

# Análise de Dashboard: LLM-Stats

## 1. Introdução e Escolha do Caso

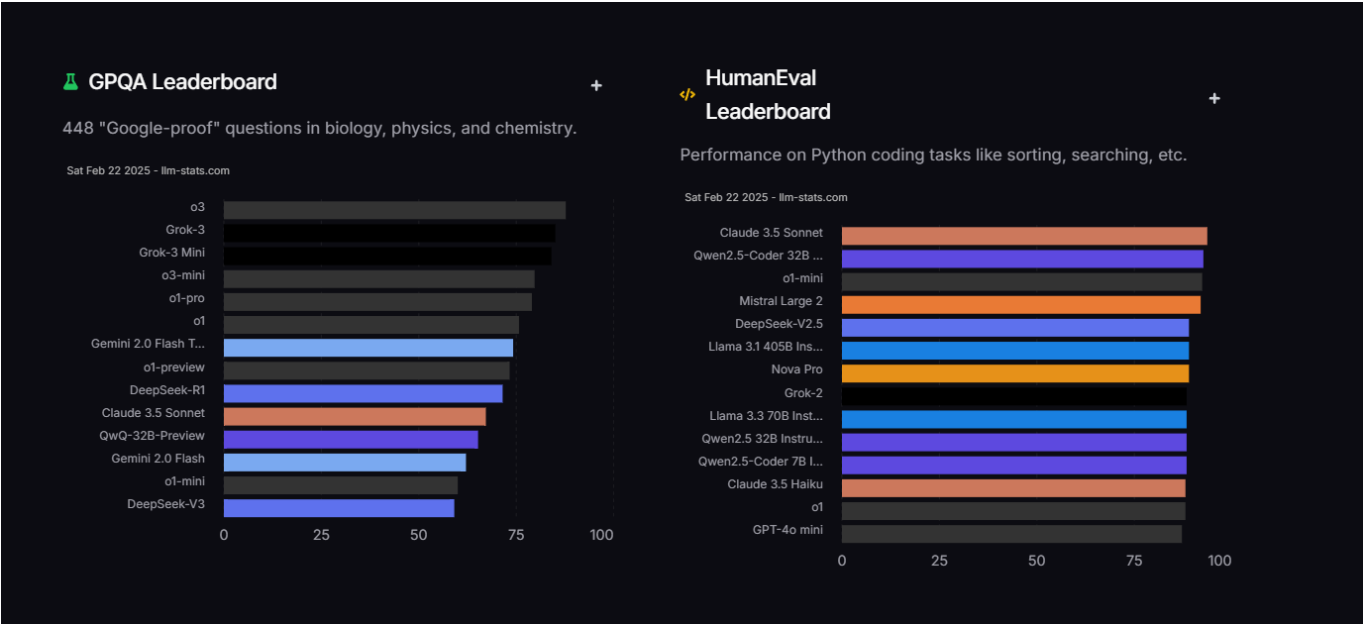
O avanço dos modelos de inteligência artificial tornou essencial a comparação entre diferentes Large Language Models (LLMs). O site LLM-Stats apresenta uma solução inovadora ao consolidar benchmarks de IA em uma única dashboard. Esta análise explora como esse dashboard organiza e apresenta os dados e como ele se compara a outras abordagens, como a da xAI.

## 2. Problema e Contexto

A crescente diversidade de modelos de IA criou um desafio para pesquisadores e empresas: comparar diferentes modelos de forma justa. Benchmarks padronizados são essenciais para avaliar o desempenho real dos LLMs. Dashboards que apresentam dados de forma transparente e intuitiva são ferramentas fundamentais para facilitar essa análise.

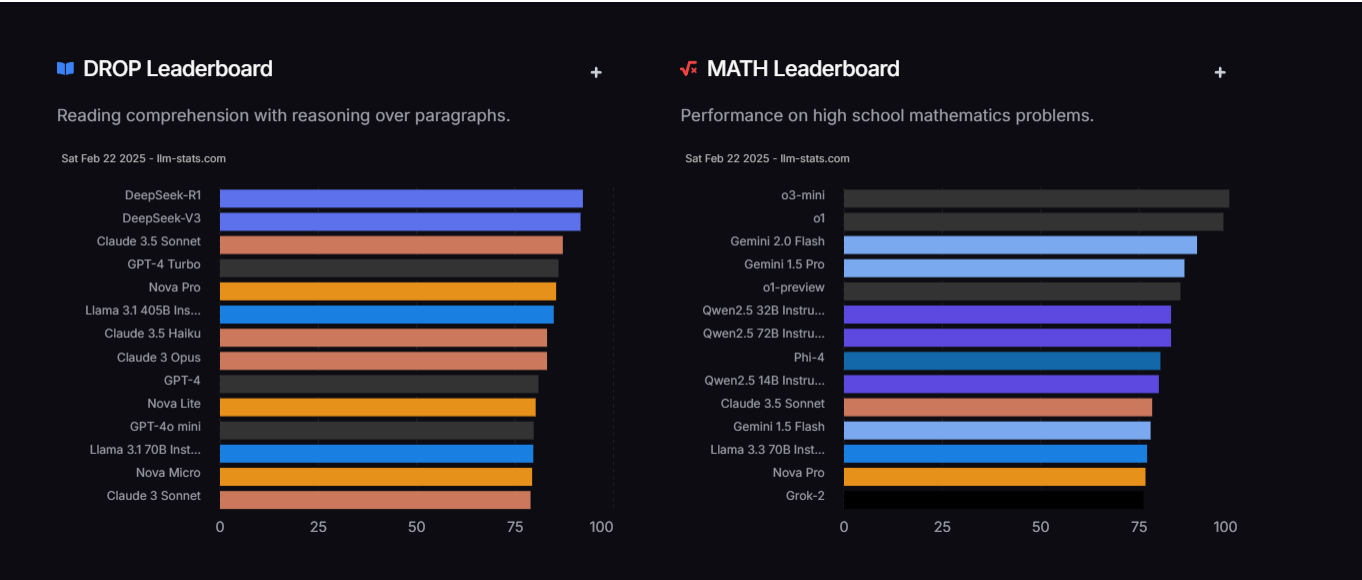
## 3. Descrição da Solução: LLM-Stats

O LLM-Stats se destaca por sua abordagem transparente e detalhada. Ele apresenta rankings de modelos com métricas diversas, utilizando gráficos absolutos que evitam manipulação de dados.



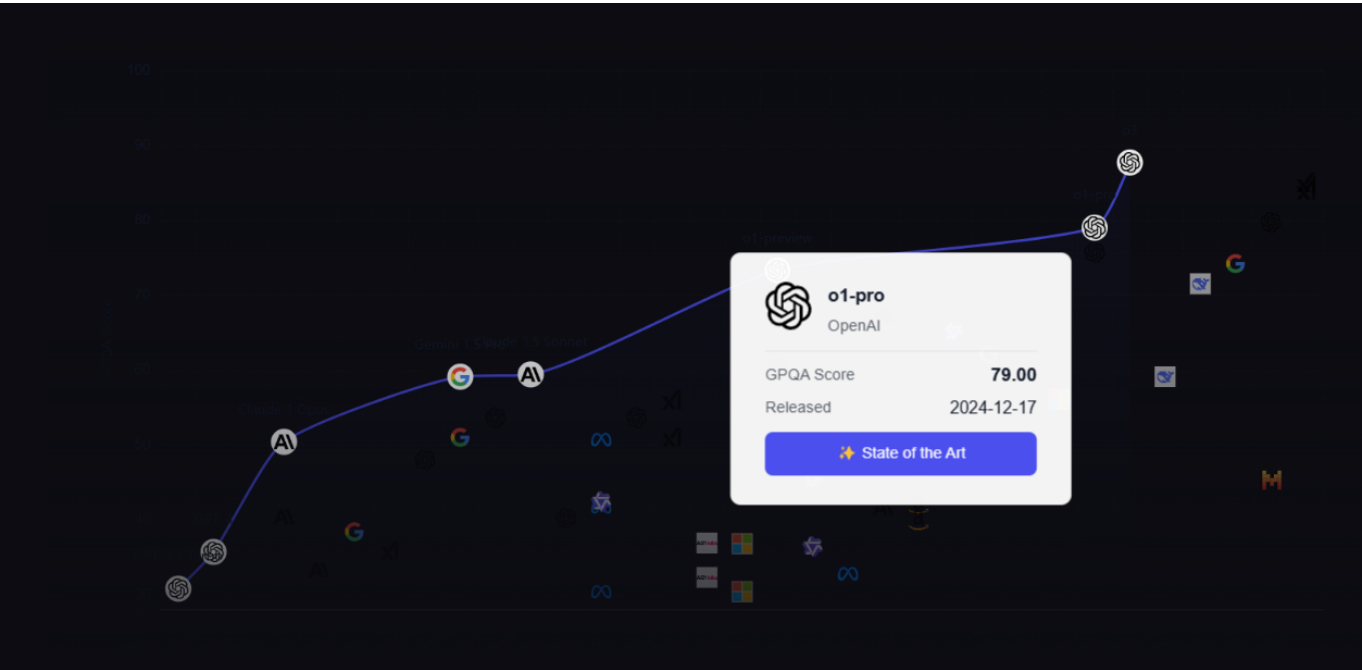
### GPQA Leaderboard

O gráfico GPQA Leaderboard apresenta a performance dos modelos em questões científicas complexas. A escolha de um gráfico de barras horizontais facilita a comparação direta entre os modelos. A escala absoluta utilizada mantém a transparência dos dados, evitando distorções.



MATH Leaderboard

Este gráfico avalia a habilidade dos LLMs em resolver problemas matemáticos. A escolha de barras coloridas para cada modelo auxilia na distinção rápida das performances. A estrutura bem definida permite uma interpretação rápida dos resultados.



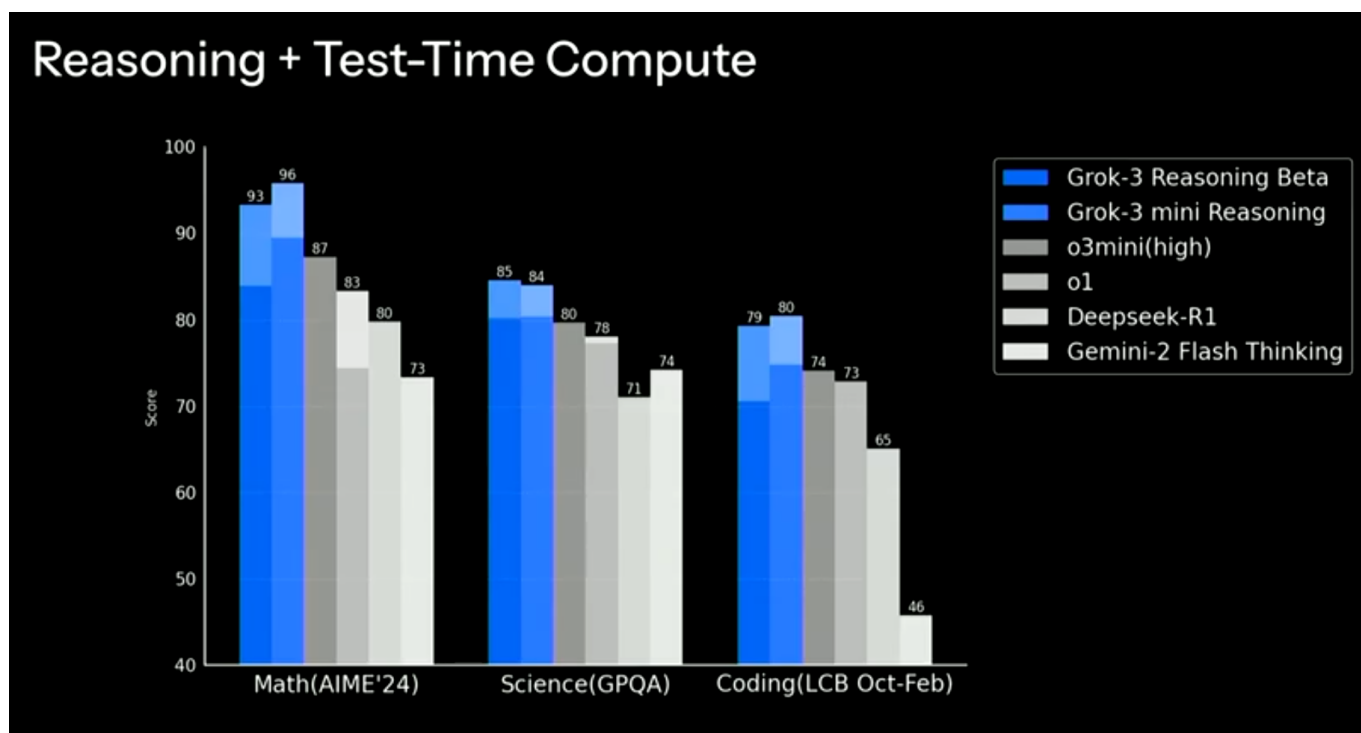
Evolução Temporal dos Modelos

O gráfico de linha ilustra a evolução da performance dos modelos ao longo do tempo. Ele permite uma visão clara do progresso e das tendências do setor, sendo um diferencial do LLM-Stats.

4. Comparação com xAI (Grok-3)

Diferente do LLM-Stats, a xAI utilizou um gráfico recortado para enfatizar a superioridade do Grok-3. Ao omitir a parte inferior do gráfico, a empresa gerou um efeito visual exagerado,

distorcendo a percepção dos dados.



## 5. Aplicação de Boas Práticas de Dashboards

De acordo com o artigo da Aela.io sobre design de dashboards, alguns princípios fundamentais são:

- **\*\*Uso de escalas absolutas\*\***: O LLM-Stats segue essa prática, garantindo que as comparações entre modelos sejam justas.
- **\*\*Hierarquia visual clara\*\***: Os rankings são organizados de forma intuitiva, priorizando os modelos mais relevantes.
- **\*\*Uso eficaz das cores\*\***: Cada modelo possui uma cor distinta, facilitando a identificação rápida.

Já no caso da xAI, a decisão de cortar o gráfico prejudica a transparência dos dados, indo contra a recomendação de clareza e integridade na visualização.

## 6. Conclusão

O LLM-Stats se destaca como um dashboard confiável e transparente para avaliação de LLMs. Ao seguir as melhores práticas descritas na Aela.io, como escalas absolutas, hierarquia visual clara e cores bem definidas, ele oferece uma experiência eficiente e útil para a comunidade de IA. Por outro lado, abordagens como a da xAI mostram como pequenas manipulações gráficas podem impactar a percepção dos dados, reforçando a importância da transparência na visualização de

informações.

## **Referências**

- <https://llm-stats.com/>
- <https://www.aela.io/pt-br/blog/conteudos/dashboard-como-criar-o-design-de-dados-e-informacoes>
- Análise comparativa de dashboards e visualização de dados em inteligência artificial.