

Boas vindas

DANIEL SÓRIA
in



Otimizando LLM's com Fine-tuning

DANIEL SÓRIA



O que é Fine-tunning?

LLM's - Large Language Models

- Modelos de gerais treinados com uma vasta quantidade de dados. (Large)
- Aprender padrão linguísticos (Language)
- Realizar várias tarefas como tradução, classificação, geração de texto.



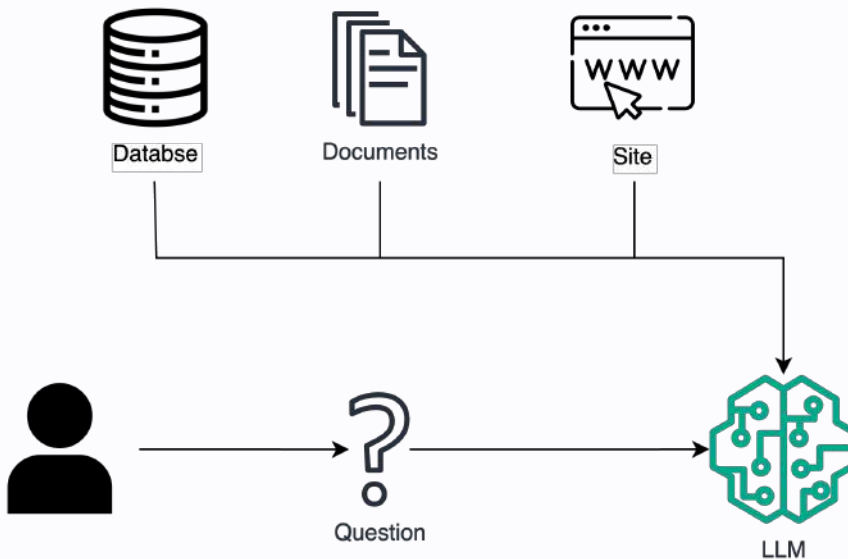
ChatGPT

LLaMA
by  **Meta**


Gemini

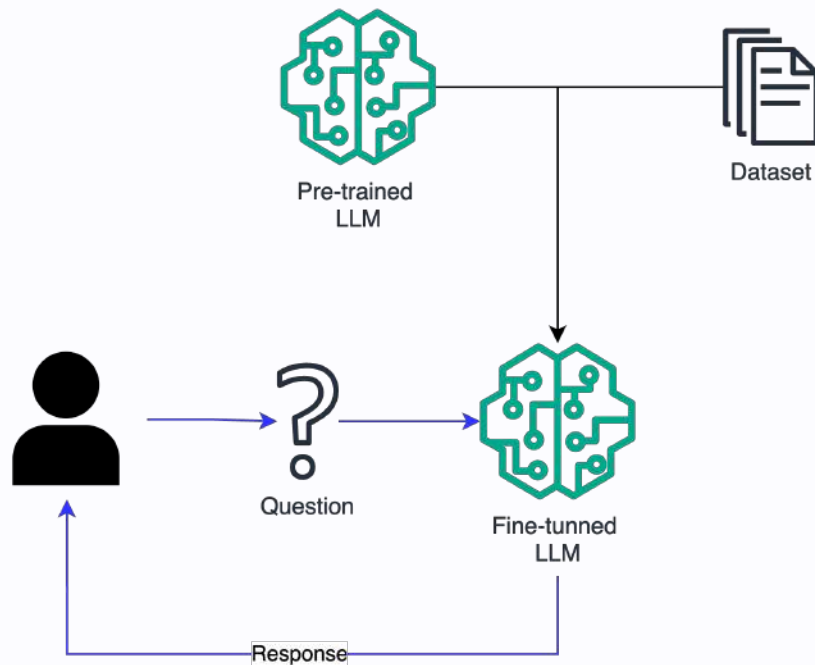
LLM's - Large Language Models

- Limitados a base (gigantesca) de treinamento
- Dados públicos (ou não) gerais
- Não atendem necessidades específicas de cada empresa/ pessoa



LLM's - Large Language Models

- Task específica
- Linguagem específica
(Medicina, jurídica)





LLM's - Large Language Models

- Task específica



Dataset


text string · <i>lengths</i>	category string · <i>classes</i>
	
georgia jagger lead airplane runway ppq british airways photos georgia jagger take famous stem sea level inventive runway...	STYLE & BEAUTY
entrepreneur john mark seltzer invaluable advice s USA t strip music producer ok soo grow san diego attend private school ge...	ENTERTAINMENT
david bowie know cancer terminal month death new documentary shed light bowie final year	ENTERTAINMENT
think paul manafort trump elect know paul manafort trump yanukovych share political brain operative name paul manafort	POLITICS
republican go cuba president obama weekend thing opportunity sell cheese	POLITICS
dr dean ornish talks stress reverse disease metric women conference stress difference stress affect health hinge handl...	WELLNESS

LLM's - Large Language Models

- Linguagem específica
(Medicina, jurídica)

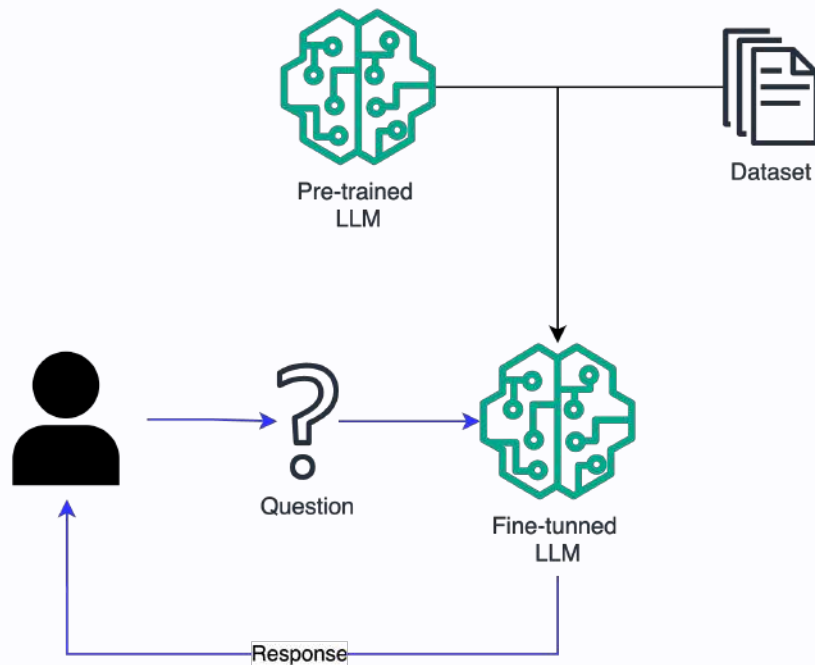


Dataset

question string · <i>lengths</i>	context string · <i>lengths</i>
	
What is the resurgent sodium current in mouse cerebellar Purkinje neurons?	FGF14 modulates resurgent sodium current in mouse cerebellar Purkinje neurons.
What isoform is a locus for inherited spinocerebellar at	Rapid firing of cerebellar Purkinje neurons is facilitated in part by a voltage-gated Na ⁺ (NaV)...
What is the name of the study that focuses on the study of the drug-	[Drug-related Anaphylactic Shocks: Under-reporting and PMSI.]
What was the code for anaphylactic shock unspecified?	Aim. To evaluate the value of research in the case-mix database to identify cases of drug-related...
What is the cost of Fever Evaluation Cost and Time to Treatment in the General Surgery Patient?	Fever Evaluation Cost and Time to Treatment in the General Surgery Patient Can be Reduced by Using a...
What is the average cost of an FAL work-up?	OBJECTIVES: Evaluation of surgical patients with fever and leukocytosis (FAL) for an infection source often...

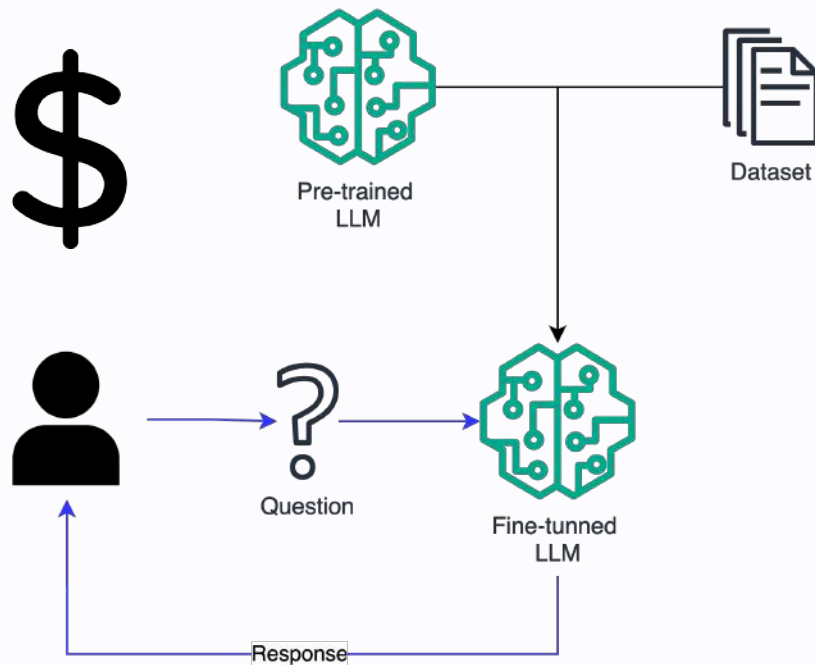
LLM's - Large Language Models

- Task específica
- Linguagem específica
(Medicina, jurídica)



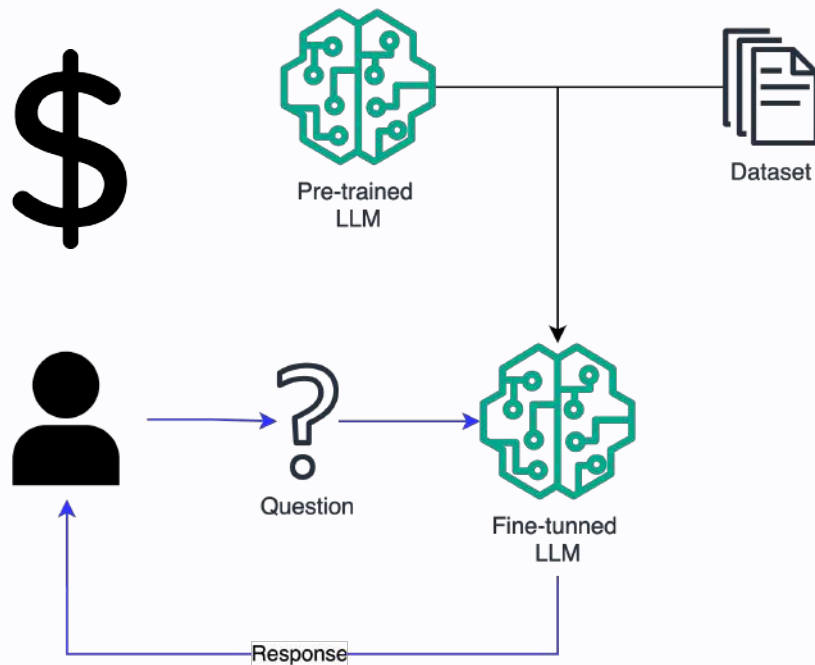
LLM's - Large Language Models

- Task específica
- Linguagem específica
(Medicina, jurídica)



LLM's - Large Language Models

- Task específica
- Linguagem específica
(Medicina, jurídica)



*Na maioria dos casos, LLM's não precisam de
Fine-tuning*

Fine-tuning via GPT

Fine-tuning via AWS

AWS Bedrock

Fine-tuning via AWS

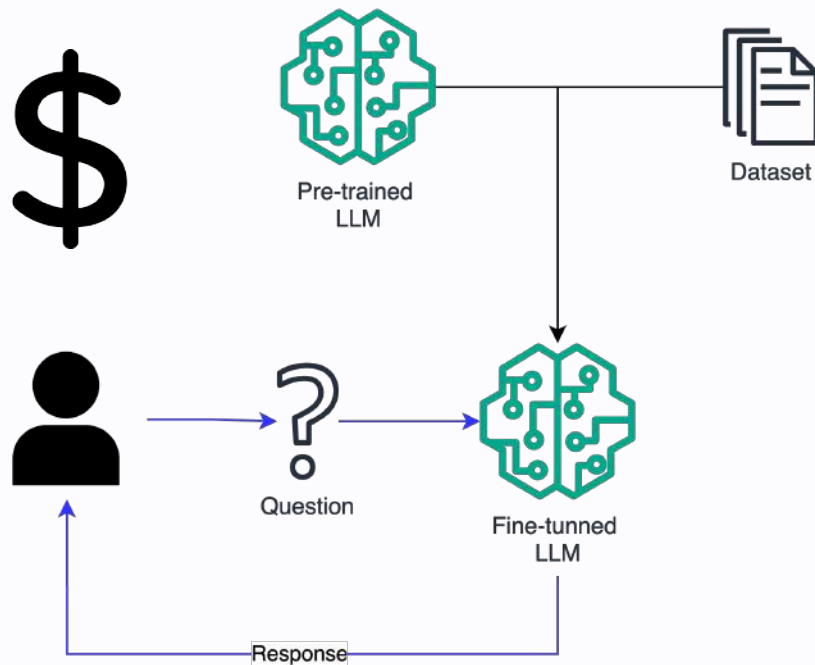


AWS: [Cupom de acesso](#)

LLM's SLM's e Outros modelos

LLM's - Large Language Models

- Task específica
- Linguagem específica
(Medicina, jurídica)



*Na maioria dos casos, LLM's não precisam de
Fine-tuning*

*Na maioria dos casos, ~~LLM's~~ não precisam de
Fine-tuning*

*Na maioria dos casos, SLM's e outros LM's precisam de
Fine-tuning*

Opções frente aos LLM's

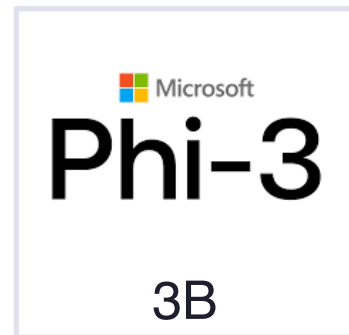


500B a 1T

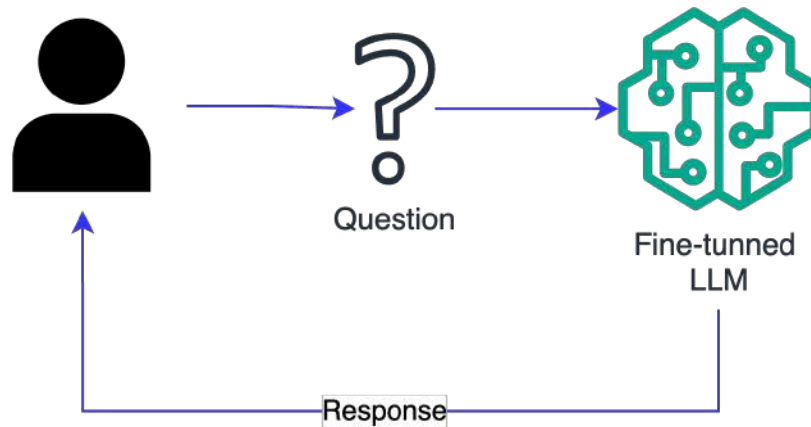
 **Llama 3**

The Llama logo consists of a blue infinity symbol. To its right, the word "Llama" is written in a black, sans-serif font, followed by a large blue number "3".

405B



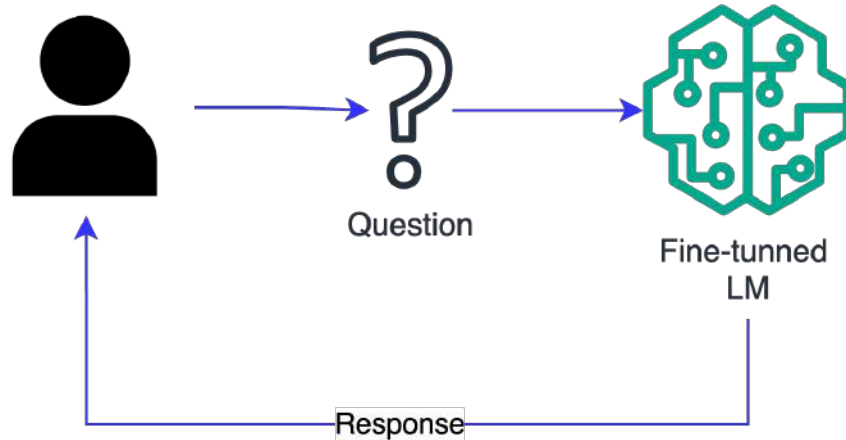
Opções frente aos LLM's



Hate Speech
No Hate Speech

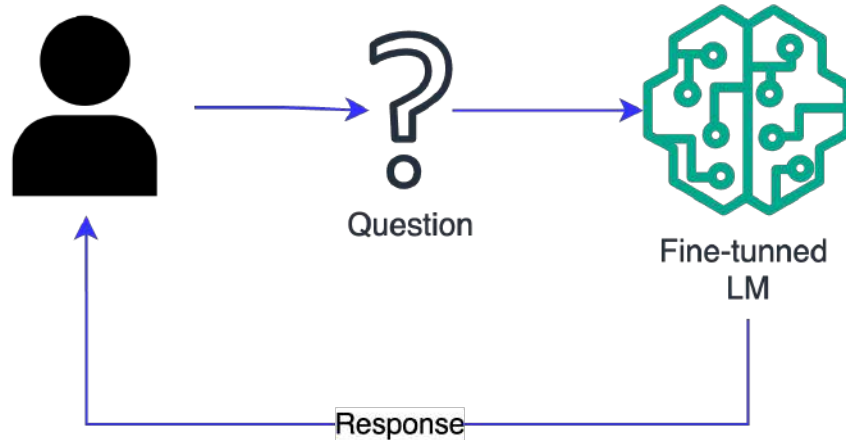
Opções frente aos LLM's

Language Model
Sem geração de texto



Opções frente aos LLM's

Language Model
Sem geração de texto



BERT - Bidirectional Transformers for Language Understanding



Fine-tuning Localmente

Fine-tuning BERT

