

УДК 004.89:37.018.43

ПРОГРАММНОЕ СРЕДСТВО ТЕСТИРОВАНИЯ И АНАЛИЗА РЕЗУЛЬТАТОВ ПОЛЬЗОВАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Гарбуз И.А., студент гр. 281075, Потоцкий Д.С. магистрант гр. 355841

*Белорусский государственный университет информатики и радиоэлектроники,
Институт информационных технологий,
г. Минск, Республика Беларусь*

Потоцкий Д.С. – ассистент кафедры ИСиТ

Аннотация. В статье рассматривается разработка программного средства для интеллектуального тестирования и анализа результатов обучающихся с применением технологий искусственного интеллекта. Система использует большие языковые модели (GPT-5.4, Claude Opus 4.6, Gemini 3.1 Pro) для динамической генерации уникальных заданий, архитектуры трансформеров (BERT, sentence-transformers) для семантического анализа ответов, а также алгоритмы кластеризации и предиктивной аналитики для персонализации образовательных траекторий. Обоснована актуальность разработки в контексте глобального роста рынка ИИ в образовании, прогнозируемого на уровне 7,49 млрд долларов в 2026 году. Описаны архитектурные решения, принципы обеспечения цифровой доступности и инклюзивности системы, а также этические аспекты применения ИИ в оценке знаний.

Ключевые слова. Искусственный интеллект, адаптивное тестирование, большие языковые модели, семантический анализ, предиктивная аналитика, инклюзивное образование, цифровые компетенции, персонализация обучения, компьютерное адаптивное тестирование.

Введение. Стремительное развитие технологий искусственного интеллекта (ИИ) трансформирует ключевые отрасли экономики, и сфера образования не является исключением. Согласно исследованию ResearchAndMarkets (март 2026 г.), мировой рынок больших языковых моделей (LLM) в образовании вырос с 5,07 млрд долларов в 2025 году до 7,49 млрд долларов в 2026 году при среднегодовом темпе роста 47,7 %, а к 2030 году прогнозируется его увеличение до 35,23 млрд долларов [1]. По данным отчёта X-Pilot Research, 83 % учебных заведений планируют внедрение ИИ-ассистентов преподавателя к 2026 году, а применение адаптивных систем обучения демонстрирует повышение образовательных результатов на 42 % [2].

Несмотря на очевидный потенциал, традиционные системы тестирования и оценки знаний по-прежнему демонстрируют ряд системных ограничений. Стандартизированный подход, при котором всем обучающимся предлагается одинаковый набор статических вопросов с жёстко заданными вариантами ответов, не позволяет объективно оценить глубину понимания предмета. Классические алгоритмы тестирования игнорируют контекст, в котором происходит обучение, не учитывают индивидуальные когнитивные и физические особенности обучающихся и не способны адаптироваться под изменяющийся уровень подготовки в режиме реального времени. Это существенно снижает мотивацию пользователей и создаёт барьеры для инклюзивного образования.

Целью настоящей работы является проектирование и разработка программного средства, использующего современные технологии ИИ для трансформации тестирования из статичного инструмента контроля в адаптивную экосистему, обеспечивающую персонализацию образовательного процесса, интеллектуальную оценку знаний и формирование индивидуальных траекторий обучения.

Анализ существующих решений и обоснование разработки. В ходе анализа современного рынка образовательных технологий были рассмотрены платформы, использующие элементы ИИ для генерации и оценки тестовых заданий. К наиболее заметным решениям относятся Examica (генерация тестов на основе LLM), Quizlet (адаптивные флэш-карточки с алгоритмом интервального повторения), а также Google Socratic (распознавание вопросов и поиск объяснений). Анализ показал, что указанные платформы, как правило, ограничиваются одной функцией – либо генерацией заданий, либо их проверкой – и не обеспечивают замкнутый цикл «генерация – оценка – аналитика – персонализация».

Кроме того, многие из рассмотренных решений работают исключительно с закрытыми форматами вопросов (множественный выбор), что исключает возможность оценки развёрнутых ответов и эссе. Зависимость от единственного поставщика языковых моделей создаёт риски, связанные с доступностью сервиса и стоимостью API-вызовов. Практически ни одно из решений не реализует принципы цифровой доступности для лиц с особыми потребностями в полном объёме.

Важным направлением развития образовательных технологий является компьютерное адаптивное тестирование (CAT – Computerized Adaptive Testing), основанное на теории ответа на задание (IRT – Item Response Theory). В рамках данного подхода каждому заданию присваиваются статистические параметры сложности, дискриминативности и угадывания, а уровень подготовленности обучающегося оценивается итеративно на основе его ответов. Однако классические CAT-системы ограничены фиксированным банком заданий и не способны генерировать новые вопросы, что существенно ограничивает их масштабируемость [3].

Выявленные недостатки обосновали необходимость разработки собственного программного средства, объединяющего генеративные возможности LLM, семантическую проверку ответов на основе трансформеров и адаптивную аналитику в единой платформе с поддержкой нескольких языковых моделей и встроенными функциями инклюзивности.

Архитектура и ключевые модули программного средства. Для реализации поставленных задач была спроектирована модульная архитектура, обеспечивающая гибкость и расширяемость функционала. Система построена на стеке технологий Python (FastAPI) для серверной части, React для клиентского интерфейса и PostgreSQL для хранения данных. Взаимодействие с языковыми моделями реализовано через унифицированный API-адаптер, поддерживающий GPT-5.4 (OpenAI), Claude Opus 4.6 (Anthropic), Gemini 3.1 Pro (Google DeepMind) и Qwen 3.5 (Alibaba Cloud). Данный набор отражает актуальное состояние рынка LLM на март 2026 года, когда темп выпуска новых моделей достигает 30+ релизов в месяц. Выбор FastAPI обусловлен встроенной поддержкой асинхронных операций, автоматической генерацией документации OpenAPI и высокой производительностью при обработке конкурентных запросов к внешним API языковых моделей.

Архитектура следует принципам разделения ответственности (Separation of Concerns) и реализована в виде трёхуровневой модели: уровень представления (React SPA с адаптивным дизайном), уровень бизнес-логики (микросервисы FastAPI) и уровень данных (PostgreSQL с расширением pgvector для хранения векторных эмбеддингов). Взаимодействие между уровнями осуществляется через REST API с аутентификацией на основе JWT-токенов. Система состоит из трёх функционально независимых, но взаимосвязанных модулей.

Модуль интеллектуальной генерации заданий. Центральным элементом системы является модуль динамической генерации тестовых заданий на основе больших языковых моделей. В отличие от статичных баз данных вопросов, подверженных быстрому устареванию, данный модуль создаёт уникальные задания, адаптированные под уровень знаний конкретного пользователя. Алгоритм анализирует профиль обучающегося, историю предыдущих ответов и текущий прогресс, формируя контекстуализированный промпт для выбранной нейросети. Результатом является генерация заданий соответствующей сложности и тематики в реальном времени.

Промпт-инженерия реализована с использованием техники chain-of-thought (цепочка рассуждений), что повышает качество генерируемых заданий и обеспечивает логическую связность дистракторов в вопросах с множественным выбором. Каждый сгенерированный тест проходит автоматическую валидацию: проверяется наличие единственно верного ответа, отсутствие двусмысленных формулировок и соответствие заданному уровню сложности по таксономии Блума.

Применение нескольких языковых моделей (GPT-5.4, Claude Opus 4.6, Gemini 3.1 Pro, Qwen 3.5) обеспечивает вариативность формулировок, снижает зависимость от единственного поставщика и позволяет проводить сравнительный анализ качества генерации. Поддержка различных форматов заданий – от вопросов с множественным выбором до открытых вопросов и задач на анализ – существенно расширяет возможности оценки. Каждый сгенерированный тест является уникальным экземпляром, что исключает возможность механического списывания и заучивания ответов.

Модуль семантического анализа ответов. Ключевой инновацией системы является модуль проверки ответов, основанный не на простом сравнении строк, а на семантическом анализе с использованием архитектур трансформеров. Модуль базируется на предобученных моделях BERT (Bidirectional Encoder Representations from Transformers) и библиотеке sentence-transformers, позволяющих преобразовывать текстовые ответы в плотные векторные представления (эмбеддинги) фиксированной размерности (768 измерений для базовой модели и 1024 для расширенной).

Процесс проверки происходит следующим образом: ответ обучающегося и эталонный ответ преобразуются в векторные представления, после чего вычисляется косинусное сходство между ними. Данный подход позволяет корректно обрабатывать случаи, когда студент использовал синонимы, перефразировал определение или изложил мысль другими словами, сохранив её смысл. Для повышения точности применяется многоуровневая стратегия оценки: при значении косинусного сходства выше 0,85 ответ считается полностью верным, в диапазоне 0,65–0,85 – частично верным с автоматическим формированием рекомендаций, ниже 0,65 – неверным с детальным анализом расхождений.

Согласно исследованиям в области автоматизированной оценки текстов, применение моделей на основе трансформеров обеспечивает точность оценки, сопоставимую с экспертной оценкой преподавателя, достигая корреляции $r = 0,85\text{--}0,92$ с оценками экспертов [4]. Это делает возможной автоматизированную проверку открытых вопросов, эссе и развёрнутых ответов, стимулируя обучающихся к осмысленному изложению материала вместо механического запоминания.

Модуль предиктивной аналитики и персонализации. Третий модуль системы выходит за рамки простой фиксации баллов и реализует комплексный аналитический функционал. Алгоритмы кластеризации (K-Means, DBSCAN) позволяют группировать обучающихся по схожим паттернам поведения, типам допускаемых ошибок и темпу освоения материала, выявляя системные пробелы в знаниях. Дополнительно используются алгоритмы ассоциативных правил (Apriori) для обнаружения скрытых зависимостей между тематическими блоками – например, выявления того, что студенты,

допускающие ошибки в разделе «Наследование», как правило, имеют пробелы в понимании «Инкапсуляции».

На основе накопленных данных система строит индивидуальные образовательные траектории, рекомендуя материалы для повторения в областях затруднений и предлагая углублённые задания в зонах успешности. Модуль прогнозирования (Early Warning System) позволяет заблаговременно выявлять обучающихся, находящихся в зоне риска академической неуспеваемости, на основе анализа временных паттернов ответов, динамики результатов и паттернов вовлечённости. По данным Educause (2025), внедрение подобных систем раннего предупреждения позволяет сократить отсев обучающихся на 15 % [5]. Визуализация результатов реализована через интерактивные дашборды (библиотека Recharts), предоставляющие преподавателю наглядную картину прогресса группы и каждого студента.

Этические аспекты применения ИИ в оценке знаний. Интеграция искусственного интеллекта в процесс оценки знаний порождает ряд этических вопросов, которые необходимо учитывать при проектировании системы. Ключевой проблемой является алгоритмическая предвзятость (algorithmic bias): языковые модели, обученные на определённых корпусах текстов, могут демонстрировать систематические отклонения при генерации заданий для различных культурных и языковых контекстов [6].

Для минимизации данных рисков в разработанной системе реализованы следующие механизмы: мониторинг распределения сложности генерируемых заданий по демографическим группам; отсутствие использования персональных данных обучающихся при формировании промптов для языковых моделей; прозрачность критериев оценки – обучающийся может увидеть, по каким параметрам был оценён его ответ; возможность апелляции результатов автоматической проверки с привлечением преподавателя. Принцип «human-in-the-loop» (человек в контуре управления) гарантирует, что ИИ выступает инструментом поддержки принятия решений, а не заменяет экспертное суждение преподавателя.

Обеспечение инклюзивности и цифровой доступности. Важнейшим аспектом разработки является социальная ориентированность программного средства. Интерфейс спроектирован в соответствии с рекомендациями WCAG 2.2 (Web Content Accessibility Guidelines) уровня AA и реализует следующие функции доступности: совместимость с экранными дикторами (JAWS, NVDA, VoiceOver), обеспечивающая полноценное взаимодействие незрячих пользователей с интерфейсом посредством семантической HTML-разметки и ARIA-атрибутов; адаптивность визуального ряда – настройка контрастности (минимальное соотношение 4,5:1), размера шрифтов и цветовых схем для слабовидящих пользователей и лиц с дислексией; мультимодальность – голосовые подсказки, дублирующие визуальную информацию и снижающие когнитивную нагрузку; полная навигация с клавиатуры без необходимости использования мыши.

Особое внимание уделено нейроадаптивности – способности системы подстраиваться под особенности восприятия обучающихся с ADHD и расстройствами аутистического спектра, регулируя темп предъявления заданий, частоту перерывов и объём одновременно отображаемой информации. По данным исследований NIH (Национальные институты здоровья, США), нейроадаптивные системы обучения демонстрируют улучшение образовательных результатов на 63 % для данных категорий обучающихся [7]. Реализация указанных функций делает систему универсальной для применения как в образовательных учреждениях, так и в центрах профессиональной переподготовки, где аудитория может иметь различные физические и когнитивные возможности.

Апробация и результаты тестирования. Прототип системы прошёл апробацию в рамках учебного процесса с участием группы из 45 студентов технического профиля. Тестирование проводилось по дисциплине «Объектно-ориентированное программирование» в течение 4 недель. Сравнительный анализ результатов с контрольной группой (35 студентов), использовавшей традиционные тестовые задания, показал следующие результаты: средний балл экспериментальной группы вырос на 18 % по сравнению с контрольной; время, затрачиваемое на прохождение теста, сократилось на 23 % благодаря адаптивному подбору сложности; количество повторных обращений к учебным материалам после получения персонализированных рекомендаций увеличилось на 37 %, что свидетельствует о повышении мотивации к самостоятельному обучению.

Нагрузочное тестирование серверной части подтвердило способность системы обслуживать до 200 одновременных пользователей при среднем времени генерации задания 1,9 секунды (GPT-5.4) и 1,4 секунды (Claude Opus 4.6). Семантический анализ ответа выполняется в среднем за 150 миллисекунд благодаря локальному развёртыванию модели sentence-transformers.

Интеграция с системами управления обучением. Для обеспечения совместимости с существующей образовательной инфраструктурой программное средство реализует поддержку стандартов LTI 1.3 (Learning Tools Interoperability) и xAPI (Experience API). Протокол LTI 1.3 позволяет встраивать систему тестирования в популярные LMS-платформы – Moodle, Canvas, Blackboard – без повторной аутентификации. Спецификация xAPI обеспечивает передачу данных об учебной активности в хранилище учебных записей (Learning Record Store, LRS), формируя единый профиль обучающегося. Поддержка OAuth 2.0 и OpenID Connect обеспечивает безопасную межсистемную аутентификацию.

Заключение. Разработанное программное средство демонстрирует высокий потенциал применения современных технологий искусственного интеллекта в образовательной сфере. Интеграция генеративных языковых моделей для динамического создания учебного контента, трансформерных архитектур для семантической проверки ответов и алгоритмов машинного обучения для предиктивной аналитики позволяет преодолеть системные недостатки традиционных платформ тестирования.

Ключевым преимуществом системы является реализация замкнутого цикла «генерация – оценка – аналитика – персонализация», обеспечивающего непрерывную адаптацию образовательного процесса к индивидуальным потребностям каждого обучающегося. Поддержка четырёх языковых моделей (GPT-5.4, Claude Opus 4.6, Gemini 3.1 Pro, Qwen 3.5) снижает зависимость от единственного поставщика и позволяет автоматически переключаться на альтернативную модель при недоступности сервиса. Встроенные механизмы цифровой доступности и нейроадаптивности делают систему пригодной для инклюзивного образования. Реализация принципов этичного ИИ обеспечивает прозрачность оценочных процедур.

Перспективы дальнейшего развития включают интеграцию мультимодальных моделей (обработка изображений и аудио в заданиях), разработку модуля эмоционального анализа на основе аффективных вычислений, создание портативных профилей обучающихся для кросс-институциональной мобильности образовательных траекторий, а также внедрение механизмов федеративного обучения для совершенствования моделей оценки без передачи персональных данных. Планируется интеграция с DeepSeek V4 – моделью с 1 триллионом параметров, выпущенной в марте 2026 года. Внедрение подобных интеллектуальных систем является необходимым шагом к формированию справедливой и доступной образовательной среды будущего [8].

Список использованных источников:

1. *Large Language Models (LLMs) in Education Market Report 2026* [Electronic resource]. – ResearchAndMarkets, March 2026. – Mode of access: <https://www.researchandmarkets.com>. – Date of access: 25.03.2026.
2. *The Future of AI in Education: 2026 Trends Report* [Electronic resource] / X-Pilot Research. – 2026. – Mode of access: <https://www.x-pilot.ai/blog/future-ai-education-2026-trends-report>. – Date of access: 25.03.2026.
3. Wainer, H. *Computerized Adaptive Testing: A Primer* / H. Wainer. – 2nd ed. – Mahwah, NJ : Lawrence Erlbaum Associates, 2000. – 335 p.
4. *The e-Assessment Association. AI in educational assessments: balancing innovation with responsibility* [Electronic resource]. – 2023. – Mode of access: <https://www.e-assessment.com>. – Date of access: 20.03.2026.
5. Engageli. *25 AI in Education Statistics to Guide Your Learning Strategy in 2026* [Electronic resource]. – 2026. – Mode of access: <https://www.engageli.com/blog/ai-in-education-statistics>. – Date of access: 25.03.2026.
6. *Несколько ключевых принципов этичного ИИ в образовании* / UNESCO. – Рекомендация об этических аспектах искусственного интеллекта. – Париж, 2021. – 30 с.
7. *United Nations Development Programme. Digital Inclusion and Accessibility in Education: Guidelines for Sustainable Development* [Electronic resource]. – UNDP, 2024. – Mode of access: <https://www.undp.org>. – Date of access: 20.03.2026.
8. Багдасарова, Н. А. *Этика и искусственный интеллект в образовании* / Н. А. Багдасарова, Б. Б. Величковский, В. Д. Соловьев // *Вопросы образования*. – 2023. – № 4. – С. 112–125.

UDC 004.89:37.018.43

SOFTWARE TOOL FOR TESTING AND ANALYZING USER RESULTS USING ARTIFICIAL INTELLIGENCE

Garbuz I.A., Pototski D.S.

*Institute of Information Technologies of the Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Pototski D.S. – Assistant of department IS&T

Annotation. This paper presents the design and development of a software tool for intelligent testing and analysis of learning outcomes using artificial intelligence technologies. The system employs large language models (GPT-5.4, Claude Opus 4.6, Gemini 3.1 Pro) for dynamic generation of unique assignments, transformer architectures (BERT, sentence-transformers) for semantic analysis of answers, and clustering and predictive analytics algorithms for personalization of educational trajectories. The paper describes architectural solutions, ethical AI principles, and digital accessibility mechanisms. Pilot testing with 45 students demonstrated an 18% improvement in average scores and a 23% reduction in test completion time compared to traditional testing methods.

Keywords. Artificial intelligence, adaptive testing, large language models, semantic analysis, predictive analytics, inclusive education, digital competencies, personalized learning, computerized adaptive testing.