

Comparaçāo de Modelos

GPT-4 e DeepSeek

O que as IAs têm a dizer?

Como foi feito:

Usando os critérios de avaliação descritos no HuggingFace, escrevemos um prompt para gerar uma tabela comparando o GPT-4 ao DeepSeek V3.

O prompt foi aplicado ao DeepSeek, ChatGPT e Gemini e cada um apresentou respostas diferentes.

A partir disso, pedimos às IAs para criarem uma pequena aplicação em Python para apresentar estas tabelas graficamente em uma página HTML.

Resultado:



Comparação de LLMs: ChatGPT, DeepSeek e Gemini

Análise baseada em diferentes fontes de dados e benchmarks

Tabela DeepSeek

Tabela ChatGPT

Tabela Gemini

Visualização Lado a Lado

Tabela DeepSeek

Critério de Avaliação	ChatGPT (GPT-4)	DeepSeek
Média Ponderada da Pontuação	85%	78%
Instruction Following Evaluation (IFEval)	85%	80%

Tabela ChatGPT

Critério de Avaliação	ChatGPT	DeepSeek
Média Ponderada da Pontuação	85%	80%
Instruction Following Evaluation (IFEval)	90%	80%

Tabela Gemini

Critério	ChatGPT	DeepSeek
Média ponderada da pontuação dos outros critérios	90%	80%
Instruction Following Evaluation (IFEval)	90%	85%