

УДК 378.147:004.891.2

<https://doi.org/10.25587/2587-5604-2026-3-61-71>

Оригинальная научная статья

Применение генеративного ИИ для разработки оценочных средств в LMS Moodle

Ю. В. Корнилов

Северо-Восточный федеральный университет имени М. К. Аммосова,
г. Якутск, Российская Федерация✉ kornilov@lenta.ru

Аннотация

Статья посвящена исследованию возможностей генеративного искусственного интеллекта для автоматизированного формирования банков тестовых заданий по материалам вузовских лекций и их последующего преобразования в формат Moodle XML. Цель работы – оценить пригодность DeepSeek и гибридной платформы Perplexity для создания содержательно корректных, педагогически обоснованных и технически импортируемых в LMS Moodle тестовых материалов. Исследование проведено на материале одиннадцати лекций дисциплины «Инфографика и визуализация в обучении». DeepSeek использовался для обработки семи лекций и генерации 70 заданий, Perplexity – четырёх лекций и генерации 40 заданий. Анализировались разнообразие типов вопросов, когнитивный уровень заданий, зависимость формулировок от контекста отдельной лекции, корректность XML-синтаксиса, настройка весовых коэффициентов ответов и особенности заданий типа Cloze. Установлено, что обе системы способны существенно ускорить создание банка вопросов, однако результаты требуют обязательной экспертной проверки. В выборке DeepSeek доля заданий, проверяющих преимущественно фактологические знания, составила 70%, тогда как в выборке Perplexity – 90%; общий показатель достиг 77,3%. При генерации DeepSeek выявлены ошибки в значениях fraction, использовании десятичного разделителя, структуре тегов answer и text, а также в содержании тега name. Файлы Perplexity импортировались без синтаксических ошибок, но отличались большей ориентацией на воспроизведение информации, наличием контекстно зависимых формулировок и ограниченной вариативностью допустимых ответов в заданиях Cloze. Сделан вывод, что техническое разнообразие форматов вопросов не обеспечивает разнообразия когнитивных действий. Эффективное применение генеративного ИИ требует детализированных промптов, автоматической валидации XML, распределения заданий по уровням таксономии Блума и последующей педагогической и психометрической экспертизы.

Ключевые слова: генеративный искусственный интеллект, большие языковые модели, автоматическая генерация вопросов, тестовые задания, Moodle XML, LMS Moodle, DeepSeek, Perplexity, электронное обучение, педагогическое оценивание.

Финансирование. Исследование не имело финансовой поддержки.

Для цитирования: Корнилов Ю. В. Применение генеративного ИИ для разработки оценочных средств в LMS MOODLE. *Вестник Северо-Восточного федерального университета им. М.К. Аммосова. Vestnik of NorthEastern Federal University. Серия «Педагогика. Психология. Философия». Pedagogics. Psychology. Philosophy». 2026: 43 (3): С. 61-71. DOI: 10.25587/2587-5604-2026-3-61-71*

Original article

Using generative AI to develop assessment tools in LMS Moodle

Iurii V. Kornilov

M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation

✉ kornilov@lenta.ru

Abstract

The article examines the use of generative artificial intelligence to automatically create test-item banks from university lecture materials and convert them into Moodle XML. The study aims to assess

the suitability of DeepSeek and the Perplexity hybrid platform for producing pedagogically meaningful and technically valid test materials importable into Moodle. The empirical material comprised eleven lectures from the course Infographics and Visualization in Education. DeepSeek processed seven lectures and generated 70 items, whereas Perplexity processed four lectures and generated 40 items. The analysis addressed item-type diversity, cognitive level, wording dependent on the context of a particular lecture, XML syntax, answer-weight settings, and the implementation of Cloze items. The findings show that both systems can substantially reduce the time needed to create a question bank, although their outputs require expert review. Fact-oriented items accounted for 70% of the DeepSeek sample and 90% of the Perplexity sample, producing an overall proportion of 77.3%. The DeepSeek-generated files contained errors involving fraction values, decimal separators, the structure of answer and text tags, and the content of the name tag. The Perplexity files were imported without detected syntax errors; however, they emphasized factual recall, included context-dependent references to lecture materials, and offered limited acceptance of alternative correct answers in Cloze tasks. The study concludes that technical diversity in item formats does not automatically ensure cognitive diversity. Effective use of generative AI therefore requires detailed prompts, automated XML validation, deliberate distribution of tasks across Bloom's taxonomy levels, and subsequent pedagogical and psychometric evaluation.

Keywords: generative artificial intelligence, large language models, automatic question generation, test items, Moodle XML, LMS Moodle, DeepSeek, Perplexity, e-learning, educational assessment

Funding. This research received no external funding.

For citation: Kornilov Iu. V. Using generative AI to develop assessment tools in the LMS MOODLE. *Vestnik of North-Eastern Federal University. Pedagogics. Psychology. Philosophy.* 2026: 43 (3): Pp. 61-71. DOI: 10.25587/2587-5604-2026-3-61-71

Введение

Тематика активного внедрения технологий искусственного интеллекта (ИИ) в современную образовательную реальность является одной из самых популярных в современном научном дискурсе. При этом вопросы внедрения ИИ в образовательных целях привлекает внимание как специалистов в области машинного обучения, так и научно-педагогического сообщества в целом [1, 2]. Постепенно интегрированными ИИ-системами, направленными на решение тех или иных организационных и, в том числе, образовательных задач, дополняются и электронные информационно-образовательные системы. «ЭИОС активно интегрирует сервисы искусственного интеллекта, что позволяет персонализировать обучение, оптимизировать управление контентом, автоматизировать оценку и анализировать образовательные данные в режиме реального времени» [11, С. 182].

Автоматическая генерация обучающих и оценочных (тестовых) материалов – одно из наиболее востребованных направлений внедрения ИИ в электронную информационно-образовательную среду. Например, формирование высококачественных вопросов по учебным текстам лежит в основе способов непрерывной проверки знаний обучающихся, а также является средством развития самостоятельного мышления. При этом нейронные подходы к генерации вопросов (Neural Question Generation, NQG) здесь весьма эффективны: они могут генерировать не только грамматически корректные и предметно-ориентированные вопросы, но и педагогически полезные – как показано в исследованиях по генерации на основе пересечения знаний [12].

В условиях широкого применения больших языковых моделей ожидается, что автоматическая генерация вопросов поможет создавать не только связанные и актуальные, но и ценные с образовательной точки зрения вопросы. Al Faraby S. и др. отмечают, что «нейронная генерация вопросов (Neural Question Generation, NQG) показала лучшие результаты в этом отношении» [12, С. 1008]. В ранних исследованиях, посвящённых автоматической генерации вопросов (Automatic Question Generation, AQG), в основном использовались подходы, основанные на правилах, когда для преобразования входного текста в вопросы

определялись конкретные правила. Позже, с появлением подходов, основанных на глубоком обучении, произошел значительный сдвиг в сторону использования нейронных методов для AQG, обычно называемых NQG. «Нейронная генерация вопросов (Neural Question Generation, NQG) – это особая форма AQG, в которой для генерации вопросов используются модели на основе нейронных сетей» [12, С. 1011]. Модели глубокого обучения продемонстрировали эмпирическое превосходство над подходами, основанными на правилах, с точки зрения автоматической оценки, которая измеряет сходство между сгенерированными вопросами и справочными вопросами [13]. Так за последние пять лет глубокое обучение стало основным методом AQG, способствующим прогрессу в этой области.

Следовательно, уже сейчас можно очертить круг ключевых направлений в области использования ИИ в образовательной практике. Это адаптивное и персонализированное обучение, сфера интеллектуальных обучающих систем, автоматизированное оценивание результатов обучения, прогнозирование успеваемости и чат-боты.

При этом такие функции, как образовательная аналитика, поддержка принятия решений и контроль за проведением экзаменов, получают значительно меньше научного внимания. Авторы также определяют приоритетные направления исследований: оценка масштабов внедрения генеративного искусственного интеллекта на практике и определение условий его эффективного использования – главными из них являются готовность преподавателей и студентов, четкое понимание возможностей инструмента и наличие индивидуальных образовательных траекторий или специализированных методологий [2].

По мнению Вихман В. В. и др. «наблюдается большое количество зарубежных публикаций, посвященных конкретным способам, методам, технологиям и возможностям применения различных субтехнологий ИИ для всех участников образовательных отношений» [1, с. 178]. Данное явление весьма закономерно ввиду ряда причин, связанных как с технологическим прогрессом и экономикой, так и с развитием педагогической науки. Также авторы отмечают, что «в отечественной литературе подобное разнообразие существенно меньше ... что может объясняться тем, что активное использование технологий ИИ за рубежом началось значительно раньше, нежели в Российской Федерации, и в зарубежном научном дискурсе уже успело накопиться значительное разнообразие исследовательских тематик» [1, с. 178].

Более конкретный горизонт применения ИИ в образовании уже определен, и, как отмечает Кузьминов Я. И. и др. «важно внедрение персонализированных образовательных траекторий, учитывающих индивидуальные образовательные потребности и возможности каждого студента» [3, с. 26]. Отдельные работы фиксируют трансформацию роли преподавателя в условиях цифровизации и рассматривают ИИ как инструмент, ускоряющий обратную связь и позволяющий адаптировать учебные материалы под разные стили восприятия информации.

Существенный вклад в рассмотрении данных вопросов вносит обзор авторов Zawacki-Richter O., Marín V.I., Bond M. и др., где прослеживается цепочка применения ИИ от автоматизации создания заданий до их проверки с предоставляемой обратной связью и предоставлением анализа полученных результатов [4]. Вместе с тем авторы Кошкина Е. А. и др. отмечают, что «скорость генерации стимульного материала является одним из главных преимуществ ИИ, но одновременно отмечаются и большая вероятность утечки персональных данных; возникновение этических и социальных проблем; технологические ограничения; повышение потребности в специальной подготовке педагогов; недостоверность информации, представленной ИИ по запросу пользователей; нарушение академической честности и т. п.» [2, с. 41].

В другом многостороннем обзоре литературы (Mustafa M. Y. и др.), проведенном в основном на материалах США и Китая, выявлено, что в системе образования применение ИИ в основном сосредоточено на поддержке преподавателей и обучающихся, и в меньшей степени для поддержки административного аппарата [5].

Сужая спектр аспектов применения ИИ в образовании следует коснуться вопросов контроля знаний, и, в частности, автоматической генерации вопросов, AQD (Automatic Question Generation). Современные модели обладают способностью генерировать вопросы различной степени сложности, начиная от простых фактологических и заканчивая вопросами, требующими более глубокого уровня когнитивной обработки информации (Oliveira R. и др.) [6].

Экспериментальные сравнения нескольких LLM (GPT-3.5, Llama 2-Chat, Flan T5 XXL) на материале университетских лекций показали, что модели без дополнительного обучения способны генерировать вопросы, оцениваемые студентами как достаточно ясные и релевантные исходному материалу, при этом GPT-3.5 продемонстрировала лучшее согласование вопроса и ответа. Вместе с тем исследование Alamoudi S. и др. указывает на сохраняющиеся ограничения в точности согласования вопрос-ответ, что делает дообучение и дополнительную настройку модели необходимыми для задач, требующих высокой точности. Сравнение шаблонного, онтологического и LLM-основанного подходов на основе метрик BERT Precision, Recall, F1 и семантической схожести показало, что генерация на основе плоского списка понятий с использованием LLM даёт наилучший баланс релевантности и разнообразия вопросов, хотя иногда порождает несвязанные или галлюцинированные формулировки, что, несомненно, требует корректировки самим преподавателем [7].

Отдельного внимания заслуживают отечественные исследования (Куручка К. С., Башаримов Ю. С.), максимально близкие к теме данной статьи. Ими описана нейросетевая модель автогенерации тестов для системы Moodle на основе анализа методических материалов, включающая модули предобработки текста, генерации вопросов, валидации и фильтрации, а также использующая архитектуру из двух LLM – основной модели-генератора и модели-эксперта для оценки качества создаваемых заданий [8].

Стоит обратить внимание на психометрическую валидность контролирующих материалов, создаваемых ИИ. В исследовании Malec W. представлен подробный анализ того, как качественно отличаются тесты, сгенерированные на базе ИИ (на примере ChatGPT) с материалами, созданными преподавателем. Выявлено, что сгенерированные ИИ вопросы заметно уступают по показателям надежности, дискриминативности, качеству предлагаемых дистракторов, что требует последующей экспертной проверки преподавателем [9].

В качестве решения отечественными исследователями предлагается верификационная модель «ИИ проверяет ИИ», где агенты имеют свои роли (генератор, редактор, верификатор), а также вводятся критерии семантической ясности, релевантности и устойчивости к перефразированию. Результаты были математически доказаны: «изначальный уровень качества генерации, составлявший 0.93 (93% заданий соответствовали критериям качества после этапа шлифовки), был повышен до уровня 0.97 после полного цикла верификации» [10, с. 10].

Отдельную группу составляют исследовательские прототипы, интегрированные напрямую с LMS, функционирующие на базе LLM для платформы Moodle, а также решения, сочетающие парсинг исходных документов (текстов лекций и других учебных материалов) с последующей генерацией структурированных данных для формирования контрольных

вопросов и тестовых заданий. При этом в опубликованной научной литературе крайне редко подробно рассматривается проблема корректного экспорта сгенерированных тестовых заданий именно в формат Moodle XML (настройка весовых коэффициентов ответов, корректный синтаксис XML, исправление ошибок и пр., а также дообучение LLM для дальнейшего упразднения таких ошибок), что указывает на существующий пробел, который призвано восполнить исследование, предложенное в настоящей статье.

Материалы и методы

В основу исследования положен метод автоматизированного формирования банка тестовых заданий на основе текстов лекций с использованием генеративных языковых моделей DeepSeek и Perplexity с последующим преобразованием результатов в формат Moodle XML. Полученные данные импортируются в LMS Moodle для применения в условиях электронного обучения, где тестовые задания используются как инструмент текущего контроля, самопроверки и сопровождения самостоятельной работы обучающихся. При проектировании подхода учитывалось, что для одного из вариантов импорта вопросов LMS Moodle использует строго определённую XML-структуру, синтаксис которой весьма требователен к содержанию.

Предлагаемый подход включает несколько последовательно реализуемых этапов: подготовку исходного лекционного текста, генерацию тестовых заданий различных типов, проверку их содержательной корректности, формирование Moodle XML-разметки и контроль импортируемости полученного файла в LMS Moodle с необходимой корректировкой промптов. В качестве типов заданий использовались различные тестовые вопросы (с множественным выбором, вопросы на сопоставление, вопросы на установление правильной последовательности и вопросы со вставкой пропущенных слов).

Результаты и обсуждение

В проведенном исследовании были совершены попытки применения больших языковых моделей для выявления их пригодности при генерации тестовых вопросов к материалам лекций в вузе по дисциплине «Инфографика и визуализация в обучении» (44.04.01 Педагогическое образование, магистерская программа «Корпоративное электронное обучение»). В основу легли результаты применения семейства LLM DeepSeek, а также гибридной платформы Perplexity. Следует обратить внимание, что Perplexity не является лишь LLM, а больше представляется как гибридная платформа, которая сочетает в себе собственные языковые модели с интеллектуальной «оркестрацией» множества сторонних ИИ-инструментов.

Каждый промпт основывался на общеизвестной последовательности: назначение роли → задача → контекст → ограничения → формат ответа → критерии успеха (пример качественного ответа) → что делать при неясности (недостаточность данных) → верификация выходных данных и исправление (опционально).

В рамках данного исследования каждая лекция загружалась в PDF-формате и сопровождалась промптом:

«Ты – преподаватель вуза с ученой степенью, специалист в области инфографики. Изучи материалы лекции в прикрепленном файле. Подготовь 10 тестовых вопросов разного типа (с множественным выбором, вопросы на сопоставление, вопросы на установление правильной последовательности и вопросы со вставкой пропущенных слов). На основании результатов сгенерируй XML-код в формате Moodle XML для импортирования этих вопросов в LMS Moodle. Проверь правильность полученного XML-кода и исправь его при необходимости».

DeepSeek показал некоторую однообразность в формулировках тестовых вопросов, прослеживалось неполное покрытие содержания лекций, а также слабое ориентирование на проверку аналитических способностей обучающихся, подменяемых проверкой фактологического знания. Большинство вопросов относилось к типу «множественный выбор», который проверяет только уровень «знание – запоминание» по таксономии Блума, и недостаточно вопросов на понимание причинно-следственных связей. Всего в рамках проводимого исследования модели DeepSeek предлагалось обработать 7 лекций по дисциплине «Инфографика и визуализация в обучении» по 10 разнотипных тестовых заданий на лекцию. Доля вопросов на проверку фактологического знания составила 70% (49 из 70 вопросов).

С позиции теории классических тестов в предложенных вариантах отсутствует градация по степени трудности (лёгкие, средние, сложные), что не позволяет дифференцировать испытуемых. Это в целом снижает уровень дискриминативности теста, что не позволяет определить студентов, изучивших материалы поверхностно, от тех, кто проработал материалы достаточно глубоко.

Также зачастую генерировался контекст, привязывающий формулировку вопроса к конкретной лекции без указания ее названия или описания раздела, которому посвящен вопрос. Например, для лекций по скрайбину DeepSeek предлагал такие вопросы, как «Назовите три разновидности скрайбинга, упомянутые в тексте лекции». Данная формулировка носит непосредственную привязку тестового вопроса к конкретной лекции, что не дает применять такие вопросы при проведении промежуточного тестирования, когда тест формируется из различных категорий банка тестовых вопросов LMS Moodle (тест охватывает ряд лекций и/или разделов). В таких случаях студент не имеет возможности корректно понять суть вопроса и дать верный ответ.

Следом за процессом генерации вопросов при организации электронного обучения возникает вопрос импортирования их в применяемую в образовательной организации LMS. В отечественных вузах чаще всего применяется популярная система с открытым кодом LMS Moodle, для которой далее будет применяться импортирование тестовых вопросов.

При всей многофункциональности LMS Moodle есть и ряд ограничений для пополнения банка тестовых вопросов. Ключевой, чаще всего применяемый преподавателями, – ручной ввод через форму создания нового вопроса. Очевидно, что для ускорения процесса наполнения банка вопросов следует использовать другие подходы, в частности импорт из внешних файлов с определенной разметкой.

В рамках описываемого исследования применялся формат Moodle XML, позволяющий на основе подготовленных вопросов генерировать средствами ИИ готовые XML-файлы, которые в дальнейшем импортировались в LMS Moodle. Пробные запросы показали, что ИИ в целом неплохо справляется с генерацией XML-кода. Однако в результате проведенных процедур по формированию XML-кода тестовых заданий для LMS Moodle были выявлены и некоторые ошибки.

Первоочередной ошибкой оказались некорректные значения `<answer fraction>` при генерации вопросов с несколькими вариантами. При загрузке XML-данных в LMS Moodle система возвращала ошибку синтаксиса. Решением проблемы для тестовых вопросов, где несколько вариантов ответов верные, был написан промпт, требующий выставление значения для каждого из них `<answer fraction="n">`, где $n=100/\text{количество правильных ответов}$ (с указанием до 3 знаков дробной части после разделителя дробной части в виде точки). Данный промпт содержал текст, требующий корректного расставления весовых коэффициентов, не вызывающих ошибку при загрузке в LMS Moodle:

«Если в тестовом вопросе типа «множественный выбор» есть 4 правильных ответа, то установи каждому правильному варианту значение $\langle \text{answer fraction}=\rangle 25 \rangle$. Если в тестовом вопросе типа «множественный выбор» есть 2 правильных ответа, то установи каждому правильному варианту значение $\langle \text{answer fraction}=\rangle 50 \rangle$. Если в тестовом вопросе типа «множественный выбор» есть 3 правильных ответа, то установи каждому правильному варианту значение $\langle \text{answer fraction}=\rangle 33.333 \rangle$ ».

Как альтернатива последнему замечанию возможно добавление отдельного дистрактора «все ответы верные» со значением $\langle \text{answer fraction}=100 \rangle$. В таком случае пропадает необходимость расчета и расставления весов между всеми правильными вариантами ответов.

Также при первых попытках генерации XML-кода ответы $\langle \text{answer} \rangle$ не были заключены в тег $\langle \text{text} \rangle$, что также было исправлено с помощью соответствующего промпта.

Существенным улучшением было исправление содержания тега $\langle \text{name} \rangle$, который по сути дублировал текст вопроса, а не указывал его описание в банке вопросов, что может вызывать у преподавателя неудобства с выбором необходимых вопросов при формировании теста с вопросами из разных категорий банка вопросов.

Также при первых попытках в сгенерированном коде значения $\langle \text{answer fraction} \rangle$ вместо разделителя целой и дробной части DeepSeek предлагал разделитель “,” вместо “.”, что также вызывало ошибку системы при попытке загрузки XML-кода в банк тестовых заданий LMS Moodle.

Подобное исследование для обработки очередных лекций по предмету «Инфографика и визуализация в обучении» (4 лекции по 10 разнотипных тестовых заданий на лекцию) проводилось на базе гибридной платформы Perplexity.

Несмотря на то, что Perplexity для генерации тестовых вопросов и их портирования в Moodle XML предлагал запустить запрос как отдельную задачу в режиме «Computer» для получения лучших результатов (режим «computer» создаёт более развёрнутый результат, используя больше инструментов и контекста), режима «Deep research» было достаточно.

Результаты показали, что вопросы на сопоставление, правильную последовательность, множественный выбор были составлены без ошибок, однако также, как и у DeepSeek, присутствовали формулировки вопросов с привязкой к конкретному тексту лекции (например, «Какое определение наиболее точно соответствует содержанию лекции?»; или «Выберите три задачи, для которых, согласно лекции, эффективно применять инфографику»). В этой связи процесс импорта XML-данных не сопровождался ошибками, что говорит о качественном соблюдении XML-синтаксиса. С тегом $\langle \text{fraction} \rangle$ Perplexity поступил иначе: было предложено 3 верных и три неверных варианта ответов. При этом значение у правильных вариантов помечалось как $\langle \text{fraction}=\rangle 33.33333 \rangle$, а у неправильных – с отрицательным знаком, $\langle \text{fraction}=\rangle -33.33333 \rangle$. Таким образом, при выборе неправильных ответов доли балла за вопрос вычитались из общей суммы набранных, что не прослеживалось у DeepSeek.

Тип вопроса со вставкой пропущенных слов (Cloze) в Perplexity был обработан не до конца корректно. В частности, порядок пропущенных слов не варьировался, в качестве правильного ответа была заложена лишь одна последовательность, считающаяся верной. То есть, при ответе на данный вопрос с указанием верных ответов, но в иной последовательности, LMS Moodle идентифицировала этот ответ как некорректный.

Сравнительно с вопросами от DeepSeek в тестах Perplexity чаще наблюдалась ограниченность формулировок и акцент на проверку фактологических знаний. Результаты анали-

за выявили достаточно высокую долю таких вопросов в количестве 90% (36 вопросов из 40). Количество вопросов, формулировка которых напрямую привязана к конкретной лекции, составила 12,5% (5 вопросов из 40). Ошибок в синтаксисе XML в ходе эксперимента выявлено не было.

Отличительной чертой в сгенерированных вопросах у Perplexity появилась градация по степени трудности – появилась разница в весовых коэффициентах (“defaultgrade”). Так, у всех вопросов типа «множественный выбор» весовой коэффициент составил 1.0000000 независимо от того, сколько правильных ответов. В вопросах типа «установление соответствия» и «правильная последовательность» данный коэффициент также составлял 1.000000. При этом у вопросов типа «вставка пропущенных слов» весовой коэффициент варьировался от 2 до 4 в зависимости от количества пропущенных в вопросе слов. Важно отметить, что данный параметр учитывается в результатах тестирования в зависимости от настроек LMS Moodle и в некоторых версиях при ручном вводе вопроса просто отсутствует.

Заключение

Проведённое исследование подтвердило практическую возможность использования генеративных языковых моделей DeepSeek и платформы Perplexity для автоматизированного создания банков тестовых заданий на основе текстов университетских лекций и их последующего представления в формате Moodle XML для ЭИОС. Применение такого подхода позволяет сократить трудозатраты преподавателя при подготовке материалов для текущего контроля, самопроверки и организации самостоятельной работы обучающихся. Формат Moodle XML обладает высокой практической значимостью, поскольку поддерживает импорт различных типов вопросов и сопутствующих данных (весовые коэффициенты) непосредственно в банк вопросов LMS Moodle.

Вместе с тем результаты показали, что автоматическая генерация не гарантирует педагогического качества получаемых заданий. В материалах DeepSeek доля вопросов, ориентированных преимущественно на воспроизведение фактологических знаний, составила 70% – 49 из 70 заданий. В материалах Perplexity данный показатель достиг 90% (36 из 40 заданий). В совокупной выборке к фактологическим были отнесены 85 из 110 вопросов, или 77,3%. Следовательно, разнообразие технических типов заданий – множественный выбор, сопоставление, установление правильной последовательности и вставка пропущенных слов – само по себе не обеспечивает разнообразия проверяемых когнитивных действий. Даже задания сложной структуры могут оставаться на уровне узнавания и воспроизведения учебной информации.

Сравнение исследуемых инструментов выявило различия в характере создаваемых результатов. DeepSeek потребовал последовательной корректировки промптов для устранения ошибок Moodle XML, связанных с использованием десятичного разделителя, структурой тегов и содержанием названий вопросов. Сформированные Perplexity файлы импортировались без выявленных синтаксических ошибок, однако созданные задания чаще были ориентированы на фактологическое знание и содержали формулировки, привязанные к неопределённому тексту лекции. Кроме того, задания типа «cloze» не всегда предусматривали вариативность допустимых правильных ответов. Выявленная разница значений «defaultgrade» характеризует распределение максимальных баллов между вопросами, но не может сама по себе рассматриваться как объективная градация их трудности.

Таким образом, качество автоматически сформированного банка вопросов определяется не только возможностями используемой модели, но и структурой промпта, наличи-

ем формализованных требований к содержанию и XML-разметке, а также обязательной экспертной проверкой результата. В промпте целесообразно явно задавать требуемое распределение заданий по когнитивным уровням, исключать контекстно зависимые формулировки, устанавливать правила формирования дистракторов и весовых коэффициентов, предусматривать альтернативные правильные ответы и требовать итоговую проверку корректности XML-кода. Генеративный ИИ в данном случае следует рассматривать не как автономную замену преподавателя, а как средство ускорения технической и методической подготовки оценочных материалов.

Ограничениями исследования являются использование материалов одной учебной дисциплины, относительно небольшой объём выборки, рассмотрение только двух ИИ-инструментов и отсутствие эмпирической проверки заданий на выборке обучающихся. Поэтому полученные результаты характеризуют прежде всего технологическую корректность и содержательные особенности генерации, но не позволяют окончательно судить о надёжности, валидности, фактической трудности и дискриминативности созданных тестов.

Перспективы дальнейшей работы связаны с расширением выборки дисциплин и языковых моделей, разработкой унифицированного многоэтапного промпта, автоматической валидацией Moodle XML и внедрением схемы «генератор – редактор – верификатор». Необходимо также апробация вопросов в реальном образовательном процессе с расчётом индексов трудности и дискриминативности, анализом качества дистракторов и последующей корректировкой банка заданий. Реализация этих направлений позволит перейти от технически корректной генерации тестов к созданию педагогически обоснованных оценочных материалов, проверяющих не только запоминание, но и понимание, применение, анализ и оценивание учебной информации.

Л и т е р а т у р а

1. Вихман В. В., Миндигулова А. А., Ромм М. В. Искусственный интеллект в образовании: обзор возможностей и ограничений. *Идеи и идеалы*. 2024;16(4):167–188. DOI 10.17212/2075-0862-2024-16.4.1-167-188.
2. Кошкина Е. А., Бордовская Д. С., Гнедых Д. С. Генеративный искусственный интеллект в высшем образовании: обзор теоретических подходов и практик применения. *Высшее образование в России*. 2025;34(6):36–57. DOI 10.31992/0869-3617-2025-34-6-36-57.
3. Кузьминов Я. И., Кручинская Е. В., Груздев И. А., Наумов А. А. Отстающие и опережающие: как студенты используют генеративный искусственный интеллект в образовательных целях. *Высшее образование в России*. 2025;34(6):9–35. DOI 10.31992/0869-3617-2025-34-6-9-35.
4. Zawacki-Richter O., Marín V.I., Bond M. et al. 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.
5. Mustafa M. Y., Tlili A., Lampropoulos G. et al. A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. *Smart Learning Environments*. 2024;11(59). DOI 10.1186/s40561-024-00350-5.
6. Oliveira R., Hernández A., Garbin M., Dos Reis J., Azevedo R. Automatic Question Generation with Large Language Models: A Survey. *ACM Computing Surveys*. 2026;58(15), Article 383:34 p. DOI 10.1145/3832773.
7. Alamoudi S., Al Khuzayem L. A., Jamal A. Optimizing Automated Question Generation for Educational Assessments: A Semantic Analysis of LLMs with Structured and Unstructured Ontologies. *Engineering, Technology & Applied Science Research*. 2025;15(3):23664–23671. DOI 10.48084/etasr.10662.
8. Курочка К. С., Башаримов Ю. С. Нейросетевая модель автогенерации тестов для студентов в системе Moodle на основе анализа методических материалов. *Цифровая трансформация*. 2025;31(3):66–75. DOI 10.35596/1729-7648-2025-31-3-66-75.

9. Malec W. A comparative study of AI-generated and teacher-constructed language tests. *Roczniki Humanistyczne*. 2025;73(11):149–163. DOI 10.18290/rh257311.8s.
10. Султанаев Я. Т., Абзалилов Р.Р., Вильданова В. Ф. Верификационная тестология: формализованная система оценки тестовых заданий для образовательных систем с искусственным интеллектом. *Human Progress*. 2025;11(9):16. DOI 10.46320/2073-4506-2025-9a-16.
11. Смык А. Ф., Ткачева Т. М. Внедрение ИИ в образование: условия и тенденции. *Балтийский гуманитарный журнал*. 2025;4:177–189. DOI 10.46845/2071-5331-2025-4-74-177-189.
12. Al Faraby S., Adiwijaya A., Romadhony A. Review on Neural Question Generation for Education Purposes. *International Journal of Artificial Intelligence in Education*. 2024;34:1008–1045. DOI 10.1007/s40593-023-00374-x.
13. Du X., Shao J., Cardie C. Learning to ask: Neural question generation for reading comprehension. In: Barzilay R., Kan M.-Y. (Eds.). *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017;1(Long Papers):1342–1352. DOI 10.18653/v1/P17-1123.
14. Maity S., Deroy A. The future of learning in the age of generative AI: Automated question generation and assessment with large language models. *Computation and Language*. 2024. DOI 10.48550/arXiv.2410.09576
15. Карданова Е. Ю. и др. Новые подходы к оцениванию: искусственный интеллект как драйвер изменений в образовании. *Современная аналитика образования*. 2025; 5 : 88.

References

1. Vikhman V. V., Mindigulova A. A., Romm M. V. Artificial intelligence in education: an overview of opportunities and limitations. *Ideas and Ideals*. 2024;16(4, pt. 1):167–188 (in Russian). DOI 10.17212/2075-0862-2024-16.4.1-167-188.
2. Koshkina E. A., Bordovskaya N. V., Gnedykh D. S., et al. Generative artificial intelligence in higher education: a review of theoretical approaches and application practices. *Higher Education in Russia*. 2025;34(6):36–57 (in Russian). DOI 10.31992/0869-3617-2025-34-6-36-57.
3. Kuzminov Ya. I., Kruchinskaya E. V., Gruzdev I. A., Naumov A. A. Lagging and leading: how students use generative artificial intelligence for educational purposes. *Higher Education in Russia*. 2025;34(6):9–35 (in Russian). DOI 10.31992/0869-3617-2025-34-6-9-35.
4. Zawacki-Richter O., Marin V. I., Bond M. et al. 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.
5. Mustafa M. Y., Tlili A., Lampropoulos G. et al. A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. *Smart Learning Environments*. 2024;11(59). DOI 10.1186/s40561-024-00350-5.
6. Oliveira R., Hernández A., Garbin M., Dos Reis J., Azevedo R. Automatic question generation with large language models: a survey. *ACM Computing Surveys*. 2026;58(15), Article 383, 34 p. DOI 10.1145/3832773.
7. Alamoudi S., Al Khuzayem L.A., Jamal A. Optimizing automated question generation for educational assessments: a semantic analysis of LLMs with structured and unstructured ontologies. *Engineering, Technology & Applied Science Research*. 2025;15(3):23664–23671. DOI 10.48084/etasr.10662.
8. Kurochka K.S., Basharymau Y.S. Neural network model for auto-generating tests for students in the Moodle system based on the analysis of teaching materials. *Digital Transformation*. 2025;31(3):66–75 (in Russian). DOI 10.35596/1729-7648-2025-31-3-66-75.
9. Malec W. A comparative study of AI-generated and teacher-constructed language tests. *Roczniki Humanistyczne*. 2025;73(11):149–163. DOI 10.18290/rh257311.8s.
10. Sultanaev Ya. T., Abzalilov R. R., Vildanova V. F. Verification testology: a formalized system for evaluating test items for educational systems with artificial intelligence. *Human Progress*. 2025;11(9):16 (in Russian). DOI 10.46320/2073-4506-2025-9a-16.
11. Smyk A.F., Tkacheva T.M. The introduction of AI in education: conditions and trends. *Baltic Humanitarian Journal*. 2025;(4):177–189 (in Russian). DOI 10.46845/2071-5331-2025-4-74-177-189.
12. Al Faraby S., Adiwijaya A., Romadhony A. Review on neural question generation for education purposes. *International Journal of Artificial Intelligence in Education*. 2024;34:1008–1045. DOI 10.1007/s40593-023-00374-x.

13. Du X., Shao J., Cardie C. Learning to ask: neural question generation for reading comprehension. In: Barzilay R., Kan M.-Y. (Eds.). *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. 2017;1(Long Papers):1342–1352. DOI 10.18653/v1/P17-1123.

14. Maity S., Deroy A. The future of learning in the age of generative AI: Automated question generation and assessment with large language models. *Computation and Language*. 2024. DOI 10.48550/arXiv.2410.09576

15. Kardanova E.Y. et. al. New approaches to assessment: artificial intelligence as a driver of change in education. *Modern Education Analytics*. 2025;5:88 (in Russian).

Сведения об авторе

КОРНИЛОВ Юрий Вячеславович – к. пед. н., доцент кафедры цифрового и технологического образования, Северо-Восточный федеральный университет имени М. К. Аммосова, г. Якутск, Российская Федерация, ORCID: 0000-0002-5801-2616, SPIN: 9087-1411, e-mail: kornilov@lenta.ru

Information about the author

KORNILOV Iurii Viacheslavovich – Cand. Sci. (Pedagogics), Associate Professor, Department of Digital and Technological Education, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russian Federation, ORCID: 0000-0002-5801-2616, SPIN: 9087-1411, e-mail: kornilov@lenta.ru

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Conflict of interests

The author declares no conflict of interests.

Поступила в редакцию / Submitted 03.07.2026

Поступила после рецензирования / Revised 31.08.2026

Принята к публикации / Accepted 21.09.2026