

УДК 004.8:517:37.016

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ КАЧЕСТВА РЕШЕНИЙ
БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ОБЩЕГО НАЗНАЧЕНИЯ
В ЗАДАЧАХ МАТЕМАТИЧЕСКОГО АНАЛИЗА

И. В. АСТАШОВА^{1, 2}, А. А. КИПКЕЕВ², Р. Р. МАСНАВИЕВ²

¹Московский государственный университет имени М. В. Ломоносова

²Российский экономический университет имени Г. В. Плеханова
Москва, Россия

Введение

Актуальность исследования обусловлена стремительным развитием технологий искусственного интеллекта (ИИ), в частности больших языковых моделей (LLM), которые становятся одним из главных трендов и катализаторов изменений в образовательной сфере (EdTech) [5]. Так, мировой рынок LLM с 2023 г. по 2025 г. показал рост с 4,5 до 8 млрд долл. США (77 %) [10].

Применение ИИ в обучении открывает новые возможности как для студентов, так и для преподавателей, выступая в роли интеллектуального ассистента в различных направлениях.

1. Для преподавателей ИИ-инструменты являются критически важным ресурсом для экономии времени и снижения профессиональной нагрузки [6]. Внедрение ИИ помогает автоматизировать рутинные задачи, такие как проверка работ, подготовка планов занятий или заполнение документации [5]. Высвобожденное время может быть перенаправлено на более глубокое взаимодействие со студентами.

2. Для студентов ключевое преимущество ИИ заключается в персонализации обучения [3]. Искусственный интеллект способен анализировать индивидуальные способности и темп усвоения материала, составляя на их основе персонализированные ответы, предоставляя мгновенную обратную связь.

В данной работе исследуются возможности ИИ для помощи студентам, а именно ИИ для ассистирования в вопросах математического анализа. Мате-

математический анализ играет фундаментальную роль в образовательном процессе студентов ИТ-направлений, а также математических, физических, экономических, инженерных и ряда других специальностей, является обязательной дисциплиной, на основе которой строятся многие последующие курсы.

Несмотря на широкий потенциал ИИ в образовании, его реалистичность и точность в решении специфических академических задач, требующих высокой строгости и математической логики (таких как задачи математического анализа), остаются предметом дискуссий. Настоящая работа посвящена оценке возможностей ИИ в решении задач математического анализа и выбору лучшей общедоступной LLM для эффективного использования в учебном процессе.

Сравнительное исследование языковых моделей для решения задач математического анализа позволит оценить их потенциал и определить возможности применения в образовательной практике. Кроме того, результаты исследования покажут, насколько модели отечественной разработки отличаются от международных аналогов по качеству решения задач.

Целью исследования является сравнительный анализ возможностей современных больших языковых моделей при решении задач математического анализа для оценки их потенциала в образовательной практике и выявления различий между отечественными и международными разработками.

Для достижения поставленной цели необходимо решить следующие **задачи**.

1. Определить критерии оценки качества решений задач математического анализа, получаемых от языковых моделей.

2. Сформировать репрезентативный набор задач по математическому анализу, охватывающий основные типы заданий.

3. Провести тестирование выбранных языковых моделей на сформированном наборе задач.

4. Сравнить качество ответов моделей.

5. Проанализировать сильные и слабые стороны рассматриваемых языковых моделей.

6. Сделать вывод о целесообразности применения языковых моделей в учебном процессе и определить, насколько отечественные модели отличаются от международных аналогов по качеству решений.

Объектом исследования являются общедоступные популярные большие языковые модели (LLM) общего назначения.

Предметом исследования является качество решения задач математического анализа современными языковыми моделями.

Для достижения целей исследования и решения поставленных задач использовались следующие **методы**.

1. Экспериментальный метод – проведение тестирования выбранных языковых моделей на специально подготовленном наборе задач.

2. Сравнительный анализ – сопоставление результатов решений разных нейросетей с целью выявления сильных и слабых сторон каждой модели.

3. Качественный анализ ответов – оценка полноты, корректности и структуры решений.

4. Опросы и анкетирование – сбор мнений студентов о восприятии результатов нейросетей и их потенциальной полезности в учебном процессе.

5. Статистическая обработка данных – количественный анализ результатов тестирования для выявления закономерностей и отличий между моделями.

Применение комплексного подхода, включающего как количественные, так и качественные методы, позволяет объективно оценить потенциал языковых моделей и их возможности использования в образовательной практике.

Структура работы. В первом разделе рассматриваются теоретические основы больших языковых моделей и их применения в образовательной и научной деятельности. Даётся обзор современных исследований в области искусственного интеллекта, а также описываются принципы функционирования языковых моделей и их роль в решении учебных задач.

Во втором разделе анализируется математический анализ как учебная дисциплина, рассматриваются типы задач, которые будут использоваться в экспериментальной части исследования, и обосновывается выбор тестового набора заданий. Также описываются критерии оценки решений, используемых нейросетями, и формулируются цели и задачи экспериментального исследования.

В третьем разделе приводится описание проведенного эксперимента, в том числе методики тестирования выбранных моделей (ChatGPT, Gemini, DeepSeek, YandexGPT и GigaChat), анализ полученных решений и их сравнение по различным параметрам. Рассматриваются результаты опросов студентов и преподавателей относительно восприятия работы нейросетей и их потенциального применения в учебном процессе.

В заключении подводятся итоги проведенного исследования, формулируются основные выводы о качестве работы отечественных и иностранных моделей при решении задач математического анализа, а также отмечаются выявленные различия между ними.

Теоретические основы

Выбор больших языковых моделей для сравнительного исследования.

Для проведения оценки потенциала больших языковых моделей (LLM) в решении задач математического анализа был сформирован набор моделей, включающий как ведущие мировые разработки, так и ключевые отечественные аналоги. Критериями отбора являлись: *возможность бесплатного пользования, широкая публичная доступность и популярность* (для оценки применимости в образовательной практике для студентов). Выбранные модели разделены на две

группы: мировые лидеры, задающие стандарты производительности, и крупнейшие отечественные разработки (табл. 1).

Табл. 1. Сравнение выбранных LLM

Модель	Разработчик	Ежемесячное количество активных пользователей (MAU), млн чел.
ChatGPT (GPT-4o)	OpenAI	500–800 [7]
Gemini (2.5 Flash)	Google	400 [9]
DeepSeek (V3.2)	Shēndù Qiúsuǒ	96,88 [8]
GigaChat (2.0)	Sber	19 [4]
YandexGPT (Alice AI*)	Yandex	43 [1]

*Примечание – * – динамически подбирается оптимальный инструмент из семейства Alice AI в зависимости от характера вопроса и требуемого формата ответа (т. е. используется не фиксированная модель)*

Выбранный пул из пяти моделей обеспечивает сбалансированное покрытие: от глобальных лидеров, устанавливающих мировые стандарты производительности (ChatGPT, Gemini), до конкурентоспособной высокопроизводительной модели (DeepSeek) и ключевых российских разработок (GigaChat, YandexGPT), что позволяет провести комплексный и релевантный сравнительный анализ.

Выбор задач для проведения исследования. В качестве источника задач был выбран сборник задач по курсу математического анализа Г. Н. Бермана [2]. Бралось по одной задаче из каждой главы.

1. Функции.
2. Предел. Непрерывность.
3. Производная и дифференциал. Дифференциальное исчисление.
4. Исследование функций и их графиков.
5. Определенный интеграл.
6. Неопределенный интеграл. Интегральное исчисление.
7. Способы вычисления определенных интегралов. Несобственные интегралы.
8. Применения интеграла.
9. Ряды.
10. Функции нескольких переменных. Дифференциальное исчисление.
11. Применения дифференциального исчисления функций нескольких переменных.

12. Многомерные интегралы и кратное интегрирование.

13. Криволинейные интегралы и интегралы по поверхности.

Получился набор из 13 задач легкой и средней сложности (табл. 2) по разным темам математического анализа.

Табл. 2. Список задач

Номер задачи	Номер задачи в [2]	Условие задачи	Ответ
1	26	$F(x) = x^2 + 6$; $\varphi(x) = 5x$. Найти все корни уравнения $F(X) = \varphi(x) $	$x_1 = -3; x_2 = -2; x_3 = 2; x_4 = 3$
2	250	$\lim_{n \rightarrow \infty} \frac{(n+1)^4 - (n-1)^4}{(n+1)^4 + (n-1)^4}$	0
3	476	Продифференцировать функцию $y = \frac{ax+b}{cx+d}$	$\frac{ad+bc}{(cx+d)^2}$
4	1250	Груз весом P , лежащий на горизонтальной плоскости, должен быть сдвинут приложенной к нему силой F . Сила трения пропорциональна силе, прижимающей тело к плоскости, и направлена против сдвигающей силы. Коэффициент пропорциональности (коэффициент трения) равен k . Под каким углом φ к горизонту надо приложить силу F , чтобы величина ее оказалась наименьшей? Определить наименьшую величину сдвигающей силы	$F_{\text{наим}} = kP / \sqrt{1+k^2}$ при $\varphi = \arctg k$
5	1661	Найти производную от функции $y = \int\limits_x^5 \sqrt{1+x^2} dx$ при $x = 0$ и $x = \frac{3}{4}$	-1; -5/4
6	1743	Найти интегралы, воспользовавшись теоремой об инвариантности формул интегрирования: $\int \frac{e^{2x}}{e^{2x} + a^2} dx$	$\frac{1}{2} \ln(e^{2x} + a^2) + C$
7	2281	Вычислить интеграл $\int\limits_0^{\pi} \sin^6 \frac{x}{2} dx$	5/16 π . Полагая $x = 2z$, преобразуем данный интеграл в $2 \int\limits_0^{\pi/2} \sin^6 z dz$

Окончание табл. 2

Номер задачи	Номер задачи в [2]	Условие задачи	Ответ
8	2566	Лемниската $(x^2 + y^2)^2 = a^2(x^2 - y^2)$ вращается вокруг оси абсцисс. Найти объем тела, ограниченного поверхностью, которая при этом получается	$\frac{\pi a^3}{4} \left[\sqrt{2} \ln(1 + \sqrt{2}) - \frac{2}{3} \right]$
9	2803	Определить области сходимости рядов $\ln x + \ln^2 x + \dots + \ln^n x + \dots$	$1/e < x < e$
10	3037	Найти частные производные функции по каждой из независимых переменных: $z = x^3 y - y^3 x$	$\frac{\partial z}{\partial x} = 3x^2 y - y^3$, $\frac{\partial z}{\partial y} = x^3 - 3y^2 x$
11	3242	$f(x, y) = x^3 + 2y^3 - xy$; разложить функцию $f(x + h; y + k)$ по степеням h и k	$x^3 + 2y^3 - xy + h(3x^2 - y) + k(6y^2 - x) + 3xh^2 - hk + 6yk^2 + h^3 + 2k^3$
12	3522	Вычислить интеграл $\iiint_{\Omega} \frac{dx dy dz}{(x + y + z + 1)^3}$, где Ω – область, ограниченная плоскостями $x = 0, y = 0, z = 0, x + y + z = 1$	$\frac{1}{2} \left(\ln 2 - \frac{5}{8} \right)$
13	3770	Вычислить криволинейный интеграл $\int_L \frac{ds}{x - y}$, где L – отрезок прямой $y = 1/2x - 2$, заключенный между точками $A(0, -2)$ и $B(4, 0)$	$\sqrt{5} \ln 2$
Дополнительная задача		Икосаэдр – один из пяти правильных выпуклых многогранников. Дан икосаэдр, каждая из 20 граней которого есть правильный треугольник со стороной $\sqrt{63}$. Пусть A и B – две наиболее удаленные друг от друга в пространстве вершины данного икосаэдра. Минимальная длина пути по поверхности икосаэдра, соединяющего A и B , равна ...	21

В качестве дополнительной была рассмотрена задача из другой области математики – стереометрии.

Математическая модель оценки и ранжирования нейросетей. Для выполнения данного исследования необходимо строго определить его методологическую базу, обозначить критерии и шкалу оценивания решений задач. Составим математическую модель.

Пусть существует множество рассматриваемых нейросетей:

$$N = \{N_j \mid j = 1, 2, \dots, m\},$$

где m – количество нейросетей.

Каждая нейросеть оценивается по множеству параметров качества решения:

$$A = \{A_i \mid i = 1, 2, \dots, k_p\},$$

где A_i – параметр качества решения; k_p – количество таких параметров.

В рамках данной работы определяются три параметра качества решения:

$$A = \{A_1, A_2, A_3\} = \{C, S, M\},$$

где C – полнота решения; S – структурированность и читаемость; M – обоснованность методов.

Каждому параметру A_i ставится в соответствие вес α_i , образуется вектор весов

$$\alpha = \{\alpha_i \mid i = 1, 2, \dots, k_p\},$$

где α_i – параметр качества решения; k_p – количество таких параметров.

В рамках данной работы определяются веса для трех параметров качества решения:

$$\alpha = \{\alpha_1, \alpha_2, \alpha_3\} = \left\{ \frac{4}{9}, \frac{2}{9}, \frac{3}{9} \right\}.$$

Значения оценок модели N_j по параметрам A от респондентов могут принимать целочисленное значение от 1 до 3:

$$A_j = \{C_j, S_j, M_j\}; \quad C_j, S_j, M_j \in \{1, 2, 3\},$$

где 1 – низкий уровень удовлетворенности; 2 – средний уровень удовлетворенности; 3 – высокий уровень удовлетворенности.

Вводится множитель корректности решения $t_{ij} \in \{0, 1\}$, который отражает правильность ответа j -й нейросети для i -й задачи: $t_{ij} = 1$, если ответ правильный, $t_{ij} = 0$, если ответ неверный.

Показатель пригодности P_j j -й нейросети для решения задач курса математического анализа уровня бакалавриата определяется как взвешенная сумма значений параметров $\sum_{i=1}^{k_p} \alpha_i A_i$ с учетом корректности t_i :

$$P_j = \sum_{q=1}^{k_t} \left[t_{jq} \sum_{i=1}^{k_p} \alpha_i A_{jqi} \right],$$

где k_t – количество задач.

Задача ранжирования заключается в упорядочении нейросетей по убыванию показателя P .

Практическая часть

Процедура сбора эмпирических данных. Эмпирический этап исследования был реализован посредством систематического тестирования пяти больших языковых моделей (LLM) на стандартизированном наборе из 13 типовых задач по математическому анализу. Задачи преимущественно относились к начальному и среднему уровням сложности и были выбраны из первых 13 глав классического задачника по математическому анализу. Дополнительная задача взята из банка Открытой международной интернет-олимпиады 2025/26 учебного года (1 тур).

Каждая задача отправлялась модели в отдельном контекстном окне в неизменном текстовом виде, без указания дополнительных пояснений или контекста. За окончательный результат принимался первый сгенерированный ответ, без последующего уточнения запроса или корректировки решения. Это обеспечивало объективную оценку способности моделей к автономному решению академических задач.

В рамках тестирования было получено 65 решений, каждое из которых проходило независимую экспертную оценку. В качестве экспертов выступили четыре студента 4-го курса направления ВШКМиС РЭУ им. Г. В. Плеханова группы 15.27Д-МО11/22б с высокой академической успеваемостью.

Респондент 1 – Куклина Анастасия Игоревна.

Респондент 2 – Мацкевич Софья Сергеевна.

Респондент 3 – Евгеньев Лев Эдуардович.

Респондент 4 – Ваганов Артем Владимирович.

Дополнительная задача рассматривалась отдельно; результаты оценки не учитывались в расчетах показателя пригодности.

Система оценки и валидация результатов. Эксперты оценивали каждое решение по трем ключевым критериям.

1. Корректность результата (С) – правильность итогового ответа.
2. Обоснованность метода (М) – логическая и методическая корректность хода решения.
3. Структурированность и читаемость (S) – полнота, ясность представления.

Каждый критерий оценивался по трехбалльной шкале (1–3 балла), что обеспечивало умеренно дискретную, но достаточную для анализа дифференциацию качества решений.

Экспертные оценки были агрегированы в матрицу данных, на основании которой для каждой модели вычислялись средние баллы по критериям и итоговый интегральный показатель в соответствии с математической моделью.

Результаты опроса

По столбцам расположены оценки для нейросетей в порядке: ChatGPT, DeepSeek, Gemini, YandexGPT, GigaChat.

Задача 1						
Респондент 1	Полнота решения (C)	2	2	3	2	2
	Структурированность и читаемость (S)	2	3	3	3	2
	Обоснованность методов (M)	1	2	3	2	2
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	2	1	3	2	1
	Структурированность и читаемость (S)	3	2	3	2	2
	Обоснованность методов (M)	3	1	3	3	3
Респондент 4	Полнота решения (C)	2	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	2
	Обоснованность методов (M)	3	3	3	3	3
	Правильность решения	1	1	1	1	1

Задача 2						
Респондент 1	Полнота решения (C)	2	3	2	2	2
	Структурированность и читаемость (S)	2	3	2	2	3
	Обоснованность методов (M)	1	2	2	1	2
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	3	3	3	2
Респондент 3	Полнота решения (C)	2	2	3	2	3
	Структурированность и читаемость (S)	3	3	1	2	2
	Обоснованность методов (M)	3	3	3	3	1
Респондент 4	Полнота решения (C)	2	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	1
	Правильность решения	1	1	1	1	0

Задача 3						
Респондент 1	Полнота решения (C)	2	2	2	3	2
	Структурированность и читаемость (S)	2	3	3	3	3
	Обоснованность методов (M)	1	2	3	2	2
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	1	2	3	2	2
	Структурированность и читаемость (S)	2	2	1	2	2

	Обоснованность ме- тодов (M)	3	3	3	3	3
Респондент 4	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность ме- тодов (M)	3	3	3	3	3
	Правильность решения	1	1	1	1	1

Задача 4

Респондент 1	Полнота решения (C)	2	3	3	1	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность ме- тодов (M)	2	2	3	2	2
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность ме- тодов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	1	3	3	2	2
	Структурированность и читаемость (S)	3	1	1	2	2
	Обоснованность ме- тодов (M)	2	3	3	2	2
Респондент 4	Полнота решения (C)	3	3	3	2	2
	Структурированность и читаемость (S)	2	3	3	2	3
	Обоснованность ме- тодов (M)	3	3	3	3	3
	Правильность решения	1	1	1	1	0

Задача 5

Респондент 1	Полнота решения (C)	2	2	3	2	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность ме- тодов (M)	2	2	3	1	2

Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	2	3	3	2	1
	Структурированность и читаемость (S)	3	3	2	2	1
	Обоснованность методов (M)	2	2	2	2	1
Респондент 4	Полнота решения (C)	3	3	3	3	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	2	3	3
	Правильность решения	1	1	1	1	0

Задача 6

Респондент 1	Полнота решения (C)	2	2	3	2	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	2	3	2	3
Респондент 2	Полнота решения (C)	2	3	3	3	3
	Структурированность и читаемость (S)	2	3	3	3	3
	Обоснованность методов (M)	2	3	3	3	3
Респондент 3	Полнота решения (C)	1	2	3	2	2
	Структурированность и читаемость (S)	3	3	2	3	3
	Обоснованность методов (M)	1	3	3	3	3
Респондент 4	Полнота решения (C)	2	2	3	2	3
	Структурированность и читаемость (S)	1	3	3	3	3

	Обоснованность методов (M)	2	3	3	2	2
	Правильность решения	1	1	1	0	1

Задача 7						
Респондент 1	Полнота решения (C)	2	2	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	2	3	2	3
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	3	3	2	3	3
	Структурированность и читаемость (S)	2	3	1	3	2
	Обоснованность методов (M)	3	3	1	3	3
Респондент 4	Полнота решения (C)	2	2	3	3	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	1	3	3	2	2
	Правильность решения	1	1	1	1	1

Задача 8						
Респондент 1	Полнота решения (C)	3	3	3	2	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	3	3	2	2
Респондент 2	Полнота решения (C)	3			3	3
	Структурированность и читаемость (S)	3			3	3
	Обоснованность методов (M)	3			3	3

Респондент 3	Полнота решения (C)	2	3	3	2	2
	Структурированность и читаемость (S)	3	1	1	2	3
	Обоснованность методов (M)	2	2	2	2	2
Респондент 4	Полнота решения (C)	3			3	3
	Структурированность и читаемость (S)	3			3	3
	Обоснованность методов (M)	3			2	3
	Правильность решения	0	0	0	0	0

Задача 9

Респондент 1	Полнота решения (C)	2	3	3	2	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	3	3	2	2
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	2	3	1	2	2
	Структурированность и читаемость (S)	3	3	2	2	2
	Обоснованность методов (M)	3	3	1	3	3
Респондент 4	Полнота решения (C)	2	3	3	2	3
	Структурированность и читаемость (S)	3	3	2	3	3
	Обоснованность методов (M)	2	3	2	2	3
	Правильность решения	1	1	1	1	1

Задача 10

Респондент 1	Полнота решения (C)	2	2	3	3	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	2	2	3	2
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	2	1	2	3	2
	Структурированность и читаемость (S)	3	2	2	3	2
	Обоснованность методов (M)	1	3	1	3	3
Респондент 4	Полнота решения (C)	3	3	2	2	2
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	2	2	3	2
	Правильность решения	1	1	1	1	1

Задача 11

Респондент 1	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	2	2	2	3	3
Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	3	3	3	1	2
	Структурированность и читаемость (S)	3	2	2	2	3

	Обоснованность ме- тодов (M)	3	3	3	1	3
Респондент 4	Полнота решения (C)	3	3	3	2	3
	Структурированность и читаемость (S)	3	2	3	3	3
	Обоснованность ме- тодов (M)	3	2	3	2	3
	Правильность решения	1	1	1	1	1

Задача 12

Респондент 1	Полнота решения (C)	2	3	3	2	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность ме- тодов (M)	2	3	3	2	3
Респондент 2	Полнота решения (C)	3	3	3	3	
	Структурированность и читаемость (S)	3	3	3	3	
	Обоснованность ме- тодов (M)	3	3	3	3	
Респондент 3	Полнота решения (C)	2	3	3	1	1
	Структурированность и читаемость (S)	3	2	2	3	1
	Обоснованность ме- тодов (M)	3	3	3	1	1
Респондент 4	Полнота решения (C)	3	3	3	2	3
	Структурированность и читаемость (S)	2	3	3	3	3
	Обоснованность ме- тодов (M)	3	3	3	2	2
	Правильность решения	1	1	1	0	0

Задача 13

Респондент 1	Полнота решения (C)	3	2	3	2	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность ме- тодов (M)	2	2	3	2	3

Респондент 2	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	3	3	3	3
	Обоснованность методов (M)	3	3	3	3	3
Респондент 3	Полнота решения (C)	2	2	3	2	2
	Структурированность и читаемость (S)	2	3	2	3	2
	Обоснованность методов (M)	3	3	3	3	3
Респондент 4	Полнота решения (C)	3	3	3	3	3
	Структурированность и читаемость (S)	3	2	3	3	3
	Обоснованность методов (M)	3	3	3	2	3
	Правильность решения	1	1	1	1	1

Аналитическая обработка результатов опроса. В результате обработки данных были получены три ключевых артефакта.

1. Критериальная оценка качества математических решений нейросетями (сравнение по параметрам C, S и M).
2. Сравнение среднего итогового балла международных и отечественных моделей.
3. Тепловая карта качества решений по тематическим блокам математического анализа (выявление сильных и слабых сторон каждой нейросети).

Анализ результатов

В ходе исследования были получены результаты, позволяющие оценить пригодность нейросетей для решения задач математического анализа, ответы нейросетей по задачам из разных тем и сравнить заграничные и отечественные нейросети в рамках проводимого исследования.

По итогам исследования был составлен ранжированный список нейросетей по убыванию значения показателя пригодности Р (табл. 3):

- 1) Gemini;
- 2) DeepSeek;
- 3) ChatGPT;
- 4) YandexGPT;
- 5) GigaChat.

Табл. 3. Оценка качества решений нейросетей по критериям

Критерий	Нейросеть				
	ChatGPT	DeepSeek	Gemini	YandexGPT	GigaChat
Полнота решения (C)	51,11	56,44	60,89	44,44	36,00
Структурированность и читаемость (S)	59,11	60,00	56,44	49,33	39,11
Обоснованность методов (M)	52,00	57,33	58,22	45,33	39,56
Итог	162,22	173,78	175,56	139,11	114,67

Для оценки эффективности различных нейросетевых моделей при решении задач математического анализа был проведен сравнительный анализ пяти систем: ChatGPT, DeepSeek, Gemini, YandexGPT и GigaChat. Оценка осуществлялась по трем критериям: полнота решения (C), структурированность и читаемость (S) и обоснованность методов (M). Итоговый показатель пригодности определялся как сумма взвешенных значений по каждому из критериев. Полученные данные представлены в табл. 3 и визуализированы на рис. 1.

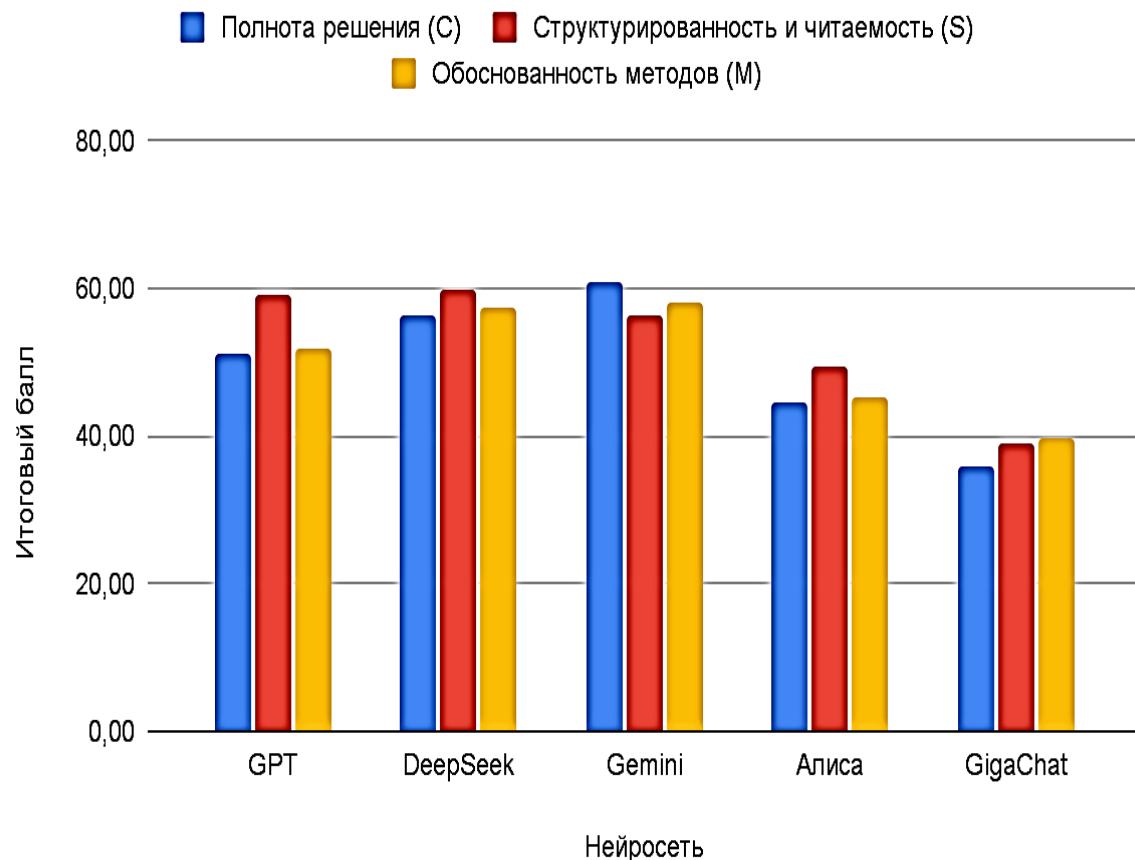


Рис. 1. Критериальная оценка качества математических решений нейросетями

Согласно представленным результатам, наивысшее значение суммарного показателя наблюдается у нейросети Gemini (175,56), что указывает на ее лидерство в исследуемой группе. Несколько ниже результат у DeepSeek (173,78), который практически не уступает лидеру. Третье место занимает ChatGPT (162,22), демонстрируя устойчивые показатели по всем критериям. Существенно меньшие значения отмечены у моделей YandexGPT (139,11) и GigaChat (114,67).

Рассмотрение распределения значений по отдельным критериям позволяет выявить специфику сильных сторон каждой модели.

По критерию «Полнота решения (C)» наилучшие результаты продемонстрировала нейросеть Gemini (60,89), что отражает ее способность полно и всесторонне раскрывать содержание задач. Далее следуют DeepSeek (56,44) и ChatGPT (51,11). После них идут отечественные разработки: YandexGPT (44,44) и GigaChat (36,00).

По показателю «Структурированность и читаемость (S)» лидером является DeepSeek (60,00), что свидетельствует о высокой степени логической упорядоченности и понятности решений. Ненамного отстают ChatGPT (59,11) и Gemini (56,44). После них идут отечественные разработки: YandexGPT (49,33) и GigaChat (39,11).

Критерий «Обоснованность методов (M)» вновь возглавляет Gemini (58,22), подтверждая ее способность к корректному выбору и применению математических подходов. Близкие результаты показывает DeepSeek (57,33), в то время как ChatGPT (52,00) занимает промежуточную позицию. После них идут отечественные разработки: YandexGPT (45,33) и GigaChat (39,56).

Визуальный анализ данных, представленных на рис. 1, позволяет отметить, что для моделей Gemini и DeepSeek характерна высокая сбалансированность по всем трем критериям – значения показателей расположены на примерно одинаковом уровне. В то же время у ChatGPT наблюдается умеренный перекос в сторону критерия структурированности, что может свидетельствовать о приоритете логичности изложения над глубиной анализа.

Модели YandexGPT и GigaChat демонстрируют сравнительно низкие и близкие по величине показатели по всем критериям, что указывает на ограниченную способность к выполнению аналитических задач математики.

Таким образом, как по критериям полноты, обоснованности и структурированности решений, так и по тематическому охвату, наилучшими моделями для решения задач математического анализа являются Gemini и DeepSeek, обладающие сбалансированными и устойчивыми характеристиками качества.

Представленная в табл. 4 информация отражает количественную оценку производительности пяти нейросетевых моделей при решении задач по 13 разделам математического анализа. Оценка, варьирующаяся от 0,00 (полное отсут-

ствие или неверное решение) до 3,00 (максимально корректный результат), выявила существенную неоднородность в их способностях.

Табл. 4. Термовая карта качества решений по тематическим блокам математического анализа

Задача	Нейросеть				
	ChatGPT	DeepSeek	Gemini	YandexGPT	GigaChat
Функции	2,44	2,36	3,00	2,64	2,42
Предел. Непрерывность	2,36	2,81	2,64	2,50	0,00
Производная и дифференциал. Дифференциальное исчисление	2,39	2,64	2,78	2,75	2,64
Исследование функций и их графиков	2,44	2,81	2,89	2,28	0,00
Определенный интеграл	2,61	2,72	2,78	2,47	0,00
Неопределенный интеграл. Интегральное исчисление	1,86	2,58	2,94	0,00	2,69
Способы вычисления определенных интегралов. Несобственные интегралы	2,47	2,69	2,61	2,83	2,75
Применение интеграла	0,00	0,00	0,00	0,00	0,00
Ряды	2,50	3,00	2,42	2,44	2,64
Функции нескольких переменных. Дифференциальное исчисление	2,53	2,44	2,39	2,89	2,44
Применения дифференциального исчисления функций нескольких переменных	2,92	2,72	2,86	2,36	2,89
Многомерные интегралы и кратное интегрирование	2,64	2,94	2,94	0,00	0,00
Криволинейные интегралы и интегралы по поверхности	2,75	2,64	2,94	2,61	2,83

Модель Gemini демонстрирует наиболее высокие и стабильные результаты, достигая максимального балла (3,00) в разделе «Функции» и показывая стабильно высокие оценки (2,78–2,94) в большинстве других разделов, включая «Неопределенный интеграл» и «Многомерные интегралы», что позиционирует ее как наиболее универсальный инструмент. Модель DeepSeek также показывает высокую компетентность, получив максимальный балл (3,00) за «Ряды» и высокие оценки (2,81) за «Предел. Непрерывность» и «Исследование функций».

GigaChat и YandexGPT демонстрируют крайне поляризованные результаты. Они достигают высоких баллов в отдельных сложных темах (напри-

мер, 2,89 у GigaChat в «Применениях дифференциального исчисления функций нескольких переменных» и 2,89 у YandexGPT в «Функциях нескольких переменных»), однако обе модели показывают нулевую производительность в нескольких других разделах.

ChatGPT занимает промежуточное положение, показывая наиболее усредненную производительность с оценками преимущественно в среднем диапазоне (2,36–2,64), за исключением заметно низкого балла (1,86) по теме «Неопределенный интеграл».

Выделяются общие закономерности:

- 1) Gemini отличается наиболее высокой универсальностью и стабильностью результатов, превосходя другие модели по большинству тем;
- 2) DeepSeek демонстрирует близкие показатели, особенно в областях, связанных с рядами и исследованием функций;
- 3) ChatGPT занимает промежуточное положение и показывает хорошие результаты в задачах, требующих применения производных и интегралов;
- 4) YandexGPT и GigaChat демонстрируют ограниченную применимость, показывая высокие результаты не по всем направлениям.

Анализ успеваемости по конкретным разделам выявляет системные проблемы. Наиболее показательным результатом является раздел «Применение интеграла», по которому все пять исследуемых моделей получили оценку 0,00. Стоит также отметить, что дополнительная задача из другой области математики – стереометрии, была решена корректно только двумя нейросетями – ChatGPT и YandexGPT. Это указывает на фундаментальный пробел в способности систем к концептуальному применению математических инструментов. Другие разделы, такие как «Многомерные интегралы» и «Неопределенный интеграл», также выявили критические пробелы у отдельных моделей (GigaChat, YandexGPT), в то время как наиболее успешно всеми моделями были решены задачи из разделов «Производная и дифференциал». В целом, проведенный анализ свидетельствует, что, несмотря на впечатляющие успехи в отдельных областях, достижение последовательной и надежной производительности в высшей математике остается для современных нейросетевых моделей открытой задачей.

Одной из задач исследования являлось сопоставление эффективности международных (ChatGPT, DeepSeek и Gemini) и отечественных (YandexGPT и GigaChat) LLM при решении задач математического анализа.

Сравнительный анализ показал существенное различие между двумя группами. Средний итоговый показатель для международных моделей составил 170,52, тогда как для отечественных – 126,89 (рис. 2). Разница в 43,63 балла отражает выраженное преимущество международных систем по совокупности оценочных критериев.

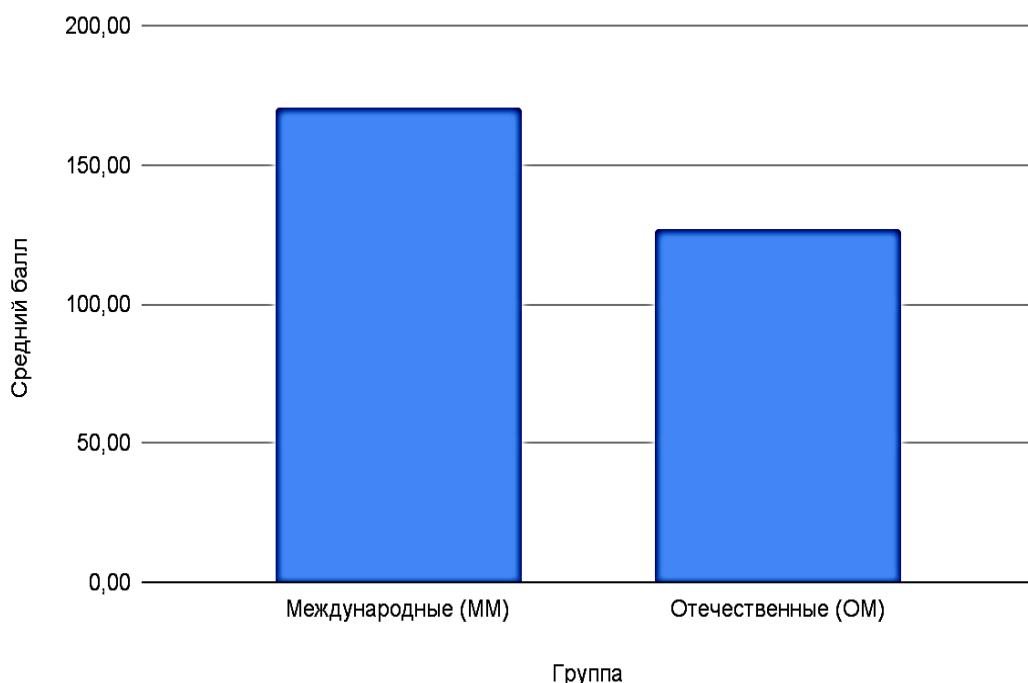


Рис. 2. Сравнение среднего итогового балла международных и отечественных моделей

Анализ указывает на то, что международные модели демонстрируют более высокий уровень точности, полноты и структурной согласованности при решении задач математического анализа. Отечественные системы характеризуются большей разрозненностью показателей и менее стабильным качеством решений при переходе между различными тематическими разделами.

Заключение

На основе представленного исследования, посвященного оценке и сравнительному анализу эффективности современных больших языковых моделей (LLM) в решении задач, можно сформулировать следующие общие выводы.

Актуальность работы подтверждается стремительным развитием технологий искусственного интеллекта и их активным внедрением в образовательную сферу. Однако, несмотря на потенциал ИИ в качестве ассистента для студентов, его надежность в решении задач, требующих строгой математической логики, оставалась предметом дискуссий.

Целью работы являлся сравнительный анализ пяти ведущих LLM (ChatGPT, Gemini, DeepSeek, YandexGPT и GigaChat) для оценки их потенциала в образовательной практике и выявления различий между отечественными и международными разработками.

Для достижения цели была разработана комплексная методология, включающая формирование репрезентативного набора из 13 задач по различным разделам математического анализа и создание математической модели оценки,

основанной на трех критериях: полнота решения (C), структурированность и читаемость (S) и обоснованность методов (M).

Ключевые результаты анализа.

1. *Ранжирование моделей.* По итогам тестирования и экспертной оценки был составлен ранжированный список нейросетей по убыванию показателя пригодности. Лидирующие позиции заняли международные модели: Gemini (175,56 балла) и DeepSeek (173,78 балла), продемонстрировавшие наиболее высокие и сбалансированные результаты. Модель ChatGPT (162,22) заняла третье место. Отечественные модели YandexGPT (139,11) и GigaChat (114,67) показали более низкие совокупные результаты.

2. *Сравнение отечественных и международных моделей.* Исследование выявило существенный разрыв между двумя группами. Средний итоговый показатель для международных моделей составил 170,52, в то время как для отечественных – 126,89. Международные системы продемонстрировали более высокий уровень точности, полноты и структурной согласованности решений. Отечественные модели характеризовались большей нестабильностью и поляризованными результатами, показывая высокую компетентность в одних разделах и нулевую в других.

3. *Выявление фундаментальных пробелов.* Наиболее важным результатом стало обнаружение системных недостатков, присущих *всем* исследуемым моделям. В разделе «Применение интеграла» все пять нейросетей получили оценку 0,00. Это указывает на фундаментальный пробел в способности ИИ к концептуальному применению математических инструментов, а не только к выполнению алгоритмических расчетов.

Проведенный анализ показал, что современные большие языковые модели, в особенности Gemini и DeepSeek, демонстрируют значительный потенциал для использования в образовательном процессе по математическому анализу.

Тем не менее их текущая применимость ограничена. Выявленные фундаментальные пробелы, особенно в задачах на концептуальное применение знаний, свидетельствуют о том, что достижение последовательной и надежной производительности, сопоставимой с человеческим уровнем математического мышления, остается открытой задачей для современных LLM.

Таким образом, на данном этапе развития ИИ может рассматриваться исключительно как вспомогательный инструмент, требующий дополнительной верификации.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Аудитория «Алисы» уже более 43 млн человек // IT Speaker : [сайт]. – 2025. – URL: <https://itspeaker.ru/news/auditoriya-alisy-uzhe-bolee-43-mln-chelovek/> (дата обращения: 25.11.2025).

2. **Берман, Г. Н.** Сборник задач по курсу математического анализа : учеб. пособие / Г. Н. Берман. – 22-е изд. – М. : Наука, 1985. – 448 с.

3. Дайджест EduTech. Как изменится EdTech в России и мире. Искусственный интеллект, виртуальная реальность и видеообучение // СберУниверситет: [сайт]. – М. : СберУниверситет, 2023. – Вып. 24. – 46 с. – URL: <https://sberuiversity.ru> (дата обращения: 25.11.2025).

4. Топ-менеджер Сбера назвал ежемесячное количество пользователей GigaChat // Lenta.ru : [сайт]. – 2025. – 20 нояб. – URL: <https://lenta.ru/news/2025/11/20/top-menedzher-sbera-nazval-ezhemesyachnoe-kolichestvo-polzovateley-gigachat/> (дата обращения: 25.11.2025).

5. AI in Edtech: Top Use Cases You Need To Know // SmartDev : [website]. – URL: <https://smartdev.com/ai-use-cases-in-edtech/> (date of access: 06.10.2025).

6. How artificial intelligence will impact K-12 teachers / J. Bryant, C. Heitz, S. Sanghvi, D. Wagle // McKinsey & Company. Public Sector Practice & Social Sector Practice. – 2020. – P. 7.

7. ChatGPT Statistics: User, Growth, Revenue & Facts // NerdyNav : [website]. – URL: <https://nerdynav.com/chatgpt-statistics/> (date of access: 25.11.2025).

8. DeepSeek Stats: User Count & Growth Stats // Backlinko : [website]. – URL: <https://backlinko.com/deepseek-stats> (date of access: 25.11.2025).

9. Gemini Stats Overview: User Count & Trends // Thunderbit : [website]. – URL: <https://thunderbit.com/blog/gemini-stats-overview> (date of access: 25.11.2025).

10. LLM statistics 2025: Comprehensive insights into market trends and integration // Hostinger : [website]. – URL: <https://www.hostinger.com/tutorials/llm-statistics#:~:text=The%20global%20LLM%20market%20was,33.7%25%20over%20the%20forecast%20period> (date of access: 25.11.2025).