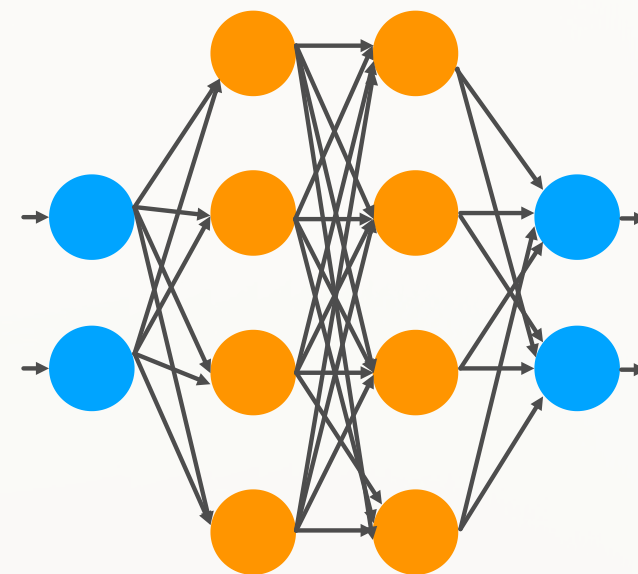
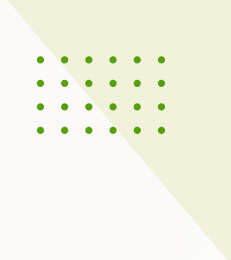


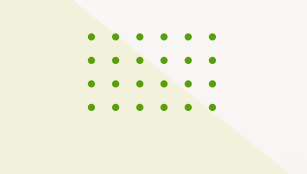


Modelos de IA Generativa





O que são modelos de
IA Generativa?

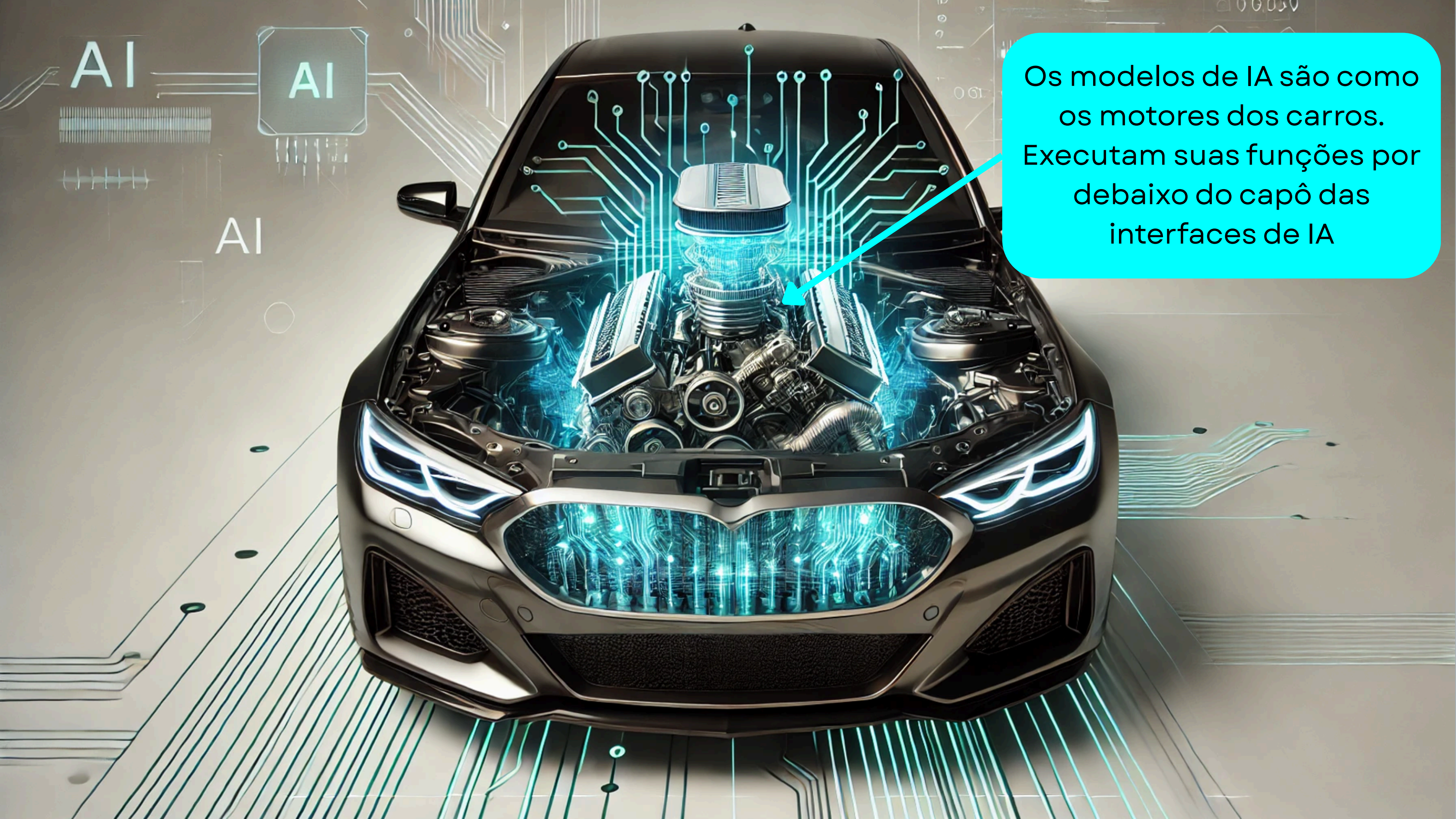


O QUE SÃO MODELOS DE IA?

Modelos de Inteligência Artificial são **algoritmos matemáticos** que aprendem padrões e tomam decisões a partir de dados, sem serem explicitamente programados para cada tarefa.

A ideia central é que, em vez de um programador escrever regras para cada cenário possível, o modelo é treinado e "aprende" essas regras ou padrões por meio da exposição a grandes volumes de dados.

O resultado de todo o processo de treinamento é o modelo em si.



Os modelos de IA são como os motores dos carros. Executam suas funções por debaixo do capô das interfaces de IA

"Assim como um mesmo modelo de carro pode ser equipado com diferentes tipos de motores, cada um com suas características de desempenho e consumo ... "

Versões

O Uno está disponível em várias versões, em 3 e 5 portas:

Uno/Mille:

- *Uno S*
- *Uno CS*
- *Uno CS Top*
- *Uno CSL*
- *Uno ELX*
- *Uno SX*
- *Uno 1.5 R*
- *Uno 1.6 R*
- *Uno Mille*
- *Mille Brio*
- *Uno Turbo i.e.*
- *Mille Eletronic*
- *EP*
- *EX*
- *Smart*
- *Mille i.e.*
- *Mille Fire*
- *Mille Economy*
- *Mille Way*
- *Grazie Mille*



"...uma mesma ferramenta de IA Generativa pode utilizar diferentes modelos de linguagem, de acordo com os objetivos e necessidades do usuário."


ChatGPT
GPT-5
GPT-4o
GPT-4.1
o3, o4
Dall-e 2
Sora


Gemini
2.5 Flash
2.5 Pro
Veo3


Grok
grok-4
grok-3
grok-3-mini


DeepSeek
deepseek-V3
deepseek-R1


Claude
Sonet4
Opus 4

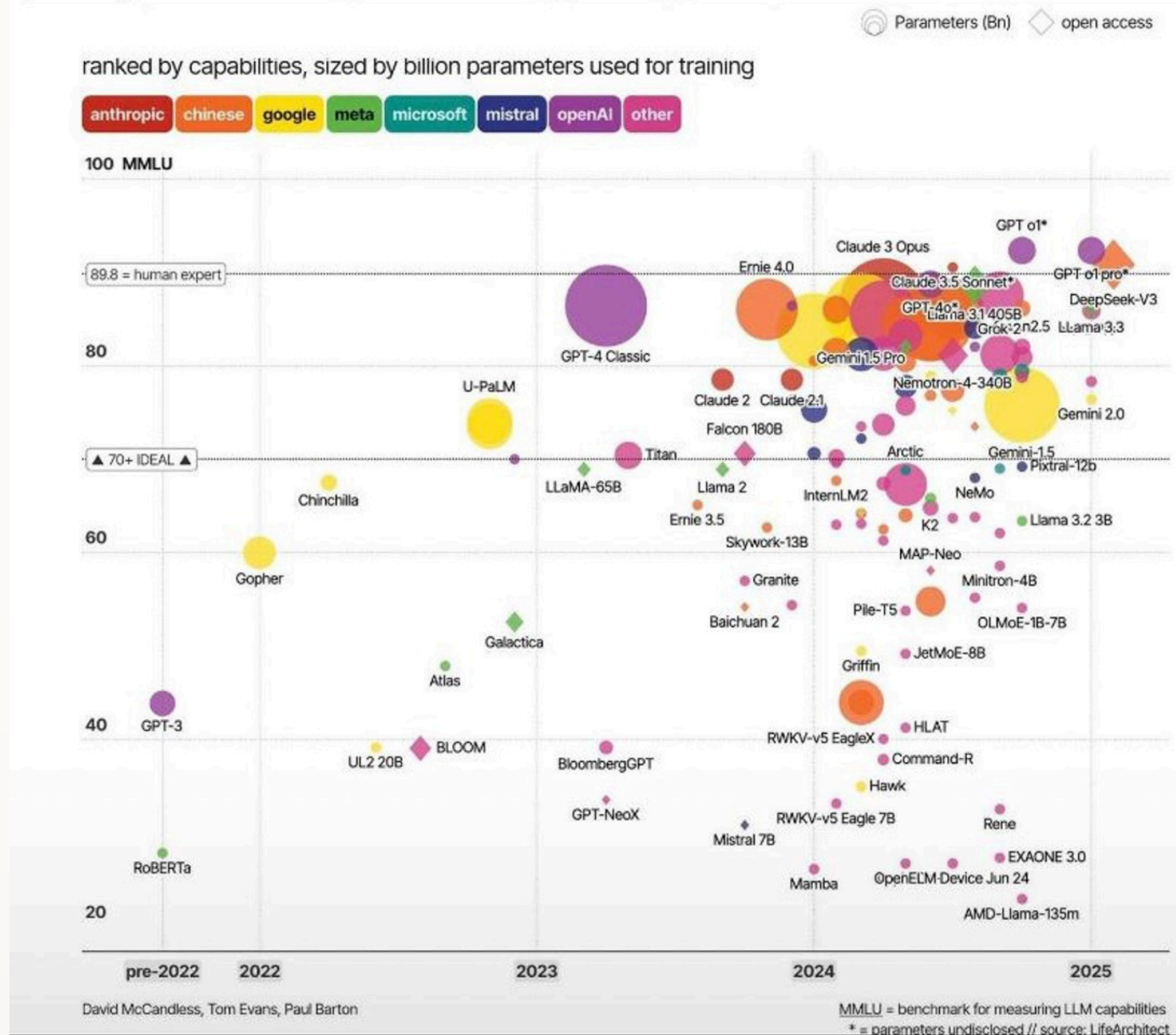

Llama
Llama4 series


Manus
Não tem
modelo
próprio


**Hugging
Face**
Uma
infinidade
de modelos
abertos

Evolução na Construção de Modelos

Major Large Language Models (LLMs)





Como os modelos são
treinados?

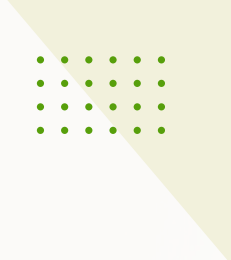


COMO OS MODELOS SÃO TREINADOS?

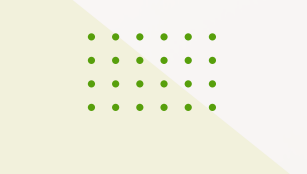




Vídeo: <https://www.youtube.com/watch?v=LPZh9BOjkQs>



O que os Modelos
NÃO são?



O QUE OS MODELOS NÃO SÃO?

Não são Bancos de Dados

Os modelos não armazenam informações de forma estruturada para consulta direta.

Não funcionam como busca exata

O usuário não faz uma pergunta esperando sempre a mesma resposta, como em uma consulta SQL ou em um motor de busca tradicional.



Não têm memória dos dados de treinamento

Os modelos aprendem padrões estatísticos, mas não conseguem "lembrar" ou "recuperar" documentos exatos usados no treinamento.

Geração é probabilística

Cada resposta é uma previsão com base nos padrões aprendidos, o que significa que, para a mesma pergunta, o modelo pode dar respostas diferentes em momentos diferentes.

Limitações dos Modelos

LIMITAÇÕES DOS MODELOS

Falta de entendimento real

O modelo **não compreende** o significado das palavras como um humano. Ele apenas calcula probabilidades de sequência de tokens com base no que foi aprendido.

Janela de contexto limitada

O modelo só consegue trabalhar com uma **quantidade restrita** de informação por vez (exemplo: até 128 mil tokens no GPT-4o).

Atualização de conhecimento

Os modelos têm um **corte de conhecimento**, ou seja, eles sabem apenas o que foi incluído até a data final de seu treinamento. Eventos recentes não são conhecidos (a menos que haja integração com fontes externas).

Alucinações (erros de resposta)

Os modelos podem **gerar informações incorretas** ou inventadas com aparência de verdade. Isso acontece porque eles funcionam por previsão de padrões, não por consulta a fontes confiáveis.

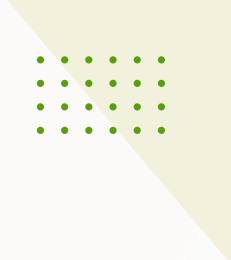
Dependência de dados de entrada

A qualidade da resposta depende muito da clareza, precisão e qualidade do **prompt** enviado pelo usuário.

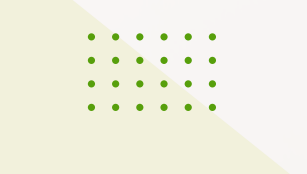
Viés e conteúdo sensível

Como foram treinados com grandes volumes de dados da internet, os modelos podem reproduzir **viéses** sociais, culturais ou até linguísticos, mesmo com filtros implementados.





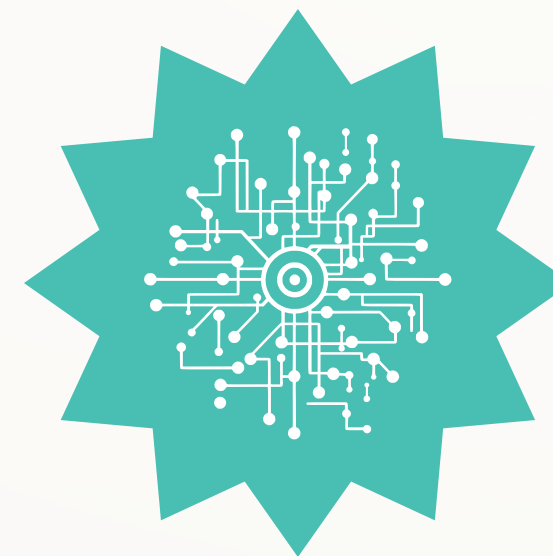
Classificações dos Modelos de IA Generativa



POR PROPÓSITO OU FUNÇÃO

Modelos Fundacionais (Foundation Models)

Modelos de grande escala, treinados em grandes volumes de dados diversos, que podem ser adaptados para muitas tarefas. Exemplos: GPT-4, Claude, LLaMA.



Modelos de Raciocínio (Reasoning Models)

Focados em realizar tarefas que exigem múltiplas etapas de lógica, como resolver problemas, inferir relações ou seguir instruções complexas. Exemplos: o3, o4, Claude 5Opus, Gemini 2.5. Pro

POR MODALIDADE

Modelos Monomodais

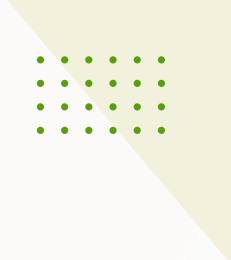
Operam com um único tipo de dado, como texto ou imagem.
Exemplos: GPT-3, DeepSeek (texto), DALL·E (imagem).



Modelos Multimodais

Compreendem e geram múltiplos tipos de dados (texto, imagem, áudio, vídeo).
Exemplos: GPT-4o (texto, imagem, áudio), Gemini 2.5, Grok 3.





Ranking de Modelos de IA Generativa

Ranking em
08 de setembro de 2025



LMarena

New Chat

Leaderboard

Take your chats anywhere

Create an account to save your chat history across your devices.

Login

Send Feedback

Report Bugs

Terms of Use

Privacy Policy

Cookies

Overall

Search by model name...

Default

Rank (UB) ↑	Model ↑↓	Score ↑↓	95% CI (±) ↑↓	Votes ↑↓	Organization ↑↓	License ↑↓
1	 gemini-2.5-pro	1456	±5	35.405	Google	Proprietary
1	 gpt-5-high	1447	±7	11.405	OpenAI	Proprietary
1	 claude-opus-4-1-20250805-thinking-16k	1447	±7	8.615	Anthropic	Proprietary
2	 o3-2025-04-16	1444	±4	40.935	OpenAI	Proprietary
2	 chatgpt-4o-latest-20250326	1443	±4	36.773	OpenAI	Proprietary
2	 gpt-4.5-preview-2025-02-27	1439	±6	15.271	OpenAI	Proprietary
2	 claude-opus-4-1-20250805	1436	±6	11.548	Anthropic	Proprietary
7	 gpt-5-chat	1426	±7	8.585	OpenAI	Proprietary
8	 grok-4-0709	1422	±6	18.239	xAI	Proprietary
8	 kimi-k2-0711-preview	1421	±5	18.588	Moonshot	Modified MIT
8	 deepseek-v3.1	1419	±9	18.111	DeepSeek	MIT

View all

Text Arena

View rankings across various LLMs on their versatility, linguistic precision, and cultural context across text.

Last Updated
Aug 28, 2025

Total Votes
4,007,423

Total Models
238

Overall

Search by model name...

Default

Rank (UB) ↑	Model ↑↓	Score ↑↓	95% CI (±) ↑↓	Votes ↑↓	Organization ↑↓	License ↑↓
1	 gemini-2.5-pro	1456	±5	35.405	Google	Proprietary
1	 gpt-5-high	1447	±7	11.405	OpenAI	Proprietary
1	 claude-opus-4-1-20250805-thinking-16k	1447	±7	8.615	Anthropic	Proprietary
2	 o3-2025-04-16	1444	±4	40.935	OpenAI	Proprietary
2	 chatgpt-4o-latest-20250326	1443	±4	36.773	OpenAI	Proprietary
2	 gpt-4.5-preview-2025-02-27	1439	±6	15.271	OpenAI	Proprietary
2	 claude-opus-4-1-20250805	1436	±6	11.548	Anthropic	Proprietary
7	 gpt-5-chat	1426	±7	8.585	OpenAI	Proprietary
8	 grok-4-0709	1422	±6	18.239	xAI	Proprietary
8	 kimi-k2-0711-preview	1421	±5	18.588	Moonshot	Modified MIT
8	 deepseek-v3.1	1419	±9	18.111	DeepSeek	MIT

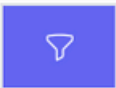
View all

Artificial Analysis Intelligence Index

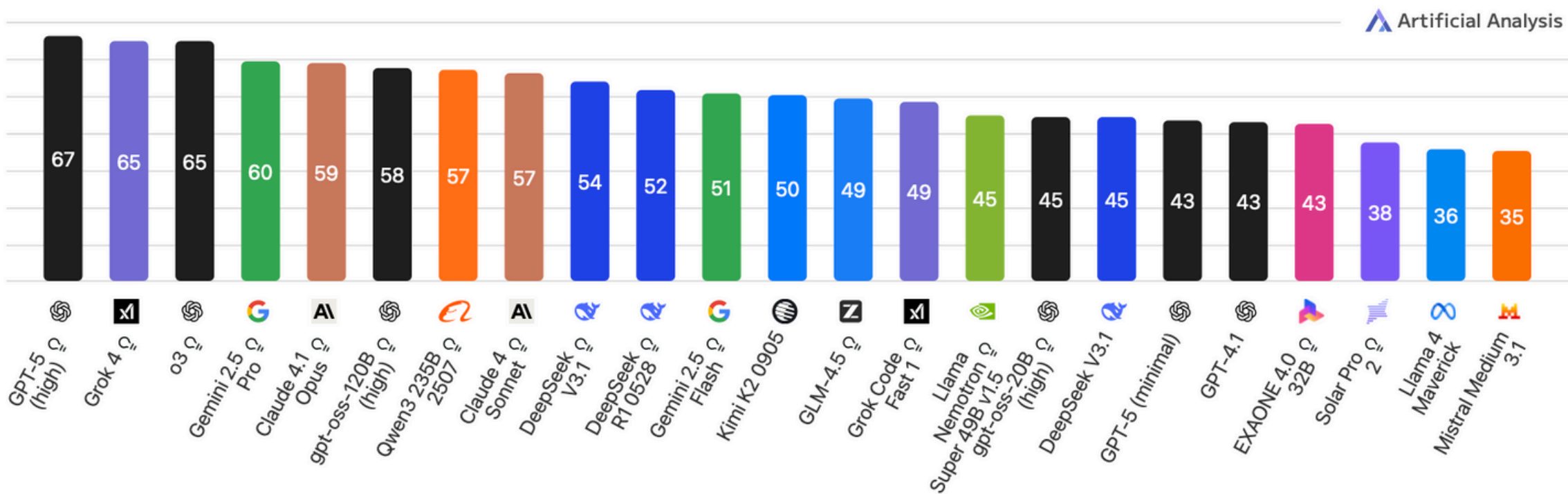
Artificial Analysis Intelligence Index v3.0 incorporates 10 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME 2025, IFBench, AA-LCR, Terminal-Bench Hard, τ^2 -Bench Telecom



23 of 267 models selected



+ Add model from specific provider



Artificial Analysis Intelligence Index: Combination metric covering multiple dimensions of intelligence - the simplest way to compare how smart models are. Version 3.0 was released in September 2025 and includes: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME 2025, IFBench, AA-LCR, Terminal-Bench Hard, τ^2 -Bench Telecom. See [Intelligence Index methodology](#) for further details, including a breakdown of each evaluation and how we run them.



E Agora!
Qual modelo eu uso?



Assim como não existe o "Melhor carro para tudo!", também não existe o "Melhor modelo para tudo!"

A escolha do modelo depende da sua necessidade: tipo de tarefa, complexidade, formato dos dados e recursos disponíveis.

Conhecer as classificações ajuda a escolher o modelo mais eficiente para cada caso.

Types of Cars



SUV



CROSSOVER



SEDAN



PICKUP TRUCK



HATCHBACK



CONVERTIBLE



LUXURY



COUPE



HYBRID/ELECTRIC



MINIVAN



SPORTS CAR



STATION WAGON

Conclusões

CONCLUSÕES



Os modelos, treinados intensivamente com vastos dados e redes neurais, aprendem padrões para gerar resultados inovadores.



É crucial entender que eles não são bancos de dados, operam probabilisticamente e **têm limitações** como contexto restrito e potencial para vieses ou "alucinações".



A escolha do modelo ideal depende da necessidade específica do usuário, não existindo um "melhor para tudo".



AUDITORIA INTELIGENTE

<http://auditoriainteligente.com.br>
[in/andrecorumba](#)