УДК 378.147:61:004.8 DOI 10.25005/3078-5022-2026-3-3-282-296

РЕЗЮМЕ

Ф.С. БАКОЕВ[1], Н.А. ЮСУПОВ[2], Н.А. ЗАРИФИ[3]

ЦИФРОВАЯ ПЛАТФОРМА ДЛЯ ПОЭТАПНОГО РАЗБОРА КЛИНИЧЕСКИХ КЕЙСОВ С ИСПОЛЬЗОВАНИЕМ БОЛЬШОЙ ЯЗЫКОВОЙ МОДЕЛИ: РАЗРАБОТКА И ПЕРВЫЙ ОПЫТ АПРОБАЦИИ

[1] Кафедра детских болезней № 1 им. профессора Х.Дж. Аминова, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино», Душанбе, Республика Таджикистан

[2] Кафедра оториноларингологии им. профессора Ю.Б. Исхоки, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино», Душанбе, Республика Таджикистан

[3] Кафедра неврологии, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино», Душанбе, Республика Таджикистан

Цель. *разработать веб-платформу для группового поэтапного разбора клинических кейсов, в которой ход рассуждения группы фиксируется, а большая языковая модель (БЯМ) используется для предварительного разбора занятия, и описать первый опыт её апробации.*

Материал и методы. *Платформа написана на TypeScript (React, Node.js/Express, PostgreSQL). Кейс раскрывается по этапам. На каждом этапе группа формулирует дифференциально-диагностические гипотезы и уточняющие вопросы, а система прослеживает их «линии» от этапа к этапу. Студенты подключаются со смартфонов по QR-коду без регистрации. БЯМ размечает текст кейса по ролям, а после занятия готовит структурированный разбор с баллами 0–100 по этапам. Баллы проходят через жёсткие ограничения, которые зависят от фактического объёма работы. Итоговую оценку ставит преподаватель. Данные апробации (23.03–01.05.2026) выгружены из базы платформы и проанализированы описательными методами.*

Результаты. *В апробации участвовали 2 кафедры, 3 факультета (медицинский, педиатрический, стоматологический) и 8 учебных групп 4–6 курсов. По одному кейсу из 7 этапов проведено 7 занятий, из них 6 завершены. Группы сформулировали 49 линий гипотез и 57 линий вопросов, в среднем 2,4–3,7 гипотезы и 2,8–4,4 вопроса на этапе. Число гипотез снижалось к завершающим этапам (с 3,4–3,7 до 2,4), а число вопросов возрастало. Ни одна линия гипотез не изменила формулировку. Вход по QR-коду опробован в одном занятии: 2 студента прислали 4 идеи, 3 из них ведущий принял в общий список. ИИ-разбор получен для 3 занятий. В обоих случаях, когда были и балл БЯМ, и оценка преподавателя, преподаватель оценил работу на 15–16 баллов выше. Затраты на один ИИ-разбор составили менее 0,01 долл. США.*

Заключение. *Платформа поддерживает все этапы CBL-занятия: подготовку кейса, порционную выдачу информации, запись гипотез и вопросов группы, участие всех студентов со смартфонов, разбор и оценку. Она технически реализуема и применима в учебном процессе. Она делает ход клинического рассуждения группы наблюдаемым и выявляет закономерности, например сужение дифференциального ряда к концу кейса. Объём данных апробации пока недостаточен для статистической оценки согласованности БЯМ и преподавателя. Для этого планируется проспективное исследование.*

Ключевые слова: *обучение на клинических случаях, клиническое мышление, медицинское образование, большие языковые модели, искусственный интеллект, автоматизированное оценивание, цифровая образовательная платформа.*

Для цитирования: *Ф.С. Бакоев, Н.А. Юсупов, Н.А. Зарифи. Цифровая платформа для поэтапного разбора клинических кейсов с использованием большой языковой модели: разработка и первый опыт апробации. Наука и образование. 2026;3(3):282-296. https://doi.org/10.25005/3078-5022-2026-3-3-282-296*

ХУЛОСА

Ф.С. БАКОЕВ[1], Н.А. ЮСУПОВ[2], Н.А. ЗАРИФӢ

ПЛАТФОРМАИ РАҚАМӢ БАРОИ ТАҲЛИЛИ МАРҲИЛАВИИ ҲОЛАТҲОИ КЛИНИКӢ БО ИСТИФОДАИ МОДЕЛИ КАЛОНИ ЗАБОНӢ: ТАҲИЯ ВА ТАҶРИБАИ АВВАЛИНИ САНҶИШ

[1] Кафедраи бемориҳои кудаконаи №1 ба номи профессор Ҳ.Ҷ. Аминов МДТ "ДДТТ ба номи Абуалӣ ибни Сино" Душанбе, Тоҷикистон

[2] Кафедраи оторинолорингология ба номи профессор Ю.Б. Исҳоқӣ МДТ "ДДТТ ба номи Абуалӣ ибни Сино" Душанбе, Тоҷикистон

[3] Кафедраи асабшиносӣ МДТ "ДДТТ ба номи Абуалӣ ибни Сино" Душанбе, Тоҷикистон

Мақсад. *таҳияи платформаи веб барои таҳлили гурӯҳӣ ва марҳилавии ҳолатҳои клиникӣ, ки раванди тафаккури гурӯҳро сабт мекунад ва модели калони забониро (МКЗ) барои таҳлили пешакии дарс истифода мебарад, инчунин тавсифи таҷрибаи аввалини санҷиши он.*

Мавод ва усулҳо. *Платформа бо TypeScript (React, Node.js/Express, PostgreSQL) сохта шудааст. Ҳолати клиникӣ марҳила ба марҳила кушода мешавад; дар ҳар марҳила гурӯҳ фарзияҳои ташхиси тафриқавӣ ва саволҳои дақиққунанда пешниҳод мекунад, ки «хатти» онҳо аз марҳила ба марҳила пайгирӣ карда мешавад. Донишҷӯён бо смартфон тавассути рамзи QR бе қайдгирӣ пайваст мешаванд. МКЗ матни ҳолатро аз рӯи нақшаҳо ҷудо мекунад ва пас аз дарс таҳлили сохтории кор бо холҳои 0–100 барои ҳар марҳила омода месозад; ба холҳо маҳдудиятҳои қатъӣ вобаста ба ҳаҷми воқеии кор татбиқ мешаванд. Баҳои ниҳоиро омӯзгор мегузорад. Маълумоти санҷиш (23.03–01.05.2026) аз пойгоҳи додаҳо гирифта шуда, бо усулҳои тавсифӣ таҳлил гардид.*

Натиҷаҳо. *Дар санҷиш 2 кафедра, 3 факултет (тиббӣ, педиатрӣ, стоматологӣ) ва 8 гурӯҳи таълимии курсҳои 4–6 иштирок карданд. Аз рӯи як ҳолати 7-марҳилагӣ 7 дарс гузаронида шуд, ки 6-тои он анҷом ёфт. Гурӯҳҳо 49 хатти фарзия ва 57 хатти савол пешниҳод карданд; шумораи фарзияҳо дар марҳилаҳои охир кам шуда (аз 3,4–3,7 то 2,4), шумораи саволҳо зиёд гардид. Воридшавӣ тавассути рамзи QR дар як дарс санҷида шуд: 2 донишҷӯ 4 ғоя фиристоданд, ки 3-тоашро роҳбари дарс қабул кард. Таҳлили МКЗ барои 3 дарс гирифта шуд; дар ҳар ду ҳолате, ки ҳам холи МКЗ ва ҳам баҳои омӯзгор мавҷуд буд, омӯзгор корро 15–16 хол баландтар баҳо дод. Хароҷоти як таҳлил камтар аз 0,01 доллари ИМА буд.*

Хулоса. *Платформа аз ҷиҳати техникӣ амалишаванда ва дар раванди таълим татбиқпазир аст; он раванди тафаккури клиникии гурӯҳро намоён мегардонад. Ҳаҷми маълумот барои арзёбии оморӣ мувофиқати МКЗ ва омӯзгор ҳоло нокифоя аст; барои ин тадқиқоти оянда ба нақша гирифта шудааст.*

Калимаҳои калидӣ: *таълим дар асоси ҳолатҳои клиникӣ, тафаккури клиникӣ, таҳсилоти тиббӣ, моделҳои калони забонӣ, зеҳни сунъӣ, баҳодиҳии худкор, платформаи рақамии таълимӣ.*

ABSTRACT

F.S. BAKOEV[1], N.A. YUSUPOV[2], N.A. ZARIFI [3]

A DIGITAL PLATFORM FOR STAGED CLINICAL CASE DISCUSSION WITH A LARGE LANGUAGE MODEL: DEVELOPMENT AND INITIAL PILOT EXPERIENCE

[1] Department of Paediatrics No. 1 named after Professor H.J. Aminov, State Educational Institution 'Abu Ali ibn Sina Tajik State Medical University', Dushanbe, Republic of Tajikistan
[2] Department of Otorhinolaryngology named after Professor Yu.B. Iskhoki, State Educational Institution 'Abu Ali ibn Sino Tajik State Medical University', Dushanbe, Republic of Tajikistan
[3] Department of Neurology, State Educational Institution 'Abu Ali ibn Sino Tajik State Medical University', Dushanbe, Republic of Tajikistan

Objective. *To develop a web platform for staged, group-based clinical case discussion that records the group's reasoning process and uses a large language model (LLM) for preliminary session debriefing, and to describe the initial experience of its pilot use.*

Material and methods. *The platform was built in TypeScript (React, Node.js/Express, PostgreSQL). A case is disclosed stage by stage; at each stage the group formulates differential-diagnostic hypotheses and clarifying questions, whose "lineages" are tracked across stages. Students join from smartphones by scanning a QR code without registration. The LLM tags case text by speaker role and, after the session, produces a structured debrief with stage scores of 0–100, which are then subjected to hard caps based on the actual amount of work. The final grade is assigned by the teacher. Pilot data (23 March – 1 May 2026) were extracted from the platform database and analysed descriptively.*

Results. *The pilot involved 2 departments, 3 faculties (general medicine, paediatrics, dentistry) and 8 study groups of 4th–6th-year students. Seven sessions were held on one seven-stage case, six of which were completed. Groups generated 49 hypothesis lineages and 57 question lineages (2.4–3.7 hypotheses and 2.8–4.4 questions per stage on average). The number of hypotheses decreased towards the final stages (from 3.4–3.7 to 2.4), while the number of questions increased. None of the hypothesis lineages was reworded. QR-code participation was tested in one session: 2 students submitted 4 ideas, 3 of which were accepted by the session leader. An LLM debrief was obtained for 3 sessions. In both sessions with both an LLM score and a teacher grade, the teacher scored the work 15–16 points higher. The cost of one LLM debrief was below USD 0.01.*

Conclusion. *The platform is technically feasible and applicable in teaching. It makes a group's clinical reasoning observable and reveals patterns such as the narrowing of the differential diagnosis towards the end of a case. The pilot data are not yet sufficient for a statistical assessment of LLM–teacher agreement, which is planned in a prospective study.*

Keywords: *case-based learning, clinical reasoning, medical education, large language models, artificial intelligence, automated assessment, digital learning platform.*

Введение

Клиническое мышление, то есть умение собрать данные о пациенте, выдвинуть и последовательно проверить диагностические гипотезы, — центральная компетенция врача и одна из самых трудных для обучения и оценивания [1–3]. Значительная доля диагностических ошибок связана не с недостатком знаний, а с когнитивными сбоями: преждевременным закрытием гипотезы, якорением, недостаточным дифференциальным рядом [4, 5]. Поэтому образовательные стратегии, которые делают процесс рассуждения явным и дают обратную связь по нему, а не только по конечному диагнозу, считаются особенно ценными [2, 6].

Обучение на клинических случаях (case-based learning, CBL) — один из наиболее распространённых способов развивать клиническое мышление на доклиническом и клиническом этапах [7, 8]. По данным систематического обзора BEME, студенты и преподаватели высоко оценивают CBL, а сам метод способствует связыванию теории с практикой и активному участию в обсуждении [7]. В отличие от проблемно-ориентированного обучения (PBL) [9], CBL предполагает более направляемое

обсуждение под руководством преподавателя, что многие студенты и педагоги воспринимают как более эффективное использование учебного времени [10].

Типичное CBL-занятие строится по модели «последовательного раскрытия» (progressive disclosure), наиболее близкой к реальной клинической работе [3]. Группа из 8–15 студентов получает первую порцию сведений о пациенте, обычно жалобы и краткий анамнез. Студенты выдвигают предварительные диагнозы (гипотезы) и решают, какие данные им нужны для их проверки: вопросы анамнеза, данные осмотра, лабораторные и инструментальные исследования. Затем преподаватель выдаёт следующую порцию информации. Группа пересматривает гипотезы, одни отбрасывает, другие уточняет и формулирует новые вопросы. Цикл повторяется до финального раскрытия диагноза, после чего проводится разбор (дебрифинг): что группа сделала правильно, где ошиблась и почему. Преподаватель выступает модератором: он направляет обсуждение, но не даёт готовых ответов [3, 8].

Проведение такого занятия с помощью бумаги, доски или слайдов требует от преподавателя одновременно решать несколько задач, и это создаёт ряд трудностей. Во-первых, подготовка: кейс нужно разбить на порции и оформить так, чтобы студенты не видели информацию раньше времени. Во-вторых, преподаватель должен одновременно модерировать дискуссию и записывать гипотезы на доске, а запись стирается при переходе к следующей порции данных. В результате ход рассуждения группы не сохраняется: после занятия остаётся лишь итоговый диагноз, но не видно, когда и почему менялись гипотезы. В-третьих, в группе говорят несколько активных студентов, а вклад остальных невидим. В-четвёртых, на содержательный разбор каждого этапа и индивидуальную обратную связь часто не хватает времени. Наконец, оценивание опирается на общее впечатление преподавателя и плохо сопоставимо между группами и кафедрами. Системы интерактивного опроса аудитории (audience response systems) повышают вовлечённость [11, 12], а виртуальные пациенты улучшают навыки клинического мышления [13, 14]. Однако первые, как правило, не связаны со структурой клинического кейса, а вторые рассчитаны на индивидуальную работу и требуют больших затрат на разработку сценариев.

Появление больших языковых моделей (БЯМ) открыло новые возможности для медицинского образования. БЯМ демонстрируют результаты на уровне проходного балла в лицензионных экзаменах [15, 16] и рассматриваются как инструмент для создания учебных материалов, персонализированной обратной связи и оценивания [17, 18]. Первые исследования показывают умеренную или хорошую корреляцию оценок БЯМ с оценками экзаменаторов при проверке открытых ответов студентов-медиков [19]. Вместе с тем известны систематические искажения моделей-«судей»: склонность к завышению оценок, чувствительность к длине и формулировкам ответа [20]. Кроме того, международные руководства подчёркивают необходимость прозрачности, контроля со стороны человека и защиты данных при использовании ИИ в образовании [21, 22]. Поэтому актуальна разработка инструментов, в которых БЯМ выполняет вспомогательную функцию, а её выводы ограничены формальными правилами и подтверждаются преподавателем.

Цель исследования

Разработать веб-платформу для подготовки и проведения CBL-занятий, которая поддерживает все этапы занятия, от подготовки кейса до разбора и оценки, фиксирует развитие гипотез и вопросов группы и использует БЯМ для подготовки предварительного разбора, и описать первый опыт её апробации в учебном процессе.

Материал и методы

Дизайн исследования

Работа выполнена в два этапа. Первый этап — разработка платформы по принципам исследования на основе проектирования (design-based research): итеративные циклы «прототип — использование на занятиях — доработка». Второй

этап — ретроспективный описательный анализ данных апробации платформы с 23 марта по 1 мая 2026 г. на двух кафедрах ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино» (ТГМУ): кафедре детских болезней № 1 им. профессора Х.Дж. Аминова и кафедре оториноларингологии им. профессора Ю.Б. Исхоки. В апробации участвовали учебные группы 4–6 курсов медицинского, педиатрического и стоматологического факультетов.

Педагогическая модель CBL-занятия на платформе

Платформа проектировалась вокруг классического цикла CBL-занятия. Каждому шагу цикла соответствует функция платформы, которая снимает с преподавателя техническую нагрузку и сохраняет результаты работы группы (табл. 1). Логика занятия остаётся прежней: группа обсуждает кейс вслух в аудитории, а преподаватель модерирует дискуссию. Платформа не заменяет живое обсуждение. Она выполняет роль «умной доски», которая выдаёт информацию порциями, хранит гипотезы и вопросы, собирает идеи студентов и готовит материал для разбора.

Таблица 1. Этапы CBL-занятия и соответствующие функции платформы

Этап CBL-занятия	Задача преподавателя	Функция платформы
Подготовка кейса	Разбить кейс на порции, сформулировать цели, подготовить ожидаемые ответы	Конструктор этапов, учебные цели этапа, ИИ-разметка текста, изображения, скрытый эталон
Запуск занятия	Собрать группу, обеспечить участие всех	Выбор кейса и группы, QR-код для входа студентов без регистрации
Первая порция данных	Показать жалобы и анамнез, не раскрывая остального	На экране отображается только текущий этап
Выдвижение гипотез и вопросов	Модерировать дискуссию, фиксировать предложения	Электронная доска гипотез и вопросов; идеи студентов со смартфонов во «входящих»
Получение новых данных	Выдать следующую порцию, сохранить сделанное	Переход к следующему этапу с автоматическим переносом гипотез и вопросов (линии)
Пересмотр гипотез	Побуждать уточнять и отбрасывать гипотезы	Редактирование и удаление гипотез, история их изменений
Финальное раскрытие	Сообщить диагноз и тактику	Финальный этап кейса
Разбор (дебрифинг)	Дать обратную связь по ходу рассуждения	Линии гипотез, ИИ-черновик разбора по этапам, вклад каждого студента
Оценивание	Выставить оценку, сравнить группы	Предварительный балл ИИ, итоговая оценка преподавателя, аналитика кафедры

Технологическая основа

Платформа реализована как веб-приложение на языке TypeScript (React, Node.js/Express) с базой данных PostgreSQL и работает в любом современном браузере,

в том числе на смартфоне, без установки программ. Сведения о кафедрах, факультетах, курсах, учебных группах, кейсах, занятиях, гипотезах, вопросах и оценках хранятся в связанных таблицах базы данных. Для функций ИИ используется программный интерфейс БЯМ в режиме структурированного вывода (JSON) с температурой 0,3. Модель задаётся в настройках. В текущей версии используется DeepSeek [23], а в период апробации применялись также модели deepseek-chat и Gemini 2.0 Flash. Общая архитектура показана на рисунке 1.

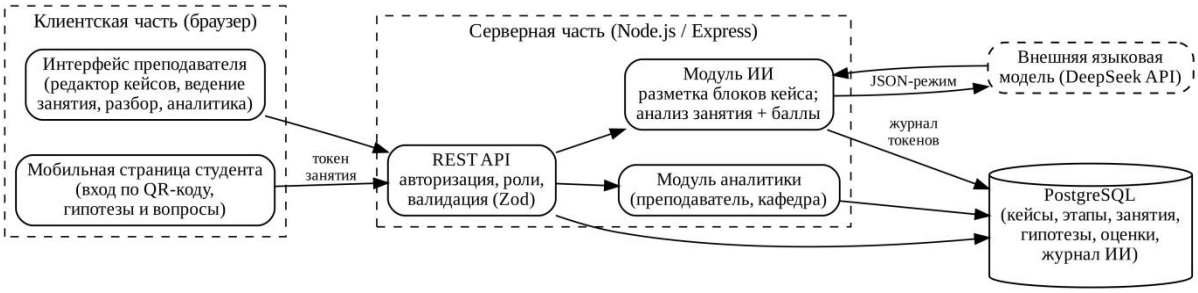

Рисунок 1. Архитектура платформы

Модель данных и педагогическая структура кейса

Кейс принадлежит кафедре и может быть привязан к факультетам и курсам. Он состоит из упорядоченных этапов. Для каждого этапа задаются название, учебные цели и набор блоков содержимого: текст, реплики пациента, записи врача, повествование или изображение. Последний этап может быть помечен как «финальное раскрытие», где сообщаются окончательный диагноз и разбор. Для каждого кейса преподаватель может заполнить «эталон» — ожидаемые гипотезы и ключевые вопросы. Эталон не показывается студентам и используется только при ИИ-анализе. Изменение кейса увеличивает номер его версии, и каждое занятие сохраняет версию, по которой оно проводилось. Это позволяет дорабатывать кейс, не искажая результаты уже проведённых занятий.

ИИ-разметка содержимого кейса

Чтобы ускорить подготовку кейсов, преподаватель может вставить исходный текст (например, из методического пособия), и БЯМ размечает его на смысловые фрагменты: повествование, реплики врача и речь пациента. Это позволяет показывать кейс в виде диалога. В инструкции модели явно запрещено сокращать, пересказывать текст или добавлять факты: разрешено только разметить исходный текст. Для защиты от ошибок модели предусмотрены три механизма. Во-первых, полученная HTML-разметка очищается по белому списку тегов. Во-вторых, проверяется полнота: если видимый текст ответа короче трети исходного (для текстов длиннее 800 символов) или ответ обрезан по лимиту длины, система вместо ответа модели использует простую разбивку исходного текста на абзацы. В-третьих, при отсутствии доступа к модели применяется та же эвристическая разметка. Результат всегда доступен преподавателю для проверки и ручного редактирования.

Проведение занятия

Преподаватель (ведущий) запускает занятие, выбирая кейс и учебную группу. Ход занятия показан на рисунке 2. На экране аудитории отображается текущий этап кейса. Группа в обсуждении формулирует гипотезы (предварительные диагнозы) и уточняющие вопросы (какие данные анамнеза, осмотра или обследований нужны), а ведущий вносит их в общий список. При переходе к следующему этапу текущий этап «отправляется», а все гипотезы и вопросы переносятся в новый этап с сохранением уникального идентификатора линии (lineageId). Группа может переформулировать гипотезу (например, «пневмония» → «внебольничная правосторонняя нижнедолевая пневмония») или удалить её. Это делает возможным отслеживание «траектории» каждой мысли: по завершении кейса каждая линия классифицируется как *устойчивая*

(формулировка не менялась до последнего этапа), *эволюционировавшая* (формулировка менялась) или *отброшенная* (отсутствует на последнем этапе). Для каждого этапа фиксируются время открытия и отправки, что позволяет рассчитать его длительность.

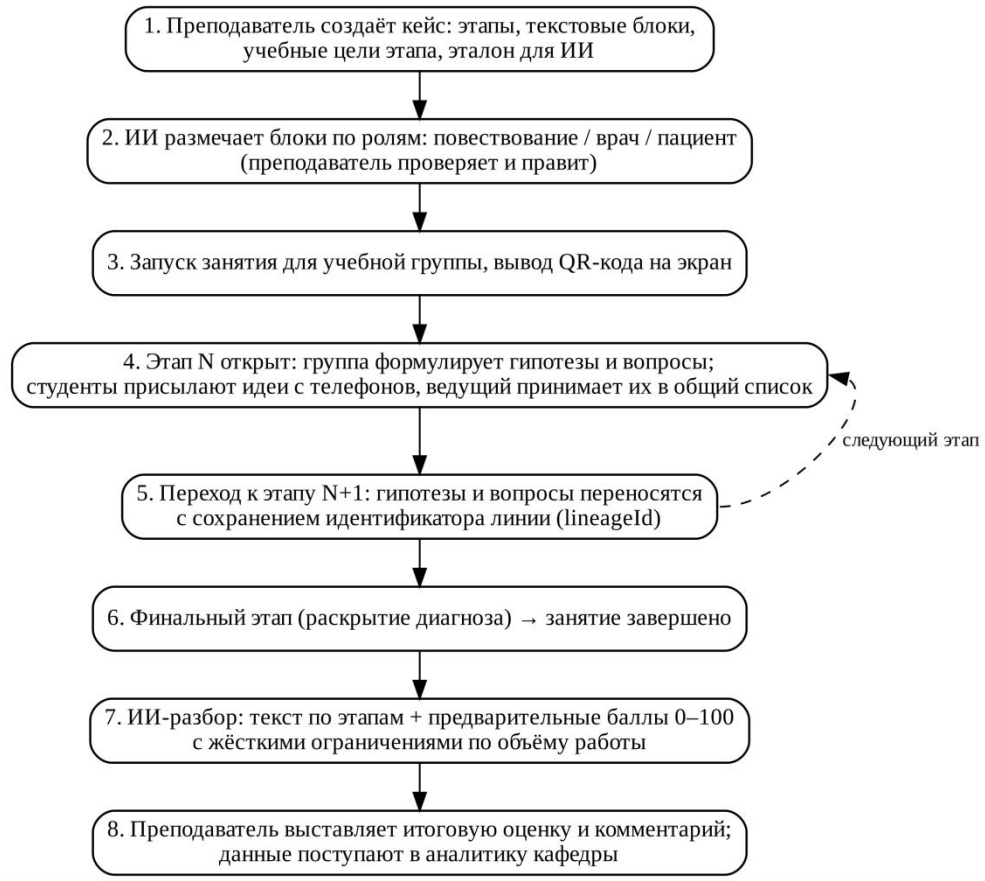

Рисунок 2. Последовательность проведения занятия на платформе

Участие студентов без регистрации

Чтобы вовлечь всю группу, а не только активных студентов, для каждого занятия генерируется короткий токен из 8 символов, а на экран выводится QR-код со ссылкой. Студент открывает ссылку на смартфоне, вводит имя и может отправлять ведущему свои гипотезы и вопросы к текущему этапу. Учётная запись не нужна: студенту присваивается случайный идентификатор, который хранится в браузере. Для защиты от злоупотреблений установлены ограничения: до 280 символов на идею, не более 8 непринятых идей на этапе и 30 идей за занятие от одного студента. Идеи поступают ведущему во «входящие», и он может одним действием принять идею в общий список группы. Факт принятия фиксируется.

Для обратной связи по индивидуальному вкладу рассчитывается ориентировочный балл студента (0–100), который зависит от числа присланных гипотез (h) и вопросов (q). Если присланы только гипотезы (1–2), балл равен 10 + 5h; если только гипотезы (3 и более) — min(40; 28 + 3(h − 3)); если только вопросы — min(28; 15 + 4(q − 1)); если и гипотезы, и вопросы — min(78; 48 + 4·min(h; 5) + 3·min(q; 4)). Дополнительно начисляется 4 балла за каждый этап участия начиная со второго (не более 12) и 2 балла за каждую принятую ведущим идею (не более 6). Балл носит информационный характер и в итоговую оценку группы не входит.

ИИ-анализ занятия и предварительное оценивание

После завершения занятия преподаватель может запросить ИИ-разбор. Модели передаются названия этапов, итоговые списки гипотез и вопросов по каждому этапу и эталон преподавателя. В системной инструкции модель выступает в роли методиста медицинского образования. Она должна вернуть один JSON-объект с тремя полями: текст разбора в формате Markdown, где для каждого этапа выделены разделы

«Положительные качества», «Отрицательные качества» и «Рекомендации»; массив баллов 0–100 по этапам; средний балл. В инструкции заданы якорные ориентиры шкалы: например, «и гипотезы, и вопросы, логика видна» — 50–70 баллов, 85 и выше — «резерв для выдающейся работы». Модели прямо указано не выставлять итоговую оценку вместо преподавателя. Версия инструкции (в пилоте — session-analysis-v4) сохраняется вместе с результатом для воспроизводимости.

Ответ модели проверяется по схеме. Если число оценок не совпадает с числом этапов, баллы отбрасываются. Затем к баллам каждого этапа применяются жёсткие ограничения, которые зависят от фактического объёма работы группы (табл. 2). Ограничения исключают «оценку за намерение», когда модель ставит высокий балл при пустом или минимальном вкладе. Средний балл пересчитывается заново. Число входных и выходных токенов каждого обращения записывается в журнал, что позволяет учитывать затраты.

Таблица 2. Жёсткие ограничения предварительного балла этапа

Работа группы на этапе (h — гипотезы, q — вопросы)	Максимальный балл	Обоснование
$h = 0$ и $q = 0$	0	нет вклада — нет баллов
$h = 1–2$, $q = 0$	20	узкий дифференциальный ряд без плана обследования
$h \geq 3$, $q = 0$	40	гипотезы не подкреплены поиском данных
$h = 0$, $q \geq 1$	25	сбор данных без диагностической гипотезы
$h \geq 1$ и $q \geq 1$	без ограничения (по оценке модели)	полноценная работа этапа

Итоговую оценку (по шкале 0–100 или в иной форме) и комментарий выставляет преподаватель, просматривая разбор, траектории гипотез и вклад отдельных студентов. После утверждения оценка фиксируется с отметкой времени.

Аналитика

Для преподавателя и заведующего кафедрой платформа формирует сводные показатели: число занятий и их статус, длительность занятий и этапов (занятия длительностью более 8 часов исключаются как незакрытые), среднее число гипотез и вопросов по этапам, распределение оценок преподавателя, БЯМ и студентов, рейтинг активности студентов. Администратор видит сводку по всем кафедрам и затраты на ИИ.

Сбор данных и статистический анализ

Данные периода апробации выгружены из базы платформы одним SQL-запросом в обезличенном агрегированном виде. Учитывались все занятия, проведённые в указанный период. Длительность занятия рассчитывали только для завершённых занятий продолжительностью от 1 до 480 мин. Длительность этапа определяли как интервал между его открытием и отправкой. Линию гипотезы считали изменённой, если в разных этапах встречалось более одной уникальной формулировки с одним идентификатором линии. Ввиду малого объёма выборки применялись только описательные методы: абсолютные числа, доли, среднее (M), стандартное отклонение (SD) и медиана (Me). Согласованность оценок БЯМ и преподавателя описывалась попарно, без проверки статистических гипотез.

Этические аспекты

Исследование основано на анализе обезличенных данных учебного процесса. Студенты при подключении указывают только имя (без фамилии и иных персональных

данных). В языковую модель передаются только тексты гипотез и вопросов, а сведения о студентах не передаются.

Результаты

Возможности платформы для проведения CBL-занятий

Разработанная платформа поддерживает полный цикл CBL-занятия: подготовку кейса, проведение занятия, разбор и оценку, а также анализ результатов на уровне кафедры. Возможности сгруппированы по участникам учебного процесса.

Для преподавателя: подготовка к занятию

– Конструктор кейса: разбиение на этапы, учебные цели для каждого этапа, отметка финального этапа с раскрытием диагноза.

– Быстрый перенос готового материала: текст из методического пособия вставляется целиком, а языковая модель размечает его как диалог врача и пациента и повествование, без изменения содержания.

– Добавление изображений (рентгенограммы, ЭКГ, фотографии кожных проявлений) в нужный этап.

– Скрытый от студентов эталон — ожидаемые гипотезы и ключевые вопросы, с которыми сравнивается работа группы.

– Привязка кейса к факультетам и курсам, версии кейса, экспорт в презентацию PowerPoint для работы без интернета.

Для преподавателя: во время занятия

– Пошаговая выдача информации: на экране аудитории виден только текущий этап, следующий открывается одним действием, и студенты не могут «заглянуть вперёд».

– Электронная доска гипотез и вопросов: преподаватель вносит предложения группы, а при переходе к новому этапу список переносится автоматически, ничего не стирается.

– Подключение всей группы по QR-коду: каждый студент может прислать свою гипотезу или вопрос с телефона, в том числе те, кто не решается говорить вслух.

– «Входящие» идеи студентов: преподаватель видит, кто что предложил, и одним нажатием переносит удачные идеи в общий список группы.

– Учёт времени по этапам и отображение хода занятия: какой этап открыт, сколько гипотез и вопросов уже сформулировано.

Для преподавателя: разбор и оценка

– Линии гипотез: наглядная картина того, какие гипотезы группа выдвинула сразу, какие уточнила, а от каких отказалась и на каком этапе. Это готовая основа для дебрифинга и обсуждения ошибок мышления.

– Черновик разбора от ИИ: по каждому этапу положительные стороны, недостатки и рекомендации, а также ориентировочный балл. Преподаватель экономит время на подготовку обратной связи и сам решает, что из этого использовать.

– Вклад каждого студента: число присланных гипотез и вопросов, принятые идеи, ориентировочный индивидуальный балл.

– Итоговая оценка и комментарий преподавателя, которые сохраняются вместе с занятием.

Для студентов

– Вход на занятие за несколько секунд, без регистрации и установки приложений: достаточно сканировать QR-код и ввести имя.

– Возможность высказаться каждому: собственные гипотезы и вопросы отправляются с телефона, а студент видит, была ли его идея принята в общий список.

– Работа в условиях, приближенных к клиническим: данные о пациенте поступают постепенно, как на реальном приёме, и требуют пересмотра гипотез.

– Структурированная обратная связь после занятия: что получилось, что нет и что делать в следующий раз, по каждому этапу кейса.

Для кафедры и администрации

– Банк кейсов кафедры, который можно повторно использовать в разных группах, на разных факультетах и курсах.

– Аналитика: число проведённых занятий, их длительность, активность групп и студентов, распределение оценок преподавателя и ИИ, сравнение групп по одному кейсу.

– Контроль затрат на ИИ и работа платформы без ИИ в автономном режиме, когда нет ключа доступа к языковой модели.

Характеристика апробации

В период апробации в системе были зарегистрированы 2 кафедры, 3 факультета (медицинский, педиатрический, стоматологический), курсы с 4-го по 6-й и 8 учебных групп. Занятия вели 3 пользователя: 1 преподаватель и 2 администратора. Для апробации использовался один опубликованный кейс из 7 этапов. Все 7 текстовых блоков кейса были размечены с помощью БЯМ. Проведено 7 занятий, из них 6 завершены, а 1 на момент выгрузки оставалось незавершённым. Основные показатели приведены в таблице 3.

Таблица 3. Показатели апробации платформы (23.03–01.05.2026)

Показатель	Значение
Кафедры / факультеты / учебные группы	2 / 3 / 8
Курсы обучения	4, 5, 6
Пользователи, ведущие занятия (преподаватель / администратор)	1 / 2
Кейсы (опубликованные); число этапов в кейсе	1 (1); 7
Текстовые блоки, размеченные БЯМ	7 из 7
Занятия: всего / завершены / в процессе	7 / 6 / 1
Длительность занятия, мин: Me; M (n = 3)	59,4; 48,6
Длительность этапа, мин: Me (n = 33)	1,7
Линии гипотез / линии вопросов, всего	49 / 57
Линии гипотез на занятие / линии вопросов на занятие, Me	2,5 / 2
Линии гипотез с изменённой формулировкой	0
Занятия со входом по QR-коду; студенты; идеи (гипотезы / вопросы)	1; 2; 4 (3 / 1)
Идеи, принятые ведущим	3 (75%)
Занятия с ИИ-разбором; балл БЯМ, M ± SD	3; 19,7 ± 21,2
Занятия с числовой оценкой преподавателя; оценка, M ± SD	4; 41,3 ± 16,5
Пары «БЯМ — преподаватель» (баллы)	44 — 60; 10 — 25
Токены на один ИИ-разбор (вход / выход); затраты	1088 / 2424; ≈ 0,0016 долл. США

Медиана длительности завершённого занятия (n = 3 с корректно зафиксированным временем) составила 59,4 мин, среднее — 48,6 мин. Медиана длительности этапа (n = 33) составила 1,7 мин. Короткая длительность этапа объясняется, вероятно, тем, что часть обсуждения проходила устно, а ведущий вносил итоговые гипотезы и вопросы в систему перед переходом к следующему этапу.

Всего группы сформулировали 49 линий гипотез и 57 линий вопросов, медиана на одно завершённое занятие — 2,5 и 2 соответственно. Динамика среднего числа гипотез и вопросов по этапам кейса показана на рисунке 3. На первых пяти этапах в среднем было 3,1–3,7 гипотезы, а на 6-м и 7-м этапах их число снизилось до 2,4. Число уточняющих вопросов, напротив, выросло с 2,8–3,2 на первых этапах до 3,8–4,4 на 5–7-м этапах. Ни одна из 49 линий гипотез не изменила формулировку: группы добавляли новые гипотезы и удаляли неподтвердившиеся, но не уточняли существующие.

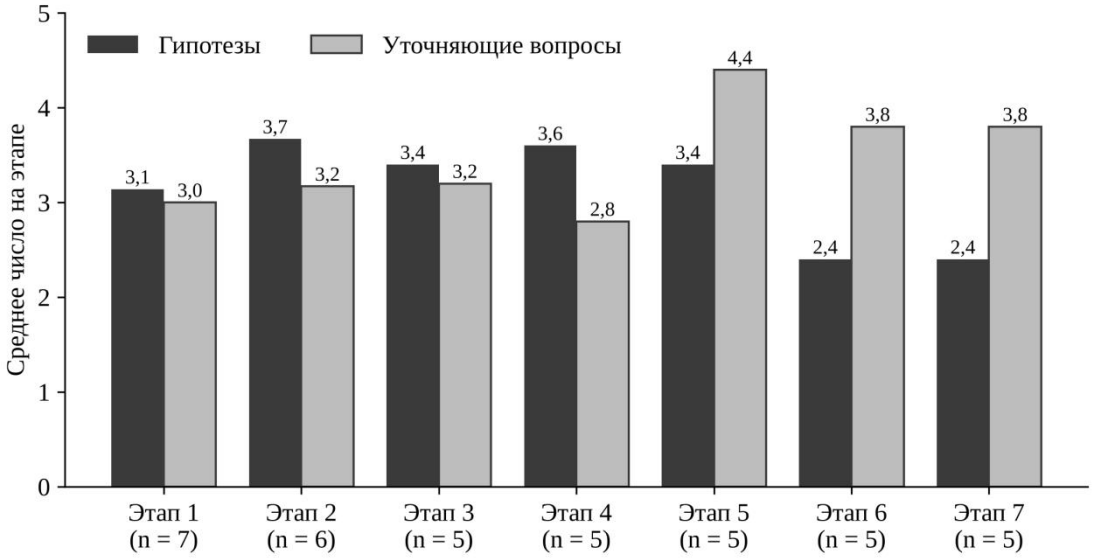

Рисунок 3. Среднее число гипотез и уточняющих вопросов на этапах кейса (по всем занятиям; n — число занятий, дошедших до этапа)

Вход по QR-коду опробован в одном занятии: 2 студента прислали 4 идеи (3 гипотезы и 1 вопрос), из которых ведущий принял в общий список 3 (75%).

Предварительные оценки языковой модели и оценки преподавателя

ИИ-разбор с баллами получен для 3 занятий. Средний предварительный балл составил 19,7 ± 21,2, что соответствует строгой шкале, заложенной в инструкцию и ограничения. Итоговую числовую оценку преподаватель выставил в 4 занятиях, средняя оценка — 41,3 ± 16,5 балла. Обе оценки имелись в 2 занятиях: 44 балла БЯМ против 60 баллов преподавателя и 10 против 25. В обоих случаях преподаватель оценил работу выше модели, на 16 и 15 баллов соответственно. Порядок занятий по уровню работы при этом совпал. Объём данных не позволяет провести статистическую оценку согласованности.

В журнале расходов, который введён в конце периода апробации, зарегистрировано 1 обращение к модели: 1088 входных и 2424 выходных токена. При тарифе 0,15 и 0,60 долл. США за 1 млн токенов прямые затраты на один ИИ-разбор составили около 0,0016 долл. США.

Обсуждение

Разработанная платформа сочетает три элемента, которые в существующих решениях обычно разделены: структуру клинического кейса с последовательным раскрытием, интерактивное участие всей аудитории и автоматизированную аналитику с элементами ИИ. Апробация показала, что платформа технически работоспособна и применима на занятиях разных факультетов и курсов.

Платформа отвечает на основные трудности проведения CBL-занятия, перечисленные во введении. Подготовка кейса ускоряется за счёт переноса готового текста и его автоматической разметки. Порционная выдача информации обеспечивается самой структурой этапов. Запись гипотез и вопросов больше не стирается, а переносится между этапами и сохраняется после занятия. Участие всех студентов поддерживается входом по QR-коду. Подготовку обратной связи облегчают черновик ИИ-разбора и линии гипотез. Сопоставимость оценивания между группами обеспечивают единые кейсы, единая шкала и аналитика кафедры. При этом роль преподавателя как модератора дискуссии сохраняется: платформа берёт на себя техническую часть, а педагогические решения остаются за ним.

Ключевая особенность платформы — фиксация хода рассуждения группы по этапам. Литература по клиническому мышлению подчёркивает, что обучать и оценивать нужно процесс, а не только итоговый диагноз [2, 3, 6]. Уже на небольшом объёме данных видна закономерность, которая соответствует гипотетико-дедуктивной модели [1]: на первых этапах группы удерживали широкий дифференциальный ряд (3–4 гипотезы), а по мере поступления данных сужали его, одновременно увеличивая число уточняющих вопросов. Такие данные могут служить основой для целенаправленной обратной связи, которая относится к наиболее сильным факторам обучения [24], и для обсуждения когнитивных ошибок, например преждевременного закрытия [4, 5].

Вместе с тем ни одна гипотеза не была переформулирована: группы предпочитали удалять неподтвердившиеся гипотезы и добавлять новые. Возможное объяснение — интерфейс, в котором удалить гипотезу и добавить новую проще, чем отредактировать существующую. С методической точки зрения уточнение формулировки (например, «пневмония» → «внебольничная правосторонняя нижнедолевая пневмония») — важный элемент клинического рассуждения. Поэтому в следующей версии планируется явное действие «уточнить гипотезу» и соответствующая подсказка преподавателю.

Вход по QR-коду без регистрации снимает организационный барьер и позволяет каждому студенту внести вклад. Это согласуется с данными об эффективности систем опроса аудитории для вовлечения [11, 12]. В отличие от типичных систем опроса, идеи студентов привязаны к конкретному этапу кейса и могут быть приняты в общий список группы, то есть становятся частью совместного рассуждения. В апробации эта функция использовалась ограниченно (1 занятие), так как она была добавлена в конце периода. Её применение в полноценных группах — задача следующего этапа.

Предварительные баллы БЯМ в обоих случаях с парными оценками оказались ниже оценки преподавателя. Это отражает сознательно выбранную стратегию: модели-«судьи» склонны к завышению оценок [20], поэтому в инструкцию включены якорные ориентиры строгой шкалы, а результат дополнительно ограничивается правилами, не зависящими от модели. Систематическое расхождение в сторону более низких баллов указывает на возможную избыточную строгость. После накопления данных её можно скорректировать калибровкой шкалы. Для сравнения: в исследовании Morjaria L. и соавт. корреляция оценок ChatGPT с оценками экзаменатора при проверке коротких ответов студентов-медиков составила 0,6–0,7 [19]. Мы сознательно не используем БЯМ как самостоятельного экзаменатора: принципы этичного использования ИИ в образовании требуют, чтобы решения, влияющие на обучающегося, принимал человек [21, 22]. В нашей архитектуре модель готовит черновик разбора и ориентировочный балл, а преподаватель утверждает или изменяет оценку.

Важное практическое преимущество — низкая стоимость. Благодаря компактному входу (только списки гипотез и вопросов) затраты на ИИ-разбор одного занятия составляют доли цента, что делает решение доступным для вузов с ограниченным бюджетом. Платформа сохраняет работоспособность и без ИИ, что важно при перебоях интернет-соединения или ограничениях доступа к зарубежным сервисам.

Ограничения

Исследование имеет существенные ограничения. Апробация проведена на одном кейсе, с небольшим числом занятий (7) и ведущих (3), поэтому результаты носят предварительный, описательный характер. Отсутствие контрольной группы не позволяет судить о влиянии платформы на учебные результаты. Для этого необходимо сравнительное исследование с оценкой клинического мышления по валидизированным методикам [25]. Пары «оценка БЯМ — оценка преподавателя» имелись лишь для 2 занятий. В период апробации модель менялась, а журнал расходов вёлся не с начала периода. Жёсткие ограничения баллов учитывают количество гипотез и вопросов, но не их клиническую обоснованность. Ответы БЯМ недетерминированы. Оценка проводилась на русскоязычном материале.

С учётом этих ограничений спланировано проспективное исследование. В нём будут использованы несколько кейсов разных дисциплин и не менее 30 занятий с парными оценками, а модель и версия инструкции будут зафиксированы. Согласованность будет оцениваться по внутриклассовому коэффициенту корреляции [26] и методу Бланда–Альтмана [27], дополнительно будет проведено анкетирование студентов и преподавателей.

Заключение

1. Разработана веб-платформа, которая поддерживает все этапы CBL-занятия. Для преподавателя она включает конструктор кейсов с ИИ-разметкой, порционную выдачу информации, электронную доску гипотез и вопросов с отслеживанием их линий, ИИ-черновик разбора и аналитику кафедры. Для студентов она обеспечивает участие со смартфона по QR-коду без регистрации и структурированную обратную связь.
2. Апробация на 2 кафедрах (3 факультета, 8 учебных групп 4–6 курсов, 7 занятий) подтвердила техническую реализуемость платформы и её применимость в учебном процессе.
3. Платформа делает ход клинического рассуждения группы наблюдаемым: зафиксированы 49 линий гипотез и 57 линий вопросов, а также сужение дифференциального ряда к завершающим этапам кейса при росте числа уточняющих вопросов.
4. Предварительные баллы языковой модели с жёсткими ограничениями оказались строже оценок преподавателя. Объём данных недостаточен для статистической оценки их согласованности, которая будет выполнена в проспективном исследовании. Затраты на ИИ-разбор одного занятия составляют доли цента.

Литература/References

1. Elstein AS, Shulman LS, Sprafka SA. Medical problem solving: an analysis of clinical reasoning. Cambridge (MA): Harvard University Press; 1978:330.
2. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. N Engl J Med. 2006;355(21):2217-25. doi: 10.1056/NEJMra054782
3. Kassirer JP. Teaching clinical reasoning: case-based and coached. Acad Med. 2010;85(7):1118-24. doi: 10.1097/acm.0b013e3181d5dd0d
4. Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Arch Intern Med. 2005;165(13):1493-9. doi: 10.1001/archinte.165.13.1493
5. Croskerry P. A universal model of diagnostic reasoning. Acad Med. 2009;84(8):1022-8. doi: 10.1097/ACM.0b013e3181ace703
6. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. Med Educ. 2015;49(10):961-73. doi: 10.1111/medu.12775
7. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. Med Teach. 2012;34(6):e421-44. doi: 10.3109/0142159X.2012.680939

8. McLean SF. Case-based learning and its application in medical and health-care fields: a review of worldwide literature. J Med Educ Curric Dev. 2016;3:JMECD.S20377. doi: 10.4137/JMECD.S20377
9. Wood DF. Problem based learning. BMJ. 2003;326(7384):328-30. doi: 10.1136/bmj.326.7384.328
10. Srinivasan M, Wilkes M, Stevenson F, Nguyen T, Slavin S. Comparing problem-based learning with case-based learning: effects of a major curricular shift at two institutions. Acad Med. 2007;82(1):74-82. doi: 10.1097/01.ACM.0000249963.93776.aa
11. Kay RH, LeSage A. Examining the benefits and challenges of using audience response systems: a review of the literature. Comput Educ. 2009;53(3):819-27. doi: 10.1016/j.compedu.2009.05.001
12. Caldwell JE. Clickers in the large classroom: current research and best-practice tips. CBE Life Sci Educ. 2007;6(1):9-20. doi: 10.1187/cbe.06-12-0205
13. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med. 2010;85(10):1589-602. doi: 10.1097/ACM.0b013e3181edfe13
14. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the Digital Health Education Collaboration. J Med Internet Res. 2019;21(7):e14676. doi: 10.2196/14676
15. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198. doi: 10.1371/journal.pdig.0000198
16. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023;9:e45312. doi: 10.2196/45312
17. Abd-alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ. 2023;9:e48291. doi: 10.2196/48291
18. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical education: potential impact and opportunity. Acad Med. 2024;99(1):22-7. doi: 10.1097/ACM.0000000000005439
19. Morjaria L, Burns L, Bracken K, Levinson AJ, Ngo QN, Lee M, et al. Examining the efficacy of ChatGPT in marking short-answer assessments in an undergraduate medical program. Int Med Educ. 2024;3(1):32-43. doi: 10.3390/ime3010004
20. Zheng L, Chiang WL, Sheng Y, Zhuang S, Wu Z, Zhuang Y, et al. Judging LLM-as-a-judge with MT-Bench and Chatbot Arena. Adv Neural Inf Process Syst. 2023;36:46595-623.
21. Masters K. Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158. Med Teach. 2023;45(6):574-84. doi: 10.1080/0142159X.2023.2186203
22. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: WHO; 2024. 98 p.
23. DeepSeek-AI. DeepSeek-V3 technical report [Internet]. arXiv:2412.19437. 2024. doi: 10.48550/arXiv.2412.19437
24. Hattie J, Timperley H. The power of feedback. Rev Educ Res. 2007;77(1):81-112. doi: 10.3102/003465430298487
25. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. Acad Med. 2019;94(6):902-12. doi: 10.1097/ACM.0000000000002618

26. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155-63. doi: 10.1016/j.jcm.2016.02.012
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10.

Сведения об авторах

Бакоев Фаррух Саиджонович — аспирант кафедры детских болезней № 1 им. профессора Х.Дж. Аминова, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино»; ORCID ID: 0000-0002-7498-2938; SPIN-код: 7365-3484; Author ID: 1079409; e-mail: baqoev.farrukh.saidjonovich@gmail.com

Юсупов Навруз Абдусаматович — ассистент кафедры оториноларингологии им. профессора Ю.Б. Исхоки, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино»;

Зарифи Нурали Абуракиб — доцент кафедры неврологии, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино»; ORCID ID: 0000-0001-5494-1218; SPIN-код: 8489-6082; Author ID: 1054132; E-mail: nur.tj-94@mail.ru

Адрес для корреспонденции

Бакоев Фаррух Саиджонович — аспирант кафедры детских болезней № 1 им. профессора Х.Дж. Аминова, ГОУ «Таджикский государственный медицинский университет им. Абуали ибни Сино»; 734026, Республика Таджикистан, г. Душанбе, ул. Сино, 29-31; тел.: +992 (900) 446498; e-mail: baqoev.farrukh.saidjonovich@gmail.com