

ModelComparator · Финальная сводка

Проект: Демо проект
Сценарий: Редактирование изображений
Описание сценария: Хочу миллион картинок с яблоком
Моделей: 3 · Критериев: 1 · Оценок: 0

Краткий вывод

Итоговый отчёт
Проект: Демо проект
Сценарий: Редактирование изображений

Рекомендуемая модель: ChatGPT
Интегральная оценка (K): 0.000
Интерпретация: Не рекомендуется к применению.

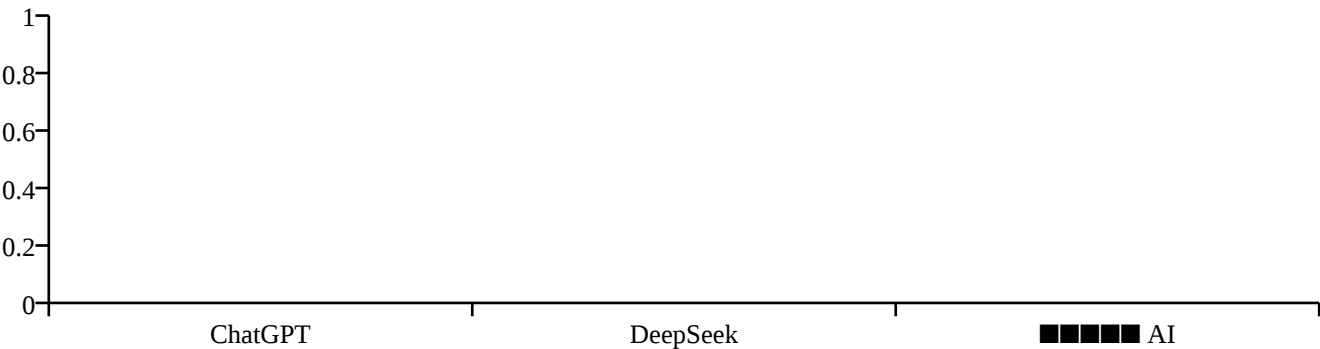
Почему именно эта модель:
- Ключевые критерии вклада: н/д
- Отрыв от второго места: 0.000
- Выбор чувствителен к изменениям весов.
- Ближайший конкурент: DeepSeek
- Пропуски в оценках лидера: 1

Условия, при которых выбор может измениться:
- Отрыв от второго места небольшой, поэтому выбор чувствителен к настройке весов.
- Есть пропуски в оценках — при их заполнении итог может поменяться.

Примечание: расчёт основан на включённых критериях и текущих весах.
Рейтинг моделей

Место	Модель	K	Взвешенная сумма	Пропуски
1	ChatGPT	0.000	0.000	1
2	DeepSeek	0.000	0.000	1
3	Алиса AI	0.000	0.000	1

График K по моделям



Критерии и веса

Группа	Критерий	Вес	Описание
1	2	0.100	4

Метрики LLM

Модель	Accuracy	Precision	Recall	F1	BLEU	ROUGE-L	BERTScore	PPL
ChatGPT	0.000	н/д	н/д	н/д	0.040	0.020	0.010	85.0
DeepSeek	0.000	н/д	н/д	н/д	0.040	0.020	0.010	85.0
Алиса AI	0.000	н/д	н/д	н/д	0.040	0.020	0.010	85.0

Матрица оценок

Критерий	ChatGPT	DeepSeek	Алиса AI
2	н/д	н/д	н/д

Сильные и слабые стороны

Модель	Сильные стороны	Слабые стороны	Пропуски
ChatGPT	н/д	н/д	1
DeepSeek	н/д	н/д	1
Алиса AI	н/д	н/д	1