

ModelComparator · Финальная сводка

Проект: Демо проект
Сценарий: Научный сценарий
Описание сценария: Расскажи о квантовой теории фотоэффекта
Моделей: 3 · Критериев: 10 · Оценок: 30

Краткий вывод

Итоговый отчёт
Проект: Демо проект
Сценарий: Научный сценарий

Рекомендуемая модель: DeepSeek
Интегральная оценка (K): 0.750
Интерпретация: Хорошая модель, подходит для большинства задач.

Почему именно эта модель:

- Ключевые критерии вклада: Точность ответа, Логичность и структура, Глубина и полнота ответа
- Отрыв от второго места: 0.075
- Выбор выглядит стабильным.
- Ближайший конкурент: Алиса AI
- Наибольшее преимущество по критериям: Логичность и структура, Обработка сложных запросов, Гибкость в интерпретации
- Где лидер уступает: Точность ответа, Глубина и полнота ответа, Контекстная согласованность
- Сильные стороны лидера: Точность ответа (оценка 4, вклад 1.20), Логичность и структура (оценка 4, вклад 0.80), Глубина и полнота ответа (оценка 4, вклад 0.80)
- Слабые стороны лидера: Адекватность формата (оценка 3, вклад 0.15), Чувствительность к контексту (оценка 3, вклад 0.15), Креативность (оценка 3, вклад 0.15)

Условия, при которых выбор может измениться:

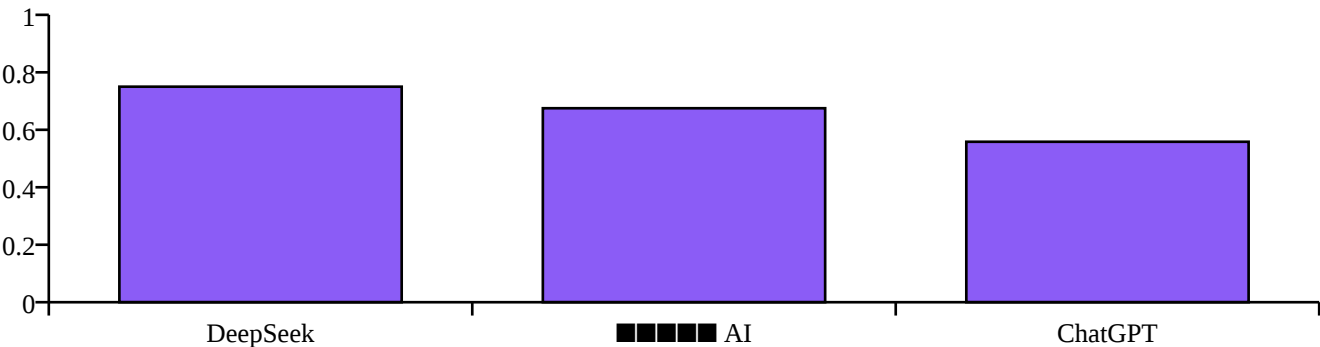
- Если повысить веса критериев, где сильнее Алиса AI (Точность ответа, Глубина и полнота ответа, Контекстная согласованность), лидер может измениться.

Примечание: расчёт основан на включённых критериях и текущих весах.

Рейтинг моделей

Место	Модель	K	Взвешенная сумма	Пропуски
1	DeepSeek	0.750	4.500	0
2	Алиса AI	0.675	4.050	0
3	ChatGPT	0.558	3.350	0

График K по моделям



Критерии и веса

Группа	Критерий	Вес	Описание
Точность	Точность ответа	0.300	Насколько точно и уместно модель отвечает на запрос.
Точность	Логичность и структура	0.200	Логичность, структурированность и последовательность ответа.
Точность	Глубина и полнота ответа	0.200	Насколько полно и глубоко раскрыта тема запроса.
Точность	Гибкость в интерпретации	0.100	Работа с неоднозначными или неточно сформулированными запросами.
Точность	Адекватность формата	0.050	Соответствие формата ответа задаче (таблица, код, шаги и т.д.).
Контекст	Чувствительность к контексту	0.050	Учет предыдущих сообщений и контекста диалога.
Контекст	Креативность	0.050	Способность предложить нестандартные решения или идеи.
Устойчивость	Обработка сложных запросов	0.100	Работа с многоэтапными и сложными задачами.
Точность	Понимание технической специфики	0.100	Корректная работа со специализированными запросами.
Контекст	Контекстная согласованность	0.050	Сохранение связности и согласованности контекста.

Метрики LLM

Модель	Accuracy	Precision	Recall	F1	BLEU	ROUGE-L	BERTScore	PPL
ChatGPT	0.540	0.633	0.400	0.490	0.526	0.533	0.539	41.8
DeepSeek	0.700	0.700	0.700	0.700	0.670	0.685	0.696	29.0
Алиса AI	0.580	0.633	0.500	0.559	0.562	0.571	0.578	38.6

Матрица оценок

Критерий	ChatGPT	DeepSeek	Алиса AI
Точность ответа	3	4	5
Логичность и структура	1	4	2
Глубина и полнота ответа	5	4	5
Гибкость в интерпретации	2	2	1
Адекватность формата	5	3	3
Чувствительность к контексту	2	3	2
Креативность	1	3	2
Обработка сложных запросов	2	5	2
Понимание технической специфики	3	4	3
Контекстная согласованность	3	3	4

Сильные и слабые стороны

Модель	Сильные стороны	Слабые стороны	Пропуски
DeepSeek	Точность ответа (оценка 4, вклад 1.20), Логичность и структура (оценка 3, вклад 0.15), Глубина и полнота ответа (оценка 4, вклад 1.00), Гибкость в интерпретации (оценка 2, вклад 0.10), Адекватность формата (оценка 5, вклад 0.25), Чувствительность к контексту (оценка 2, вклад 0.10), Креативность (оценка 1, вклад 0.05), Обработка сложных запросов (оценка 2, вклад 0.20), Понимание технической специфики (оценка 3, вклад 0.30), Контекстная согласованность (оценка 3, вклад 0.15)	Недостаточная креативность (оценка 1), Низкая гибкость в интерпретации (оценка 2), Слабая контекстная согласованность (оценка 3)	Пропуск анализа сложных запросов, отсутствие поддержки мультимедиа
Алиса AI	Точность ответа (оценка 5, вклад 1.50), Глубина и полнота ответа (оценка 5, вклад 1.00), Адекватность формата (оценка 3, вклад 0.15), Чувствительность к контексту (оценка 2, вклад 0.10), Понимание технической специфики (оценка 3, вклад 0.30), Контекстная согласованность (оценка 4, вклад 0.20)	Низкая логичность и структура (оценка 1), Слабая гибкость в интерпретации (оценка 1), Низкая креативность (оценка 2), Низкая обработка сложных запросов (оценка 2)	Пропуск анализа сложных запросов, отсутствие поддержки мультимедиа
ChatGPT	Глубина и полнота ответа (оценка 5, вклад 1.00), Точность ответа (оценка 3, вклад 0.90), Логичность и структура (оценка 1, вклад 0.05), Гибкость в интерпретации (оценка 2, вклад 0.20), Адекватность формата (оценка 5, вклад 0.25), Чувствительность к контексту (оценка 2, вклад 0.10), Креативность (оценка 1, вклад 0.05), Обработка сложных запросов (оценка 2, вклад 0.20), Понимание технической специфики (оценка 3, вклад 0.30), Контекстная согласованность (оценка 3, вклад 0.15)	Недостаточная логичность и структура (оценка 1), Низкая гибкость в интерпретации (оценка 2), Слабая креативность (оценка 1), Низкая обработка сложных запросов (оценка 2)	Пропуск анализа сложных запросов, отсутствие поддержки мультимедиа