос QM - неверно: причина неверного ответа;
- ...;

Комментарий преподавателя: короткий итоговый комментарий от лица преподавателя.

```

Убедиться, что сгенерированный тобой ответ строго структурирован согласно образцу, содержит список комментариев на неправильные ответы и итоговый комментарий

*Rис. 6. Продолжение рис. 5: фрагмент запроса пользователя в Gigachat*

*Fig. 6. Continuation of Fig. 5: Fragment of a user request in Gigachat*

**Критериальный анализ** основан на жёстко заданных пороговых значениях: нижний порог 20 баллов, верхний порог 85 баллов по 100-балльной шкале. Результат записывается в колонку *anomaly-criteria* в виде булева значения. Пороговые значения могут пересматриваться в зависимости от специфики курса и уровня подготовки группы.

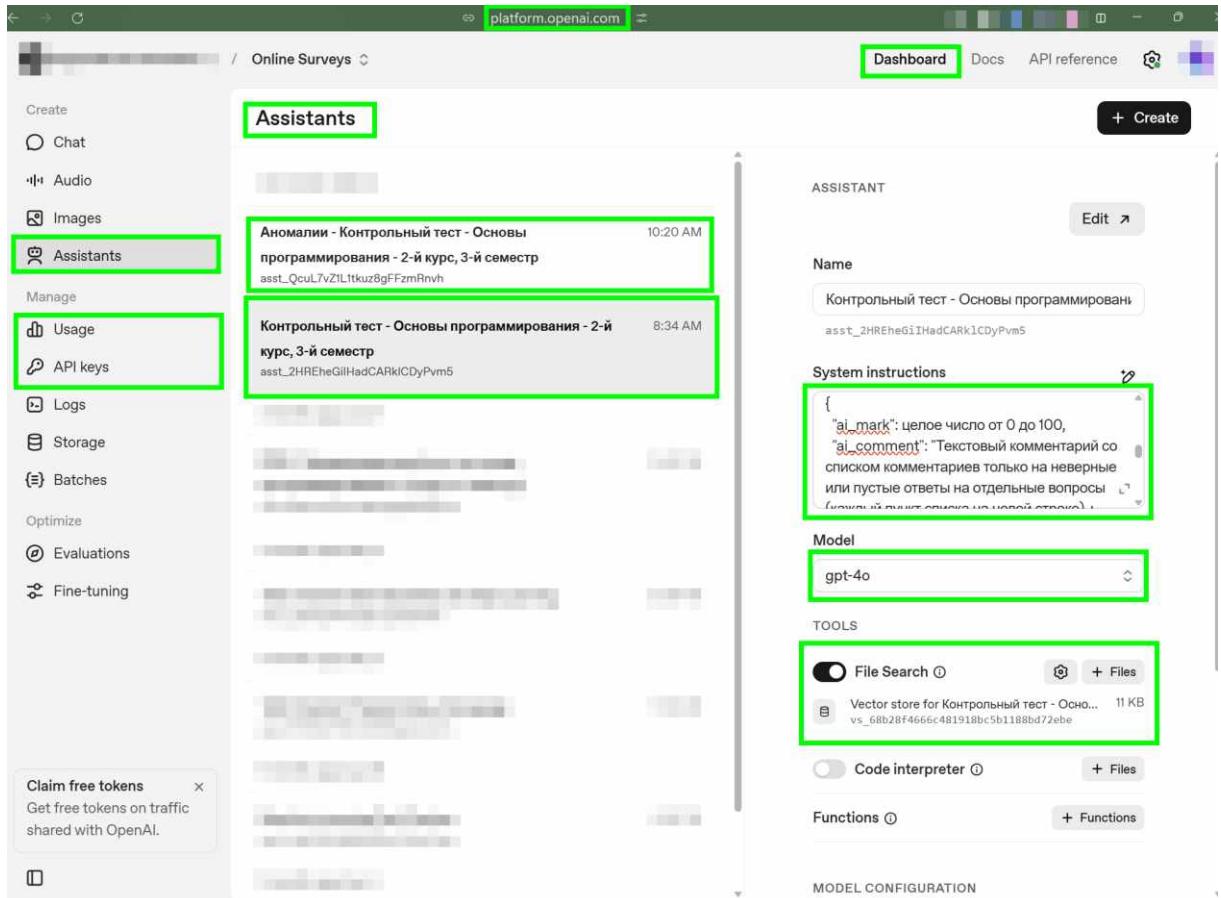


Рис. 7. Панель управления LLM-ассистентов OpenAI

Fig. 7. OpenAI LLM-assistant control panel

**Интеллектуальный LLM-анализ** представляет здесь высший уровень анализа. Большая языковая модель получает всю собранную информацию: ответы студента, предварительную оценку и булеву метку *anomaly-criteria*. На основе такого комплексного набора данных модель принимает финальное решение о наличии аномалии и генерирует свою булеву метку аномальности *anomaly-final* с непустым текстовым объяснением *anomaly-final-comment*, если LLM-ассистент принимает итоговое решение о наличии аномалии. Модель при этом возвращает простой JSON-документ с указанными меткой и комментарием.

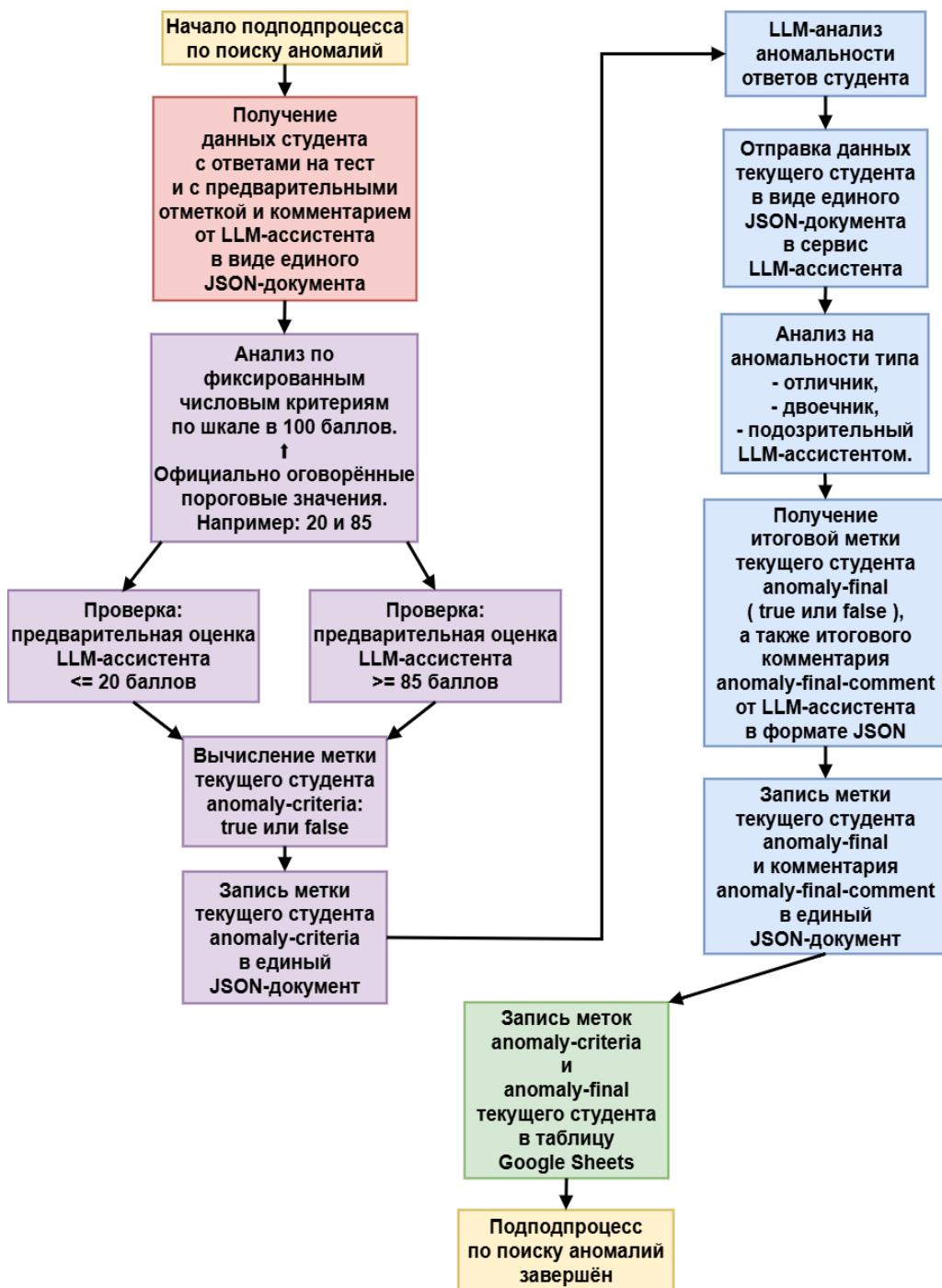


Рис. 8. Подподпроцесс многоуровневого анализа аномалий

Fig. 8. Sub-subprocess of multilevel anomaly analysis

**Схема таблицы для записи результатов.** Структура данных электронной таблицы включает различные группы колонок: основные данные (id, name, email, timestamp, q1, q2, ...), результаты LLM (ai-mark, ai-comment), метки аномалий (anomaly-

criteria, anomaly-final, anomaly-final-comment), оценки преподавателя (tutor-mark, tutor-comment). Полная структура представлена в таблице.

*Таблица.* – Структура данных электронной таблицы Google Sheets

*Table.* – Google Sheets data structure

| Категория  | Колонка                            | Описание  |
|--|------------------------------------|---|
| <b>Основные данные</b>   | <code>id</code>                    | ID ответа из Yandex.Forms   |
|  | <code>name</code>                  | ФИО студента  |
|  | <code>email</code>                 | Email студента  |
|  | <code>timestamp</code>             | Дата и время отправки теста (формат ISO8601)  |
|  | <code>q1, q2, q3...</code>         | Ответы на вопросы теста (по одной колонке на вопрос)  |
| <b>Результаты LLM</b>  | <code>ai-mark</code>               | Предварительная оценка от LLM-ассистента (0-100 баллов)   |
|  | <code>ai-comment</code>            | Предварительный текстовый комментарий LLM: список комментариев по неверным ответам и завершающий комментарий после списка   |
| <b>Метки аномалий для выявления</b><br>- отличников,<br>- двоечников,<br>- “странных”. | <code>anomaly-criteria</code>      | Метка аномальности по фиксированным критериям: $\leq 20$ или $\geq 85$ баллов; может быть <code>true</code> или <code>false</code>  |
|  | <code>anomaly-final</code>         | Итоговая метка аномальности на основе LLM-анализа. LLM-ассистент пишет <code>true</code> или <code>false</code> в своём JSON-фрагменте на основе единого JSON-документа с предыдущими метками, личными данными и ответами студента  |
|  | <code>anomaly-final-comment</code> | Комментарий от LLM об аномалии в текстовой форме или <code>none</code> , если LLM-ассистент решил окончательно, что это не аномалия. LLM-ассистент пишет либо комментарий, либо <code>none</code> в своём JSON-фрагменте, на основе единого JSON-документа с предыдущими метками, личными данными и ответами студента |
| <b>Оценки преподавателя</b>  | <code>tutor-mark</code>            | Финальная оценка преподавателя. Необязательное поле, если преподаватель согласен с ИИ   |
|  | <code>tutor-comment</code>         | Комментарий преподавателя. Необязательное поле, если преподаватель согласен с ИИ  |

Специальный шаг в Albato для маппинга JSON→Google Sheets обеспечивает корректное преобразование данных с ключами ai\_mark, ai\_comment в колонки таблицы ai-mark, ai-comment, anomaly-criteria, anomaly-final, anomaly-final-comment (рис. 9).

| A1 | I                         | J                 | K           | L   | M                | N   |                       |
|----|---------------------------|-------------------|-------------|---|------------------|---|-----------------------|
| 1  | q5                        | q6                | ai-mark     | ai-comment  | anomaly-criteria | anomaly-final                                   | anomaly-final-comment |
| 2  | быстрая сортир des wef xq |                   |             | <p>- Вопрос 4: Указан неверный SQL-оператор DELETE</p> <p>- Вопрос 6: Ответ не является корректным кодом на</p> <p>64 Итоговое заключение: Студент демонстрирует хоро</p>   | TRUE             | Странный: Ответы на q1-q3 правильные            |                       |
| 3  | быстрая сортир            | def find_max(lst) | if not lst: | <p>Все ответы студента верны.</p> <p>- Вопрос 1: Правильный ответ - B. O</p> <p>- Вопрос 2: Правильный ответ - B. C</p> <p>- Вопрос 3: Правильные ответы - A. I</p> <p>- Вопрос 4: Правильный ответ - C. SI</p> <p>- Вопрос 5: Правильный ответ - быст</p> <p>- Вопрос 6: Код функции корректен и</p> | 100              | Итоговое заключение: Студент демо               |                       |
| 4  | не помню                  | def find_max(lst) | if not lst: | <p>100 Итоговое заключение: Студент демо</p> <p>- Вопрос 4: Указан неверный SQL-оператор UPDATE</p> <p>- Вопрос 5: Ответ 'не помню' не является правильны</p>   | 79               | Итоговое заключение: Студент демонстрирует хоро |                       |
| 5  |                           |                   |             |   |                  |   |                       |

Рис. 9. Заголовки и пример строк в Google Sheets для проверки преподавателем

Fig. 9. Headers and sample rows in Google Sheets for tutor review

**Описание онлайн-формы для проведения тестирования.** Как уже было сказано ранее, для сбора данных используется специализированная веб-форма, созданная в Yandex.Forms. **\*\*Форма онлайн-теста\*\*** служит основной точкой входа для студентов. Она содержит персональные данные (ФИО, email) и тестовые вопросы разных типов (рис. 10). Встроенные баллы Yandex.Forms должны (как правило) отключаться и игнорироваться разрабатываемой системой такого типа, поскольку основную оценку по собственным настраиваемым критериям производят LLM-ассистенты и дополнительные баллы от сервиса Yandex.Forms в JSON данных могут сильно исказить результаты их работы.

**Интерпретация результатов и ограничения.** Выявляемые аномальные ответы студентов трактуются как кандидаты на ручную проверку преподавателем с возможным подтверждением с его стороны какой-либо альтернативной метрикой через дополнительное собеседование. Также преподавателем может периодически фиксироваться необходимость в изменениях порогов аномальных оценок для ответов на тест при явных систематических ложных срабатываниях, когда доля выявленный аномалий оказывается слишком высокой и доля ручного труда резко повышается.

Содержание +

Страница 1

Вопрос

q1. Какая времененная сложность у алгоритма бинарного поиска в отсортированном массиве?

Добавить пояснение Добавить изображение

Ответы

Плитка Звёзды В один ряд

Баллы

A. O(n)

Идентификатор 1756329828535

Скрытый ответ

B. O(log n)

Идентификатор 1756329828536

Скрытый ответ

C. O(n<sup>2</sup>)

Идентификатор 1756329828537

Скрытый ответ

D. O(1)

Идентификатор 1756329828538

Скрытый ответ

+ Добавить вариант

Настройки

Сортировка ответов

Без сортировки (selected)

По алфавиту

В случайному порядке для каждого пользователя

Идентификатор вопроса q1

Рис. 10. Редактор Yandex.Forms для создания контрольного теста для студентов

Fig. 10. Yandex.Forms editor for creating a test for students

Также ограничения данного подхода включают зависимость от качества и доступности внешних API, требования к базовым навыкам работы с веб-интерфейсами у преподавателя, к базовым знаниям языка JavaScript для работы с JSON-документами и их преобразования в другие JSON-документы, дополнительную проработку вопросов оптимизации на масштабы сотен и тысяч студентов. Очень важно учитывать тарифы и стоимость операций при эксплуатации связок (интеграций) на платформе Albato, а также стоимость расхода токенов различных моделей GPT от OpenAI или LLM-моделей от Sber Gigachat.

**Практическая значимость и применение.** Предлагаемая система представляет готовый к внедрению технологический прототип, который преподаватели могут адаптировать для создания своих собственных систем тестирования в различных

образовательных контекстах. Применение для этого интеграционной платформы Albato позволяет создавать гибкие системы без масштабного привлечения ИТ-специалистов.

Представленный прототип частично или в значительной степени может освобождать преподавателей от рутинной работы по проверке тестов, позволяет сконцентрироваться на индивидуальной работе с особыми «аномальными» группами студентов, анализе сложных случаев и совершенствовании образовательного процесса.

**Направления будущих исследований.** Перспективы дальнейшего развития описанного технологического решения включают, введение новых методов анализа аномалий по межквартильному размаху (IQR) в общей итоговой выборке студентов, а также изыскания в области более совершенных моделей детектирования аномалий.

Еще одним планируемым развитием представленного подхода является реализация отдельной дополнительной автоматизации на базе собственного сервиса (backend), который должен отвечать за автоматизированную массовую рассылку по электронной почте отчетов с окончательными оценками с учетом возможный комментариев и оценок преподавателя по тем студентам, по которым он сочтет нужным это сделать.

Особым направлением, на которое стоит обратить внимание являются разработки по нечётким критериям оценивания ответов студентов на тест [7].

Предполагается также исследовать возможные интеграции с существующими системами прокторинга для более надежного выявления академической нечестности и разработку алгоритмов адаптивного тестирования, где сложность последующих вопросов зависит не только от правильности предыдущих ответов, но и от результатов анализа аномалий в реальном времени [8].

Возможно, потребуется исследование этических аспектов использования ИИ в образовательной оценке, включая вопросы конфиденциальности, объективности и прозрачности алгоритмов [9, 10]. Необходима разработка отраслевых стандартов для оценки ИИ-систем в образовании и практических руководств по их внедрению.

## **Заключение**

Исследование демонстрирует концептуальную и практическую состоятельность нового подхода к автоматизации контрольно-оценочной деятельности в образовании. Разработан прототип интеллектуальной системы, который построен на комбинации генеративных языковых моделей, интеграционных платформ и многоуровневого анализа аномалий, что отчасти представляет ответ на ключевые вызовы современного

массового образования. При этом такая система не стремится заменить преподавателя, а призвана усилить его возможности, обеспечивая персонализацию и гуманизацию образования в цифровую эпоху.

#### **Список использованных источников** (на языке оригинала)

1. Рудинский И.Д. Принципы интеллектуального автоматизированного тестирования знаний. *Материалы международной конференции-выставки «Информационные технологии в образовании»*. 2001; Москва. Москва; МО РФ, 2001.
2. Crompton H., Burke D. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*. 2023;20:22. doi:10.1186/s41239-023-00392-8.
3. Рудинский И.Д., Клеандрова И.А. Концепция количественного оценивания объективности педагогического тестирования знаний. *Информатика и образование*. 2003;(12):100–104.
4. Emirtekin E. Large language model-powered automated assessment: a systematic review. *Applied Sciences*. 2025;15:5683. doi:10.3390/app15105683.
5. Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*. 2023;110:102274. doi:10.1016/j.lindif.2023.102274.
6. Jukiewicz M. The future of grading programming assignments in education: the role of ChatGPT in automating the assessment and feedback process. *Thinking Skills and Creativity*. 2024;53:101522. doi:10.1016/j.tsc.2024.101522.
7. Рудинский И.Д. Структурные основы тестологии. 2-е изд., испр. Москва: Горячая линия–Телеком; 2015. 244 с.
8. Давыдова Н.А., Рудинский И.Д. Автоматизированный синтез тестовых заданий для систем педагогического контроля знаний. *Информатизация образования и науки*. 2013;1(17):77–90.
9. Zawacki-Richter O., Marín V.I., Bond M., Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*. 2019;16:39. doi:10.1186/s41239-019-0171-0.
10. Malouff J.M., Thorsteinsson E.B. Bias in grading: a meta-analysis of experimental research findings. *Australian Journal of Education*. 2016;60(3):245–256. doi:10.1177/0004944116664618.

## References (на английском языке)

1. Rudinsky I.D. Principles of intelligent automated knowledge testing. *Proceedings of the International Conference-Exhibition “Information Technologies in Education”*; 2001; Moscow, Russia. Moscow; 2001. (In Russian)
2. Crompton H., Burke D. Artificial intelligence in higher education: the state of the field. *Int J Educ Technol High Educ.* 2023;20:22. doi:10.1186/s41239-023-00392-8.
3. Rudinsky I.D., Kleandrova I.A. Concept of quantitative assessment of objectivity in pedagogical knowledge testing. *Informatics and Education.* 2003;(12):100–104. (In Russian)
4. Emirtekin E. Large language model-powered automated assessment: a systematic review. *Appl Sci.* 2025;15:5683. doi:10.3390/app15105683.
5. Kasneci E., Sessler K., Kuchemann S., Bannert M., Dementieva D., Fischer F., et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ.* 2023;110:102274. doi:10.1016/j.lindif.2023.102274.
6. Jukiewicz M. The future of grading programming assignments in education: the role of ChatGPT in automating the assessment and feedback process. *Think Skills Creat.* 2024;53:101522. doi:10.1016/j.tsc.2024.101522.
7. Rudinsky I.D. Structural foundations of testology. 2nd ed., rev. Moscow: Goryachaya Liniya–Telekom; 2015. 244 p. (In Russian)
8. Davydova N.A., Rudinsky I.D. Automated synthesis of test items for systems of pedagogical knowledge control. *Informatization of Education and Science.* 2013;1(17):77–90. (In Russian)
9. Zawacki-Richter O., Marín V.I., Bond M., Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int J Educ Technol High Educ.* 2019;16:39. doi:10.1186/s41239-019-0171-0.
10. Malouff J.M., Thorsteinsson E.B. Bias in grading: a meta-analysis of experimental research findings. *Aust J Educ.* 2016;60(3):245–256. doi:10.1177/0004944116664618.