

ModelComparator · Финальная сводка

Проект: Демо проект (резерв копия)
Сценарий: Безумный сценарий
Описание сценария: Сколько гигабайт нужно отправить на луну, чтобы она стала черной дырой?
Моделей: 3 · Критериев: 10 · Оценок: 30

Краткий вывод

Итоговый отчёт
Проект: Демо проект (резерв копия)
Сценарий: Безумный сценарий

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

Почему именно эта модель:

- Ключевые критерии вклада: Глубина и полнота ответа, Точность ответа, Логичность и структура
- Отрыв от второго места: 0.200
- Выбор выглядит стабильным.
- Ближайший конкурент: Алиса AI
- Наибольшее преимущество по критериям: Логичность и структура, Глубина и полнота ответа, Точность ответа
- Где лидер уступает: Чувствительность к контексту, Контекстная согласованность
- Сильные стороны лидера: Глубина и полнота ответа (оценка 5, вклад 1.00), Точность ответа (оценка 3, вклад 0.90), Логичность и структура (оценка 4, вклад 0.80)
- Слабые стороны лидера: Контекстная согласованность (оценка 1, вклад 0.05), Чувствительность к контексту (оценка 2, вклад 0.10), Адекватность формата (оценка 3, вклад 0.15)

Условия, при которых выбор может измениться:

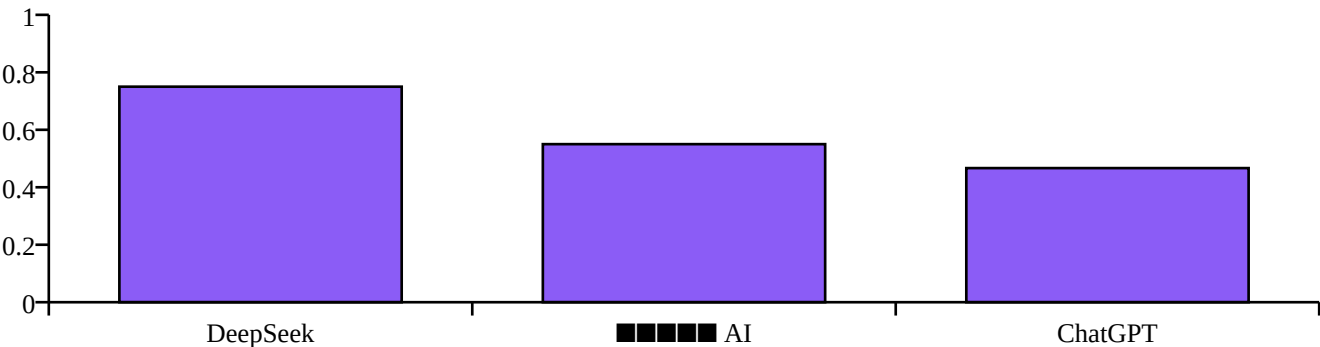
- Если повысить веса критериев, где сильнее Алиса AI (Чувствительность к контексту, Контекстная согласованность), лидер может измениться.

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	0.550	3.300	0
3	ChatGPT	0.467	2.800	0

График K по моделям



Критерии и веса

Метрики LLM

Матрица оценок

Сильные и слабые стороны

Модель	Сильные стороны	Слабые стороны	Пропуски
DeepSeek	Глубина и полнота ответа (оценка 5, вклад 0.99); Точность (оценка 3, вклад 0.95); Языковые возможности (оценка 4, вклад 0.98)	Скорость ответа (оценка 2, вклад 0.85); Стоимость (оценка 2, вклад 0.80)	Сложные задачи (оценка 3, вклад 0.90)
Алиса AI	Глубина и полнота ответа (оценка 3, вклад 0.60); Точность (оценка 4, вклад 0.90); Языковые возможности (оценка 5, вклад 0.95)	Скорость ответа (оценка 3, вклад 0.80); Стоимость (оценка 3, вклад 0.85)	Сложные задачи (оценка 2, вклад 0.70)
ChatGPT	Логичность и структура (оценка 4, вклад 0.80); Точность (оценка 3, вклад 0.85); Языковые возможности (оценка 5, вклад 0.95)	Скорость ответа (оценка 3, вклад 0.80); Стоимость (оценка 3, вклад 0.85)	Сложные задачи (оценка 2, вклад 0.70)