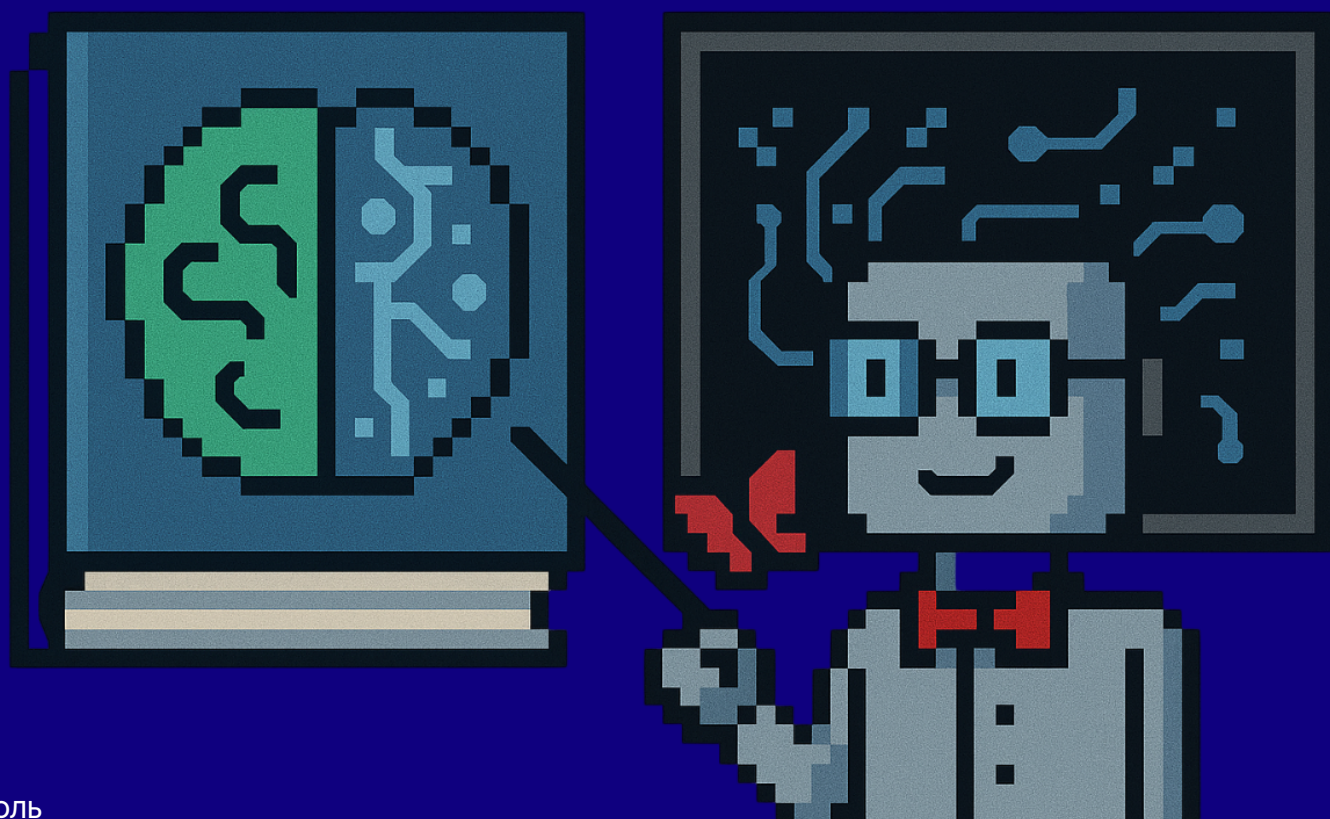


ИИ вместо профессора: прорыв в образовании



Авторы

Андрей Щербенок

директор Школы перспективных исследований ТюмГУ, директор программ бакалавриата Школы управления SKOLKOVO

Александр Диденко

руководитель Лаборатории ИИ Школы управления SKOLKOVO

Ульяна Раведовская

директор Центра образовательных разработок на основе технологий искусственного интеллекта ТюмГУ

Ася Филатова

научный сотрудник Центра образовательных разработок на основе технологий искусственного интеллекта ТюмГУ

Александр Балакир

инженер по машинному обучению Лаборатории ИИ Школы управления SKOLKOVO, аспирант ИБДА РАНХиГС

Владислав Запылихин

инженер по машинному обучению Лаборатории ИИ Школы управления SKOLKOVO, аспирант ИБДА РАНХиГС

Анна Шабанова

исследователь-координатор Лаборатории ИИ Школы управления SKOLKOVO

Софья Матвеева

медиатор, ассистент Школы перспективных исследований ТюмГУ

Оглавление

Введение	4		
Executive summary	5		
Исследование	6		
Эксперимент 1			
ИИ-профессор в действии: как искусственный интеллект меняет обучение	7		
Эксперимент 2			
Научная точность в ответах ИИ-моделей: какую LLM выбрать?	11		
Эксперимент 3			
ИИ-оценщик: когда машина может быть субъективной?	14	ПРИЛОЖЕНИЕ 1	
Эксперимент 4		Графики второго эксперимента	20
Как архитектура модели влияет на качество ответов: роль междисциплинарного подхода и RAG	16	Графики третьего эксперимента	22
		ПРИЛОЖЕНИЕ 2	
		Критерий оценки «Конкретность»	23
Заключение	18	Источники	24

Введение

Технологии искусственного интеллекта (ИИ) стремительно меняют образовательный ландшафт и предполагают множество вариантов использования — от персонализированных помощников до автоматизированных систем оценки знаний. [Согласно совместному исследованию Яндекса и ВШЭ](#), 40% преподавателей признают, что генерируют программы дисциплин с помощью GPT, а еще 20% — активно используют ИИ для создания заданий по их предметам. Студенты также активно пользуются ИИ: больше половины обучающихся сдавали задания с его помощью и около 60% использовали его для ВКР.

Несмотря на активное проникновение ИИ в образовательный процесс, до сих пор его использование ограничивалось вспомогательными функциями. Однако полноценный ИИ-преподаватель, который берет на себя роль лектора и становится единственным источником знаний, — это принципиально новый этап, который был реализован в Школе управления СКОЛКОВО, в коллаборации с ТюмГУ.

В 2024 году в рамках бизнес-бакалавриата Школой управления СКОЛКОВО и МФТИ был запущен курс «Биология поведения человека», где вместо профессора обучением студентов занимался ИИ-ассистент Роберт. Он был разработан Лабораторией ИИ Школы управления СКОЛКОВО и ТюмГУ, и его познания в биологии, генетике и нейроэндокринологии не уступали уровню экспертов. ИИ-профессору помогал человек-медиатор. Уникальное отличие этого эксперимента заключалось в том, что медиатор лишь направлял процесс, но не имел экспертных знаний по предмету, что делало ИИ единственным источником информации.

Как это было? Вместо лекций в привычном формате студенты взаимодействовали с Робертом, изучая ключевые принципы биологии поведения, механизмы влияния генетики и среды на личность, а также взаимосвязь эндокринной и нервной систем. Медиатор направлял процесс, помогал студентам глубже погружаться в тему и поддерживал их интерес к обучению.

В ходе работы исследователи Лаборатории ИИ Школы управления СКОЛКОВО оценили, как ИИ-персона повлияла на освоение профессиональной терминологии студентами, изучили качество создаваемого контента и целесообразность применения языковых моделей (Gemini 1.5 Pro, ChatGPT-4o, Claude 3.5 Sonnet, YandexGPT Pro 4 поколения, GigaChat Pro) в образовании. Определение эффективности изучения материала курса после внедрения ИИ проводилось в сравнении с академическими результатами студентов прошлого потока — когда еще преподавал человек.

Результаты оказались многообещающими: студенты, обучавшиеся с помощью ИИ, показали лучшие по сравнению с их предшественниками, изучавшими предмет традиционными методами, результаты. Более того, исследование позволило определить оптимальные языковые модели, обеспечивающие наилучший баланс между качеством контента и затратами на его генерацию.

Executive summary

1

ИИ ставит под вопрос эффективность традиционных методов обучения и проверки знаний, таких как запоминание и письменные экзамены. Large Language Models (LLM) могут легко генерировать готовые ответы, поэтому важно сосредоточиться не на воспроизведении информации, а на развитии критического мышления, анализе и решении реальных задач. Вместо пассивного обучения, где человек просто слушает и запоминает, нужен подход, при котором ИИ помогает лучше понимать и применять знания на практике. **Безынициативная «позиция слушателя» несовместима с эффективным использованием ИИ.**

Выводы для бизнеса:

Внедрение такого формата полезно не только в школах и вузах, но и в корпоративном обучении. Чтобы люди действительно развивались, важно вовлекать их в поиск решений, работу с данными и самостоятельный анализ. Организации, которые адаптируют свои образовательные программы к этому принципу, подготовят специалистов, способных эффективно использовать ИИ как инструмент, а не просто копировать его ответы.

2

Использование ИИ в роли профессора меняет привычный образовательный опыт, особенно у тех, кто привык, что преподаватель дает ответы на все вопросы. Неопределенность и необходимость самостоятельно искать ответы могут вызывать стресс, дискомфорт и пассивную агрессию к медиатору у студентов. Исследователи заметили, что удаление из образовательного процесса преподавателя провоцирует предвзятость и недоверие к ИИ. При этом субъективные переживания участников могут вступать в противоречие с объективными показателями их образовательных успехов.

Выводы для бизнеса:

При первом внедрении ИИ в корпоративное или иное обучение надо быть готовым к негативной реакции обучающихся на изменение форматов. Следует фиксировать внимание не только на субъективном опыте сотрудников или студентов, но прежде всего — на их образовательных результатах. В условиях сопротивления технологиям внутренние ощущения, основанные на впечатлениях и удовлетворенности, могут искажать оценку эффективности программы. Поэтому важно подсвечивать их позитивные результаты и умение перестроиться на новый формат.

3

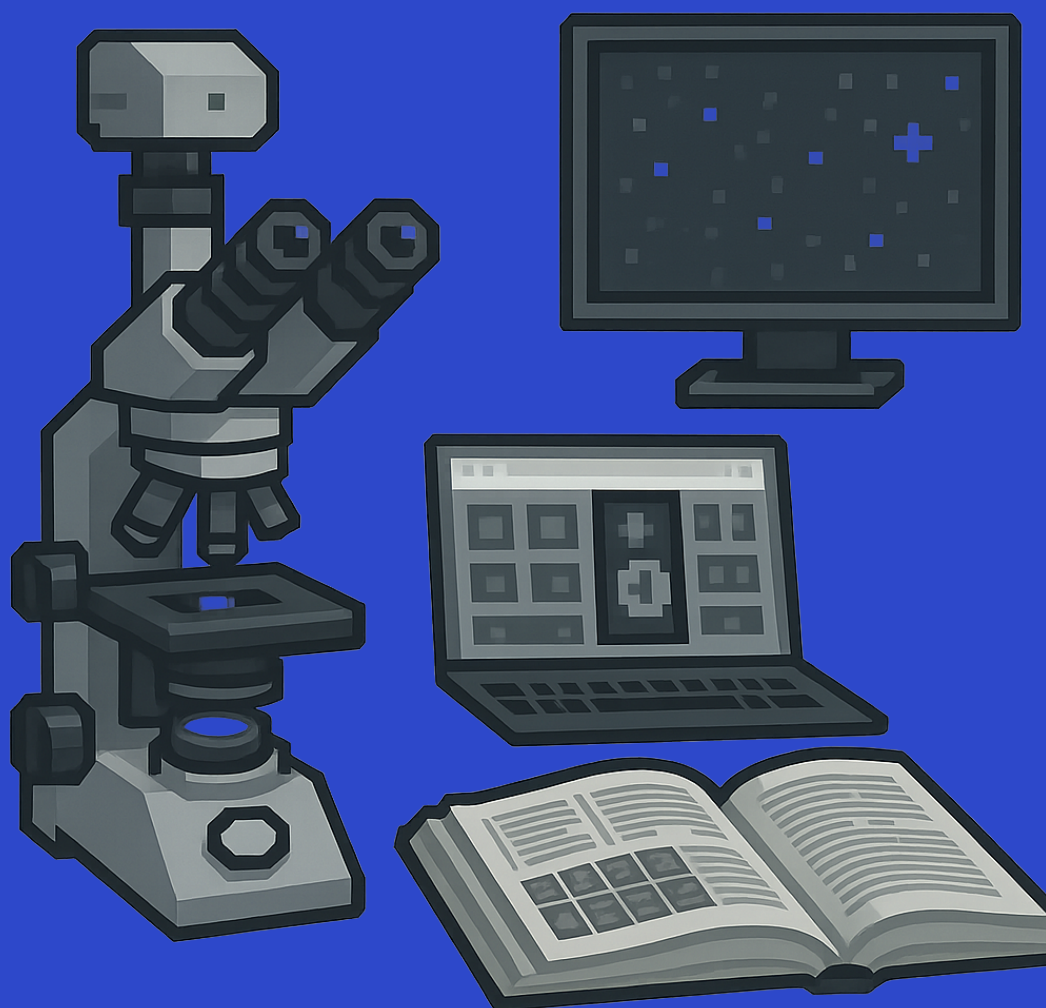
Большие языковые модели демонстрируют различный уровень точности и глубины при выполнении узкопрофильных задач. Поскольку многие проблемы требуют междисциплинарного комплексного подхода, когнитивная архитектура чат-бота (например, основанная на принципах критической оценки и интеграции ответов с разных методологических позиций) может оказаться эффективнее универсальных LLM. Это связано со способностью таких систем адаптироваться к контексту, анализировать данные из разнородных источников и учитывать многоплановые аспекты сложных вопросов.

Выводы для бизнеса:

Для повышения эффективности LLM при решении уникальных бизнес-задач стоит инвестировать в специализированные архитектуры моделей, такие как системы с Retrieval Augmented Generation (RAG), поскольку они способны демонстрировать более высокую продуктивность и точность исполнения, чем готовые решения без дополнительной модификации. При тестировании LLM следует комбинировать экспертные и автоматизированные оценки, чтобы выявлять возможные системные ошибки. Также не стоит забывать о том, что, как и люди, LLM могут ошибаться (галлюцинировать), поэтому всегда надо проверять ответы моделей, чтобы они не повлияли на качество обучения.

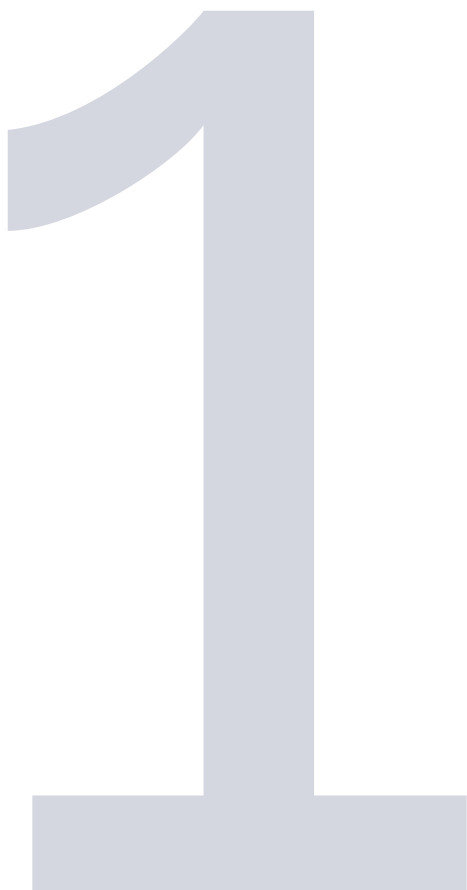


Исследование





ЭКСПЕРИМЕНТ



ИИ-профессор в действии: как искусственный интеллект меняет обучение

Можно ли сделать освоение сложных дисциплин более эффективным с помощью искусственного интеллекта? Исследователи решили проверить это в рамках интенсивного курса «Биология поведения человека» и провели эксперимент, в котором сравнили два формата обучения — традиционный и инновационный. Два потока курса идеально подошли для сравнения. Они были практически идентичны: оба интенсива проводились для 42 первокурсников, которые не обладали обширными знаниями о предмете до поступления, никто из участников не сдавал ЕГЭ по биологии, а опыт работы с ИИ был ситуационным, характерным для выпускников школ последних лет. Длительность обучения

составила шесть 7-часовых учебных дней и в 2023-м, и в 2024-м. Единственная измененная переменная эксперимента — формат обучения. К чему это привело, рассмотрим ниже.

В 2023 году курс «Биология поведения человека» преподавался в традиционном формате: лектор читал лекции, проводил семинары, использовал классические методы подачи материала, но уже в следующем году обучение было организовано по принципиально иной модели, в центре которой оказался специализированный ИИ-ассистент Роберт. ИИ-ассистент взаимодействовал со студентами исключительно на русском языке и выполнял роль основного источника информации. Роберт содержал в себе сложную архитектуру, позволяющую ему быть экспертом одновременно во многих дисциплинах, связанных с биологией и поведением (более подробно его архитектура разобрана во втором эксперименте). В рамках курса студенты взаимодействовали с Робертом через различные форматы индивидуальной и групповой работы. Центральным элементом совместной работы было активное вопрошание, которое позволяло студентам проверять и развивать свои размышления: ИИ-ассистент был активным участником свободных дискуссий и структурированных форматов — дебатов, кейс-анализа, семинаров *close-reading* и пленарных сессий, проводимых по заранее подготовленному плану. Во всех этих и других форматах Роберт выступал как интеллектуальный партнер: помогал раскрыть неявные детали биологических процессов, уточнять понятия, формулировать аргументы и углублять анализ обсуждаемых тем.

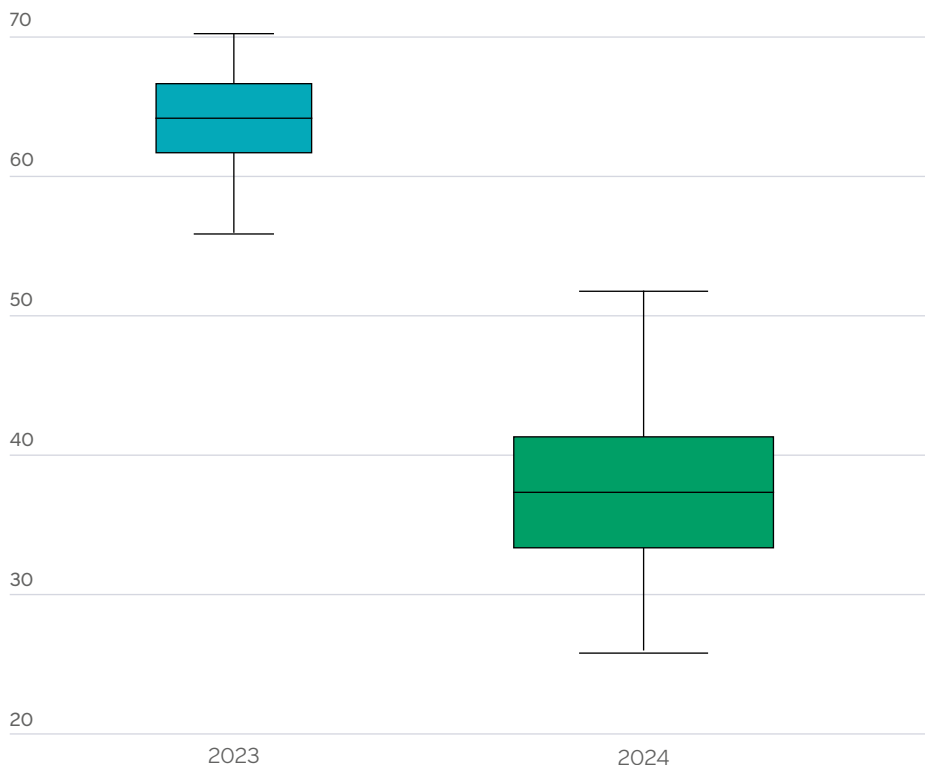
По завершении курса в обеих группах были проведены личные рефлексивные интервью, в ходе которых студентам предлагалось ответить на серию ключевых вопросов, касающихся приобретенных знаний, общего понимания биологии поведения человека как научной дисциплины и осознания влияния биологических механизмов на поведение. Интервью позволяют увидеть не только то, что студенты запомнили, но и то, насколько глубоко они поняли и связали новые знания между собой, в отличие от тестов, проверяющих в основном запоминание фактов (именно поэтому выбор методологов курса пал на них). Собранные ответы были оцифрованы, переведены в текст и переданы для анализа. Для изучения различий между интервью двух групп данные были преобразованы в векторные представления, что позволило провести анализ их семантической близости к профессиональным источникам на основе косинусных расстояний. Подобный подход набирает популярность в академической литературе в последние годы. В частности, в работах Wilson et al. (2022), Wang et al. (2022) и Segarra-Faggioni et al. (2023) методы автоматизированной семантической оценки усвоения знаний сопоставлялись

с экспертными оценками, было показано, что автоматизированная оценка имеет высокую валидность и при этом независима от субъективизма экспертов. В соответствии с этими практиками в данном исследовании семантическая близость текстов студентов к профессиональным источникам интерпретировалась как объективный индикатор не только запоминания фактической информации, но и осмысленного, более глубокого усвоения понятийной структуры дисциплины. В обеих группах эталоном сравнения служили учебные пособия и научно-популярная литература по тематике курса.

Анализ результатов показал, что ответы студентов, обучавшихся с использованием ИИ, отличались статистически значимым сокращением семантического расстояния между их тезаурусом и эталонными текстами. Это означает, что ответы студентов, которым преподавал ИИ, покрывали больше тем, которые были в материале курса, чем те, кто учился с обычным преподавателем. Кроме того, в ИИ-группе наблюдалась большая точность в использовании научной терминологии, а также более полное и детальное воспроизведение ключевых идей курса. На графиках №1 и №2 показаны расстояния между центроидом материалов (иначе говоря, сутью и центром образовательной программы) и центроидом ответов студентов на интервью. Визуализация подтверждает, что между массивом материалов и массивом студенческих ответов расстояние весьма велико, но с внедрением ИИ в учебный процесс в следующем году аналогичное расстояние сократилось **на целых 60%!** Это говорит о том, что в интервью 2024 года студенты охватили больше тем, которые были изучены в течение курса, и смогли более подробно о них рассказать, нежели студенты предыдущего, более «человечного» потока.

График №1.

Сравнение расстояния между центроидами материалов курса и ответами студентов на интервью в 2023 и 2024 году



Однако наряду с положительными результатами выявились и интересные различия в оценке самого формата обучения. В отличие от студентов 2023 года, которые в основном положительно или нейтрально отзывались о курсе, участники эксперимента, обучавшиеся с ИИ, высказывали более критичное отношение к образовательному процессу. Студенты испытывали сложности в переходе на гибридный формат обучения, но это дало свои плоды в виде их академических успехов. Сопротивление новизне ИИ обязательно нужно учитывать во внедрении ИИ-решений.

ЭКСПЕРИМЕНТ



Научная точность в ответах ИИ-моделей: какую LLM выбрать?

Развитие больших языковых моделей открыло новые возможности для генерации научного контента, но остается важный вопрос: насколько эти модели точны в специализированных областях знаний? В рамках исследования ученые решили оценить, какая из популярных LLM — Gemini 1.5 Pro, ChatGPT-4o, Claude 3.5 Sonnet, YandexGPT Pro 4 поколения и GigaChat Pro — наиболее эффективно отвечает на вопросы по биологии поведения человека. Оценку проводили эксперты-биологи, профессионалы в своей отрасли, которых для исследования выбрали коллеги из ТюмГУ.

В исследовании участвовали пять языковых моделей, каждая из которых была интегрирована в чат-бот Роберт. Сложная архитектура состояла из 7 векторных баз данных, основанных на глубоких знаниях по семи дисциплинам: генетика, зоосоциология, нейронаука и эндокринология, молекулярная генетика, патология, поведенческая эволюция и нейролингвистика. Данная структура напоминает консилиум экспертов из разных дисциплин, а архитектура с такими базами данных называется RAG.

Retrieval-Augmented Generation (RAG) — это архитектура искусственного интеллекта, при которой модель не только генерирует ответы, но и обращается к внешним базам знаний для поиска дополнительной информации. Такой подход позволяет получать более точные и содержательные ответы, особенно в сложных и узкоспециализированных темах.

При поступлении вопроса бот оценивает релевантность дисциплин к запросу посредством подсчета средней косинусной близости между ними. После вычислений отбираются только те дисциплины, показатель соответствия которых больше 0,6 по косинусной близости. Чтобы найти компромисс между точностью и скоростью ответа, отбираются топ-3 дисциплины в рейтинге по соответствию вопросу. После этого бот передает вопрос первой векторной базе данных (RAG), получает от первого «эксперта» ответ, который потом передается на критику во вторую, где «эксперт» по второй дисциплине вносит свои комментарии и дополняет ответ с точки зрения своей сферы. Третий RAG работает аналогично. После формирования ответа «консилиумом» экспертных RAG три ответа интегрируются в один текст, и уже этот финальный ответ дается пользователю.

Для обеспечения репрезентативности выборки специалисты по биологии поведения совместно с ИИ составили 1782 вопроса, и это позволило устранить потенциальные тематические пробелы и расширить спектр затрагиваемых проблем. Далее все вопросы предлагались каждому из пяти чат-ботов с разными моделями, а после их ответы анонимизировались и рандомизировались. В результате эксперты получали таблицу, в которой для каждого вопроса предлагалось пять анонимных ответов, от них же требовалось независимо выбрать лучший ответ (никто из экспертов не знал, какая модель стоит за каким ответом).

Оценка проводилась на основе 485 экспертных суждений, по результатам которых рассчитывался средний рейтинг для каждой LLM. В итоговом «турнире» моделей исследователи увидели значимые различия между моделями: **лучшие результаты продемонстрировала YandexGPT**, сгенерировав наиболь-

YandexGPT Pro 4

27,6

Claude 3.5 Sonnet

23,4

ChatGPT-4o

21,0

GigaChat Pro

18,8

Gemini 1.5 Pro

9,2

График №2

Рейтинг больших языковых моделей на основе оценки их ответов экспертами в биологии поведения

шее число ответов, признанных экспертами наиболее релевантными и детальными, а **последнее место заняла модель Gemini**.
Графики по каждому критерию отдельно — в [Приложении 1](#).

Экспертная оценка показала, что YandexGPT, Claude и ChatGPT наиболее подходят для использования в задачах, которые требуют повышенной научной точности и полноты ответа. Тем не менее модели постоянно совершенствуются, и вопрос о том, какие алгоритмы обеспечивают наилучшее качество научного контента, все еще остается открытым.

ЭКСПЕРИМЕНТ



ИИ-оценщик: когда машина может быть субъективной?

А что если ИИ попробует сам оценить свою же работу? Будет ли он более объективен, чем человек-эксперт, или, наоборот, даст себе поблажку? В третьем эксперименте мы узнали, насколько хорош ИИ-оценщик.

В качестве исходного материала использовалась таблица ответов чат-ботов с междисциплинарной архитектурой, полученная в ходе предыдущего эксперимента. Для их автоматической оценки была разработана модель, использующая тщательно подобранные промпты (пример промпта, оценивающего корректность ответов модели, приведен в Приложении 2), ориентированные

на **четыре ключевых критерия**. Первым из них стало соответствие научному стилю речи и способу аргументации — оценивалось, насколько ответы воспроизводят характерные особенности речи, логики объяснения и аргументации, свойственные научным работам. Вторым критерием выступила междисциплинарность — качественный ответ должен содержать отсылки к различным концептуальным рамкам и научным дисциплинам. В дополнение к этому анализировалась понятность — насколько ответ хорошо структурирован и доступен для понимания, а также его конкретность, то есть точность формулировок и степень соответствия поставленному вопросу. ИИ-оценщиком выступила **GPT-4o Mini**. Модель была выбрана, поскольку хорошо воспринимает инструкции, а также настроена на работу с выводом ответа в формате JSON. **По аналогии с оценкой экспертов, модель не знала, какой ответ был дан какой моделью, и оценивала независимо.**

По итогам оценки в общем рейтинге **наибольшее число высоких оценок получила модель ChatGPT**, тогда как **YandexGPT показала наименьший результат**. Однако при детальном анализе критериев выявились значимые различия: **GigaChat превзошел другие модели по понятности и конкретности ответов, а Claude продемонстрировал наибольшее соответствие научному стилю речи и способу аргументации.**

Интересно, что обобщенные результаты, полученные в ходе работы экспертной группы и GPT-4o Mini, оказались противоречивыми. В то время как эксперты признали YandexGPT моделью, предоставляющей наиболее качественные ответы, GPT-4o Mini отдал предпочтение своему сородичу — ChatGPT-4o. Этот разрыв в выводах поднимает важный вопрос о границах использования языковых моделей не только как источников информации, но и как инструментов ее оценки. При использовании ИИ важно принимать во внимание, что **алгоритмическая оценка ответов остается столь же субъективной, сколь и экспертное мнение.**

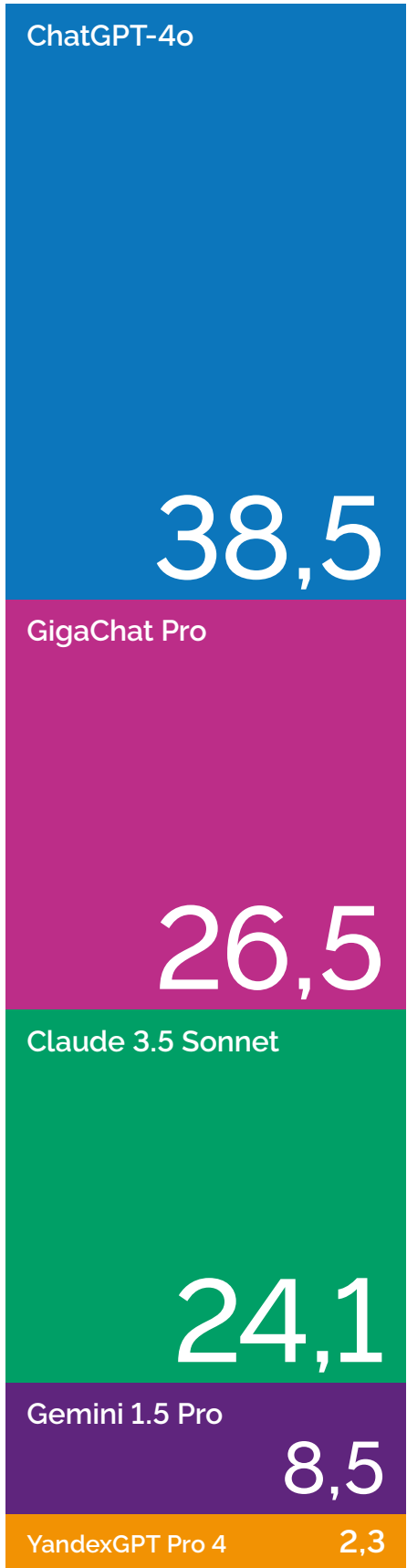


График №3
Суммарный рейтинг LLM по числу побед по всем предложенным критериям.

ЭКСПЕРИМЕНТ



Как архитектура модели влияет на качество ответов: роль междисциплинарного подхода и RAG

Идея проведения следующего эксперимента возникла в ходе второго потока курса «Биология поведения человека». Стало очевидно, что основным конкурентом ИИ-ассистента Роберта, обладающего специализированной междисциплинарной архитектурой, является не профессор, а семейство универсальных больших языковых моделей от OpenAI. Студенты самостоятельно принимали решение об их использовании, сопоставляли ответы Роберта и моделей OpenAI и часто субъективно оценивали ответы моделей как более интеллектуально доступные, лаконичные и точные. Возник принципиальный вопрос: может ли «чистая» модель без сложной архитектурной надстройки

давать более качественные ответы, чем система, основанная на нескольких Retrieval-Augmented Generation (RAG)? Чтобы проверить это, исследователи провели прямое сравнение модели OpenAI o3-mini-high с победителями предыдущего эксперимента. Для анализа использовался тот же набор вопросов, что и ранее, что обеспечило максимальную сопоставимость данных. Оценка проводилась на основе четырех ключевых критериев (соответствие научному стилю речи и способу аргументации, междисциплинарность, понятность и конкретность) с использованием специально настроенной языковой модели-оценщика. Такой подход позволил не только зафиксировать общее качество ответов, но и выявить, в каких аспектах архитектура оказывает наибольшее влияние на результат.

График №4

Суммарный рейтинг по числу побед по всем предложенным критериям между Робертом с ChatGPT-4o и чистым ChatGPT-o3-mini.

Роберт с ChatGPT-4o

ChatGPT-o3 mini

69,5

30,5

Результаты эксперимента показательны: OpenAI o3-mini-high продемонстрировала результаты хуже как в общей, так и в критериальной оценке по сравнению с лучшими чат-ботами из третьего эксперимента (графики по критериям №10-14 — в Приложении 1). Особенно заметными оказались различия в междисциплинарности и конкретности ответов (графики №11 и №13). Это параметры, в которых специализированные модели с архитектурой, интегрирующей различные концептуальные рамки и критические механизмы, показали явное превосходство. Это подтверждает гипотезу о том, что **наличие сложной когнитивной структуры в языковой модели, включая возможность интеграции знаний из различных дисциплин и использование RAG, значительно улучшает качество генерируемых ответов в отличие от работы «чистой» модели.**

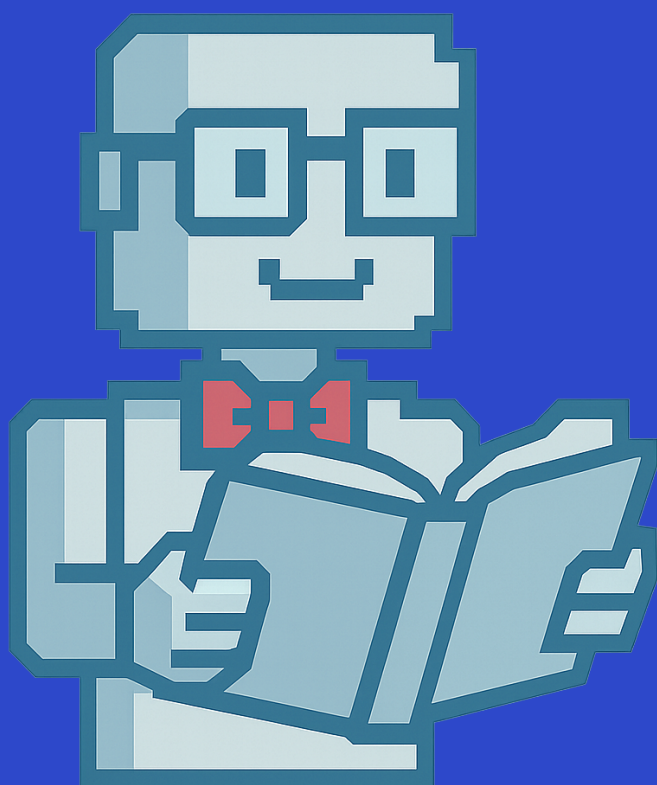
Заключение

ИИ трансформирует саму суть и логику обучения. Эпоха, в которой ценили зубрежку и четкое следование инструкциям, подходит к концу. В условиях цифровой среды и растущей сложности задач становится решающим не то, что человек знает, а то, как он мыслит: способен ли интерпретировать данные, выстраивать причинно-следственные связи, адаптироваться, формулировать решения в условиях неопределенности. ИИ в этой системе — не замена преподавателю или эксперту, а партнёр по мышлению: он провоцирует, усложняет, подталкивает к самостоятельному анализу. Его нельзя использовать пассивно — с ним нужно вступать в продуктивное интеллектуальное взаимодействие. И в этом его главная ценность.

Для бизнеса это не тренд, а необходимость. Перепроектирование корпоративного и иного обучения должно опираться не только на новые технологии, но прежде всего — на новые цели. Не настроить на зазубривание, а научить учиться, задавать вопросы, разбирать противоречивую информацию, строить суждения в сотрудничестве с ИИ. Это требует изменения в поведении и восприятии: необходим отказ от пассивной позиции слушателя и выбор активной роли и в обучении, и в жизни. Да, это вызовет сопротивление, особенно у тех, кто привык к линейной системе, где есть один «правильный» ответ. Но именно те компании, которые преодолеют этот рубеж, создадут у себя среду, где люди не просто осваивают ИИ, а растут за счёт него. А это уже не просто про эффективность обучения — это про устойчивость, инновационность и интеллектуальное превосходство в новой экономике.



Приложения





ПРИЛОЖЕНИЕ 1

Графики второго
эксперимента

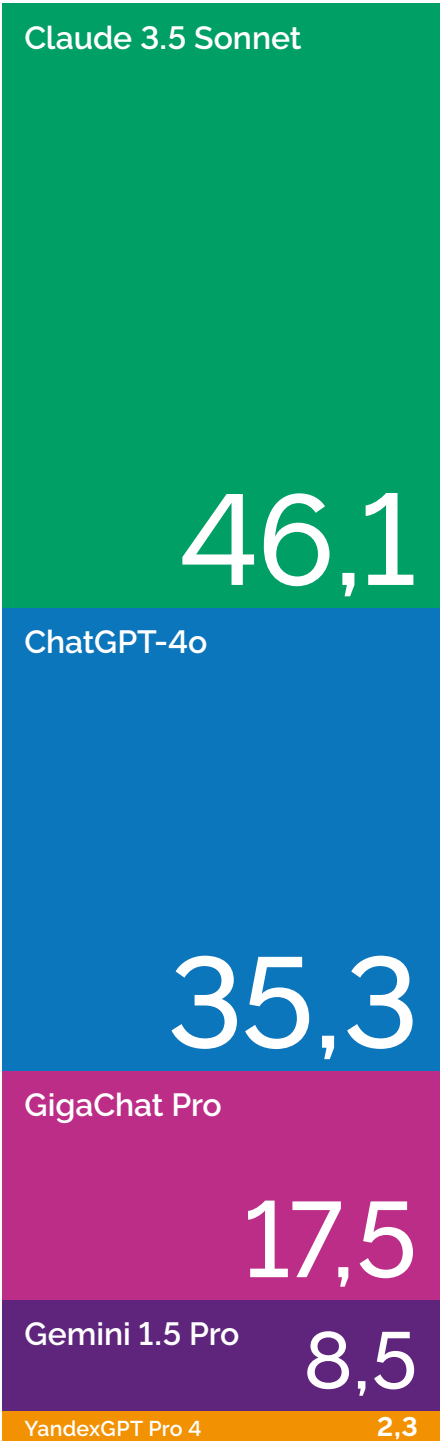


График №5
Рейтинг LLM по числу побед
по критерию «Соответствие
научному стилю речи и способу
аргументации».

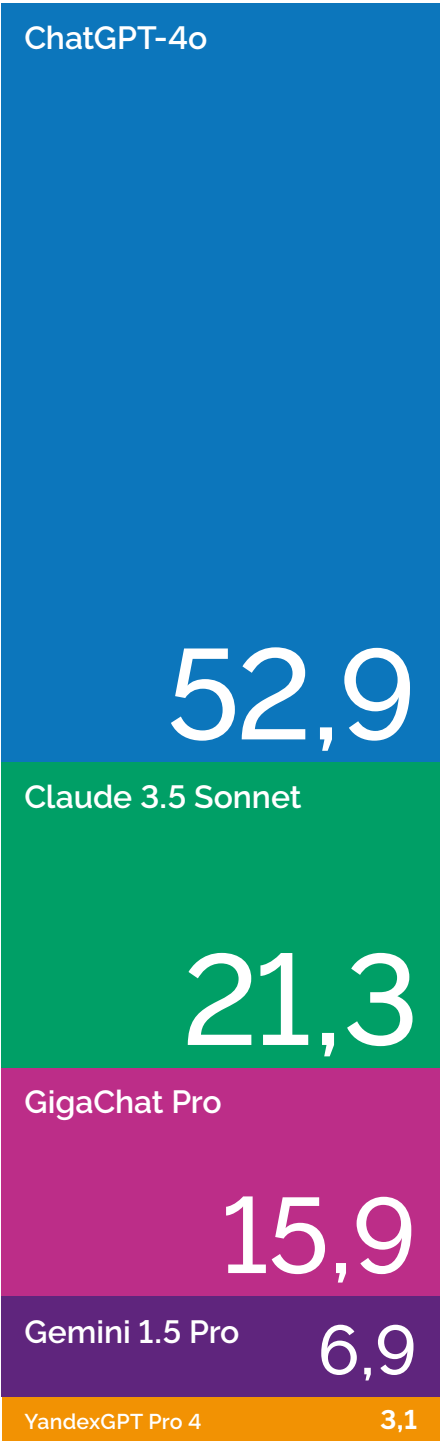


График №6
Рейтинг LLM по числу побед
по критерию
«Междисциплинарность».

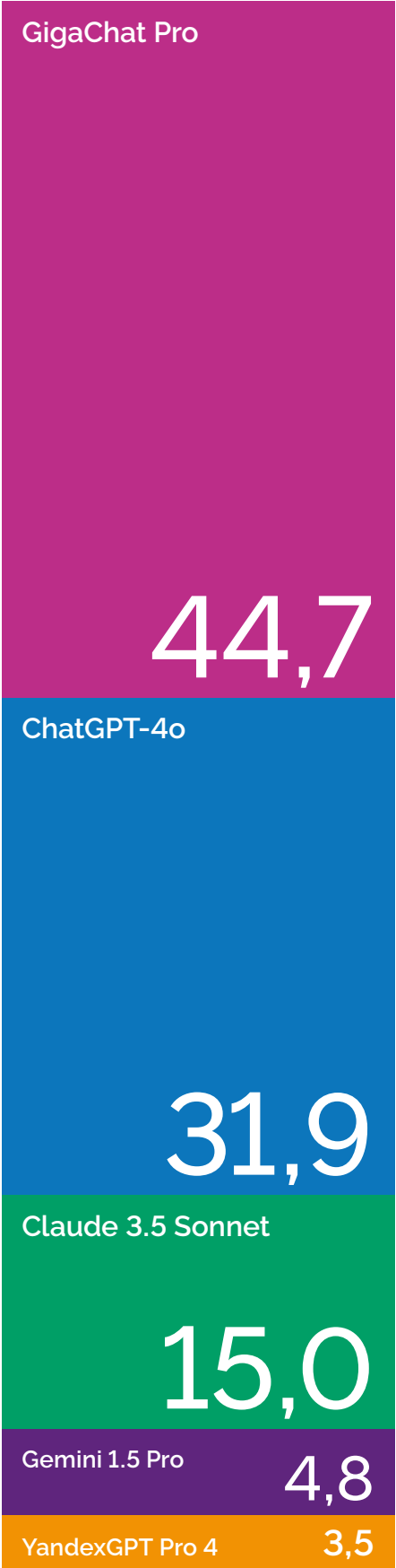


График №7
Рейтинг LLM по числу побед
по критерию «Понятность».



График №8
Рейтинг LLM по числу побед
по критерию «Конкретность»



По четырем ключевым критериям результаты были следующие:

Графики третьего эксперимента

График №9

Рейтинг LLM по числу побед по критерию «Соответствие научному стилю речи и способу аргументации» (победитель рейтинга по критерию из Эксперимента 3 и OpenAI o3-mini-high).



График №10

Рейтинг LLM по числу побед по критерию «Междисциплинарность» (победитель рейтинга по критерию из Эксперимента 3 и OpenAI o3-mini-high).



График №11

Рейтинг LLM по числу побед по критерию «Понятность» (победитель рейтинга по критерию из Эксперимента 3 и OpenAI o3-mini-high)



График №12

Рейтинг LLM по числу побед по критерию «Конкретность» (победитель рейтинга по критерию из Эксперимента 3 и OpenAI o3-mini-high).



ПРИЛОЖЕНИЕ 2

Пример промпта, используемый в 3 эксперименте для определения «Конкретности» ответов моделей:

«Ты — критичный эксперт по оценке качества ответов. Твоя задача — проанализировать шесть ответов на заданный вопрос и оценить их по критерию «конкретность».

Критерий оценки «Конкретность»

Расшифровка критерия: Необходимо оценить, насколько ответ точен, конкретен и соответствует вопросу. Оценивай именно конкретность. Не оценивай никакие другие аспекты данных ответов.

Формат оценивания:

Тебе необходимо выделить лучший ответ и худший ответ и предоставить номер лучшего ответа и номер худшего, а также предоставить свои рассуждения, почему этот ответ является лучшим, а другой худшим.

Рассуждения:

Перед тем как сформулировать итоговый ответ, проверь, рассуждая шаг за шагом:

- 1 Убедись, что ты оценил все пункты данного критерия.
- 2 Проверь, что твои интерпретации и предположения логичны и обоснованы текстом ответа.
- 3 Критикуй себя на каждом шаге рассуждений.
- 4 Подкрепляй свои рассуждения цитатами из ответов.
- 5 Формат ответа:
- 6 Оформи свой анализ и оценки в виде JSON-объекта с полями:

```
{{  
  «best_answer»: «Только номер лучшего ответа»,  
  «worst_answer»: «Только номер худшего ответа»,  
  «reasoning»: «Обоснование выбора лучшего и худшего  
ответа»  
}}
```

ИСТОЧНИКИ

- 1 Segarra-Faggioni, V., Ratté, S., & De Jong, F. (2023). Semantic similarity analysis between the student's reports with the authoritative sources. In Proceedings of the 14th International Conference on Education Technology and Computers (ICETC '22) (pp. 599–605). Association for Computing Machinery. <https://doi.org/10.1145/3572549.3572645>
- 2 Wang, Q. (2022). The use of semantic similarity tools in automated content scoring of fact-based essays written by EFL learners. *Education and Information Technologies*, 27(9), 13021–13049. <https://doi.org/10.1007/s10639-022-11179-1>
- 3 Wilson, J., Pollard, B., Aiken, J. M., Caballero, M. D., & Lewandowski, H. J. (2022). Classification of open-ended responses to a research-based assessment using natural language processing. *Physical Review Physics Education Research*, 18(1), 010141. <https://doi.org/10.1103/PhysRevPhysEducRes.18.010141>
- 4 Somers, R., Cunningham-Nelson, S., & Boles, W. (2021). Applying natural language processing to automatically assess student conceptual understanding from textual responses. *Australasian Journal of Educational Technology*, 37(5), 98–115. <https://doi.org/10.14742/ajet.7121>

ailabskolково



Пишем о том,
где место ИИ в бизнесу.

Исследуем и изучаем —
подписывайтесь, чтобы
быть в точке инноваций