Embarcadero Conference 2024

Inovação faz parte do nosso DNA!

Técnicas de Engenharia de Prompt para desenvolvedores



{Carlos Eduardo Paulino }

- Desenvolvedor Delphi há mais de 25 anos.
- Criador dos canais @DelphiCleanCode
- Dev Backend com sistemas ERP e CRM 3º Setor
- Dev componentes TAsaas, TPagSeguro, TMercadoPago, TPicPay etc.

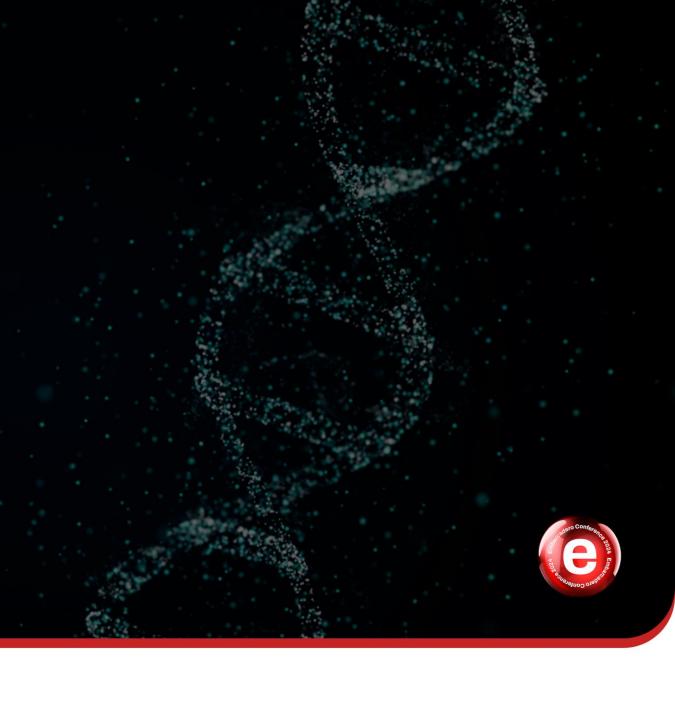




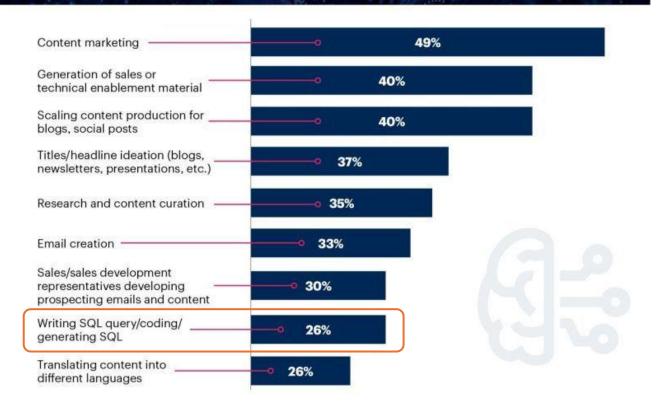








Generative Al use cases



Q. What generative AI use case(s) is your company currently implementing/ planning to implement? Source: Gartner



^{© 2024} Gartner, Inc. and/or its affiliates. All rights reserved. 2952678.

Agenda

- IA Uma visão geral
- LLM conceitos e aplicações
- Arquitetura Transformers
- Engenharia de PROMPT
- Principais Técnicas de PROMPT
- Exemplos e dicas



Artificial Intelligence

A IA envolve técnicas que equipam os computadores para emular o comportamento humano, permitindo-lhes aprender, tomar decisões, reconhecer padrões e resolver problemas complexos de uma forma semelhante à inteligência humana.

Machine Learning

Criação e utilização de algoritmos avançados para detectar padrões em grandes conjuntos de dados, permitindo que as máquinas aprendam e se adaptem.

Deep Learning

Utilização de redes neurais para processamento aprofundado de dados e tarefas analíticas. DL aproveita múltiplas camadas de redes neurais artificiais para extrair recursos de alto nível de dados brutos de entrada, simulando a maneira como o cérebro humano percebe e entende o mundo.

Generative Al

IA generativa é um subconjunto de modelos de Deep Learning que gera conteúdo como texto, imagens ou código com base nas informações fornecidas. Treinados em vastos conjuntos de dados, estes modelos detectam padrões e criam resultados sem instrução explícita, utilizando uma combinação de aprendizagem supervisionada e não supervisionada.

Artificial Intelligence

Machine Learning

Deep Learning

Generative Al

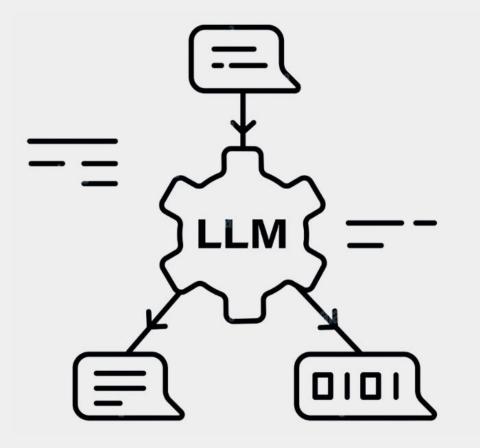
LLM (Large Language Model)

LLM é um modelo de linguagem avançado que utiliza algoritmos de processamento de linguagem natural para entender e gerar texto.

Ele é treinado com grandes volumes de dados textuais, para aprender a estrutura da linguagem e as relações semânticas entre palavras e frases.

Por exemplo, o GPT (Generative Pre-trained Transformer) é um modelo de linguagem avançado que utiliza a arquitetura de transformadores para processar texto de maneira eficiente, capturando dependências de longo alcance.

Ele é treinado em tarefas de previsão de palavras ou frases, permitindo que gere texto coerente e de qualidade com base no contexto fornecido



Onde um LLM é aplicado?



- **Geração de texto**: Gerar automaticamente resumos de texto, artigos, redações e até mesmo diálogos fictícios.



- **Tradução automática**: Traduzir texto entre diferentes idiomas, fornecendo resultados mais precisos e fluentes.



- **Assistentes virtuais**: Chatbots e sistemas de perguntas e respostas são alimentados por LLMs para entender e gerar respostas coerentes em tempo real.



- **Análise de sentimentos**: Identificação de textos como positivos, negativos ou neutros. Isso é útil em áreas como análise de mídias sociais e pesquisa de opinião.



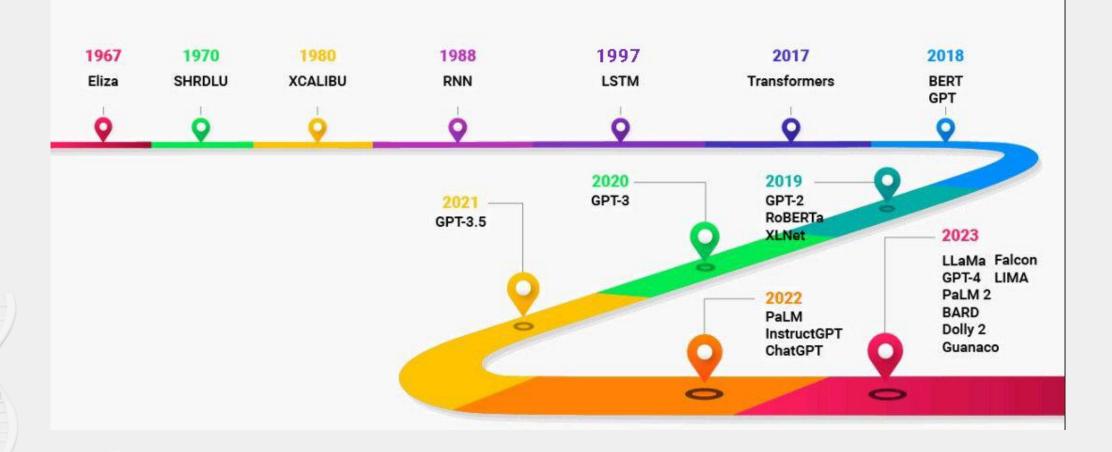
- **Recomendação de conteúdo**: Plataformas de streaming, comércio eletrônico e mídias sociais utilizam LLMs para recomendar conteúdo personalizado com base nos interesses e comportamentos dos usuários.



- **Correção automática**: Teclados e editores de texto com sugestões de palavras e corretor ortográficos e gramaticais.



Evolution of Large Language Models



1988 - 1997 - RNN e LSTM

RNNs são um tipo de arquitetura de rede neural projetada para lidar com dados sequenciais, como texto.

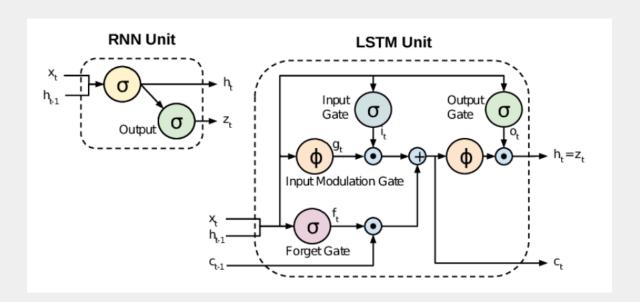
Elas são capazes de manter uma "memória" interna que permite processar sequências de entrada e gerar sequências de saída.

No entanto, as RNNs tradicionais sofrem de um problema conhecido como "vanishing gradient", que dificulta a aprendizagem de dependências de longo prazo em sequências de texto.

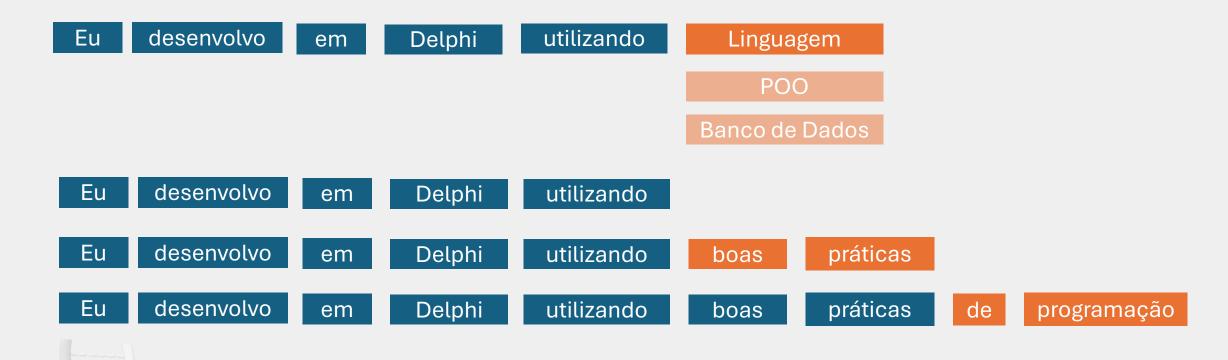
LSTMs foram propostas. Elas são uma variação das RNNs que incorpora mecanismos de "portas" para controlar o fluxo de informações na rede.

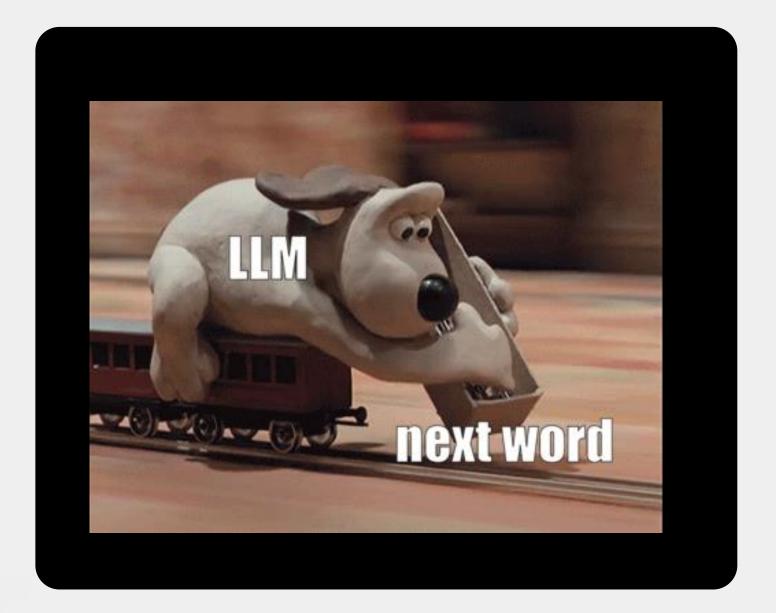
Isso permite que as LSTMs capturem dependências de longo prazo em dados sequenciais de forma mais eficaz, tornando-as ideais para tarefas de processamento de linguagem natural.





1988..1997 - RNN e LSTM







2017 – Arquitetura Transformer

Os Transformers são um tipo de modelo baseado em atenção que gerencia melhor o processamento de sequências de dados, como texto.

A arquitetura Transformer permitiu grandes avanços na geração de texto, resultando em modelos como o GPT.

Esses modelos podem compreender e gerar texto com uma coerência, relevância e diversidade sem precedentes, impactando diretamente aplicações como tradução automática, geração de conteúdo e chatbots.



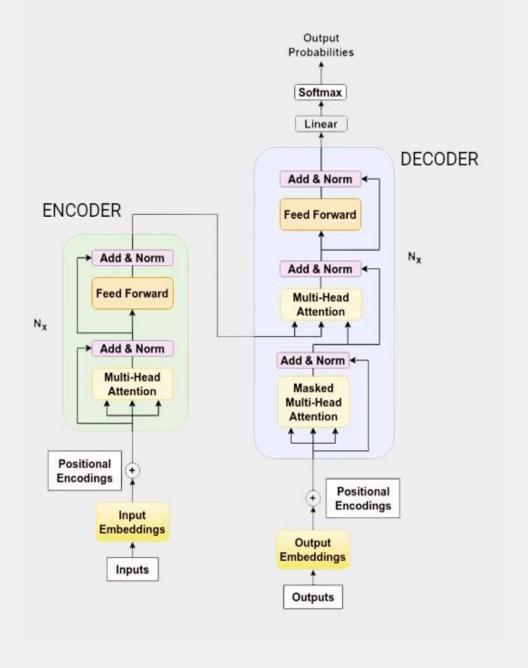


2017 – Arquitetura Transformer

Os Transformers são um tipo de modelo baseado em atenção que gerencia melhor o processamento de sequências de dados, como texto. A arquitetura Transformer permitiu grandes avanços na geração de texto, resultando em modelos como o GPT.

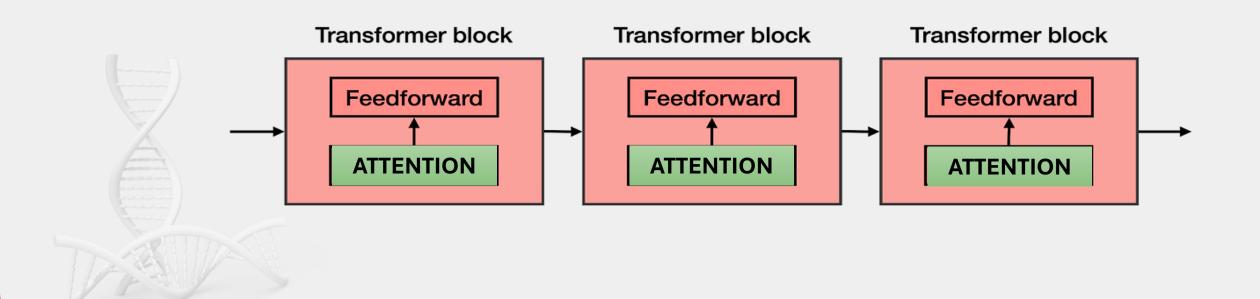
Esses modelos podem compreender e gerar texto com uma coerência, relevância e diversidade sem precedentes, impactando diretamente aplicações como tradução automática, geração de conteúdo e chatbots.

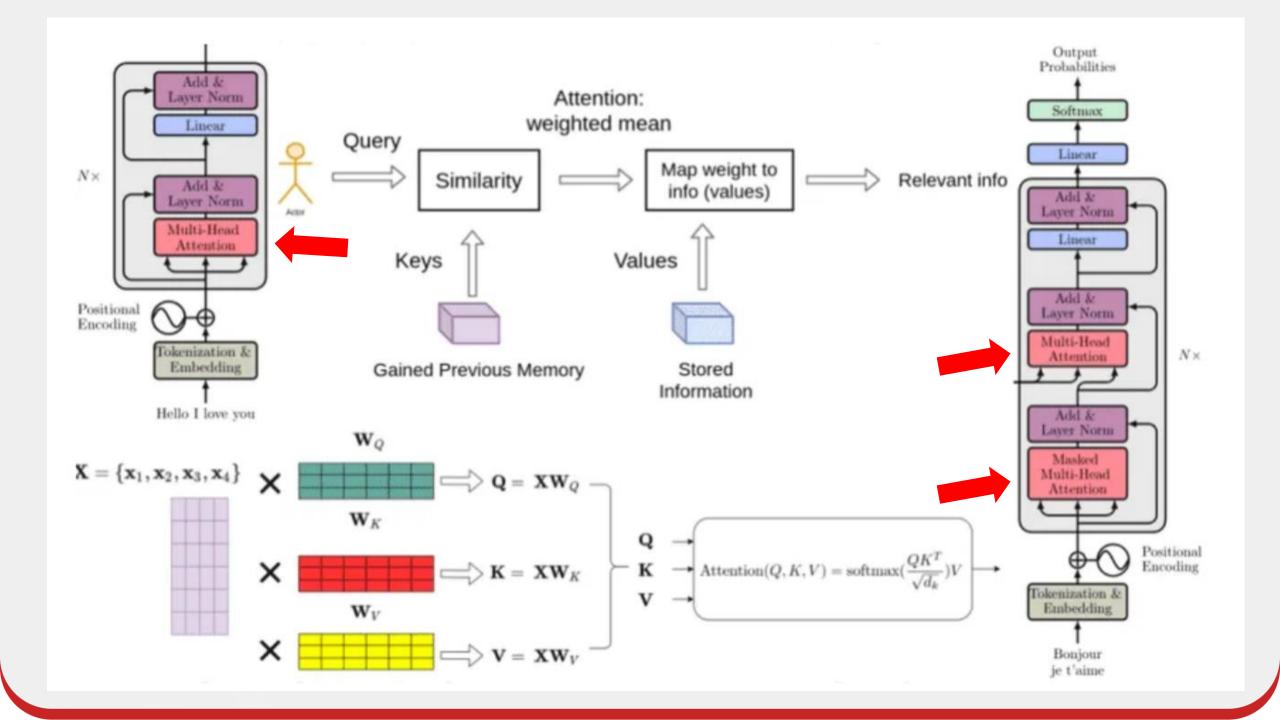


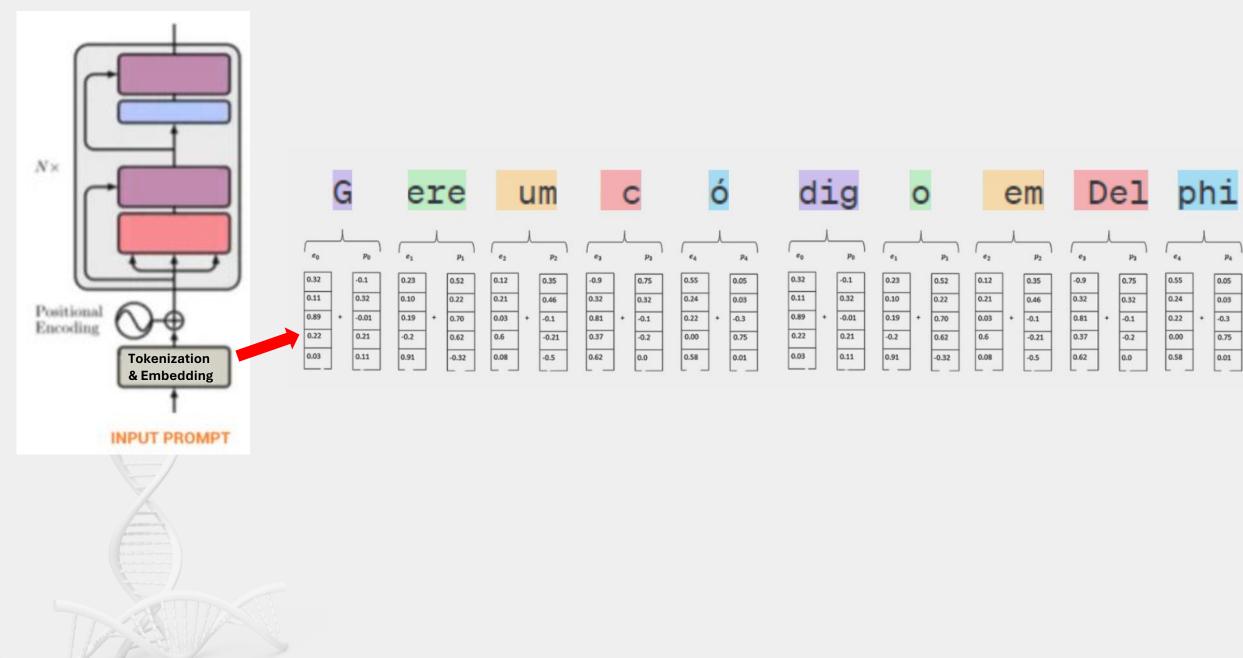


Em 2017 um dos marcos mais importantes para a GenAl aconteceu:

A publicação do artigo "Attention is All You Need" por pesquisadores do Google Brain, marcou uma grande evolução nos modelos de linguagem, porque os Transformers conseguem capturar contextos mais amplos e realizar paralelismo em seu processamento, o que resulta em um treinamento mais eficiente em grandes conjuntos de dados







 p_4

0.05

0.03

0.75

0.01

+ -0.3

0.55

0.24

0.22

0.00

0.58

PROMPT

Gere método em Delphi para geração de um gráfico, utilizando o componente TChart, considerando os seguintes <valores>...</valores>

Tokens Characters

33 130

GPT-4

Gere método em Delphi para geração de um gráfico, utilizando o componente TChart, considerando os seguintes <valores>...</valores>

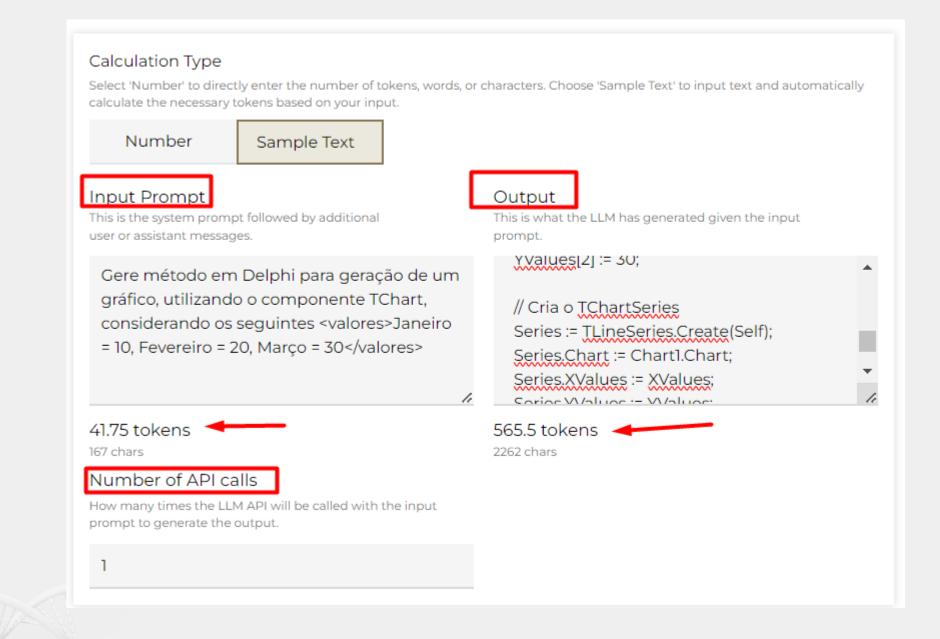
Tokens

Characters

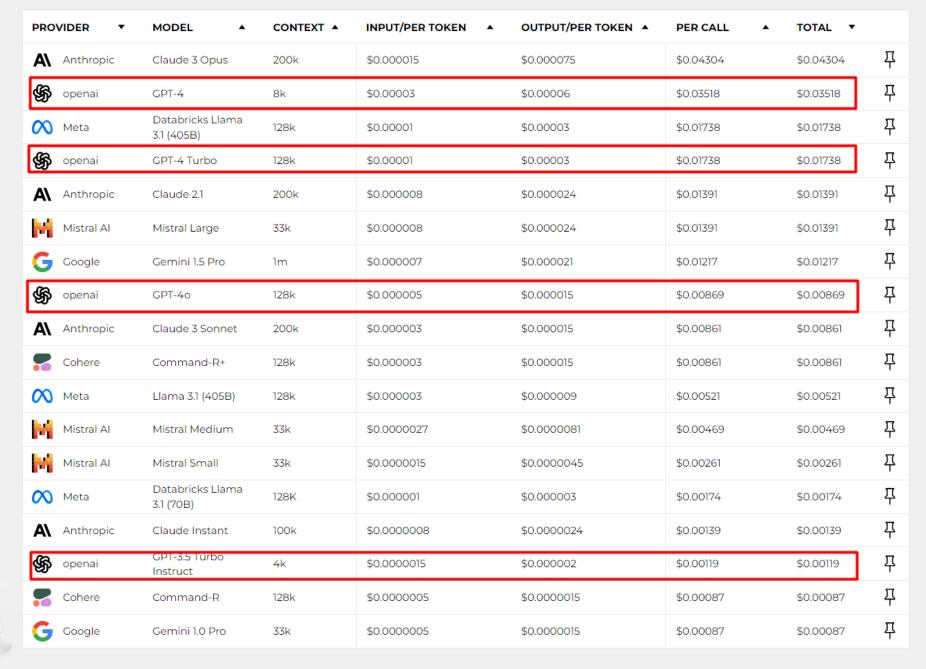
45 130

GPT-3

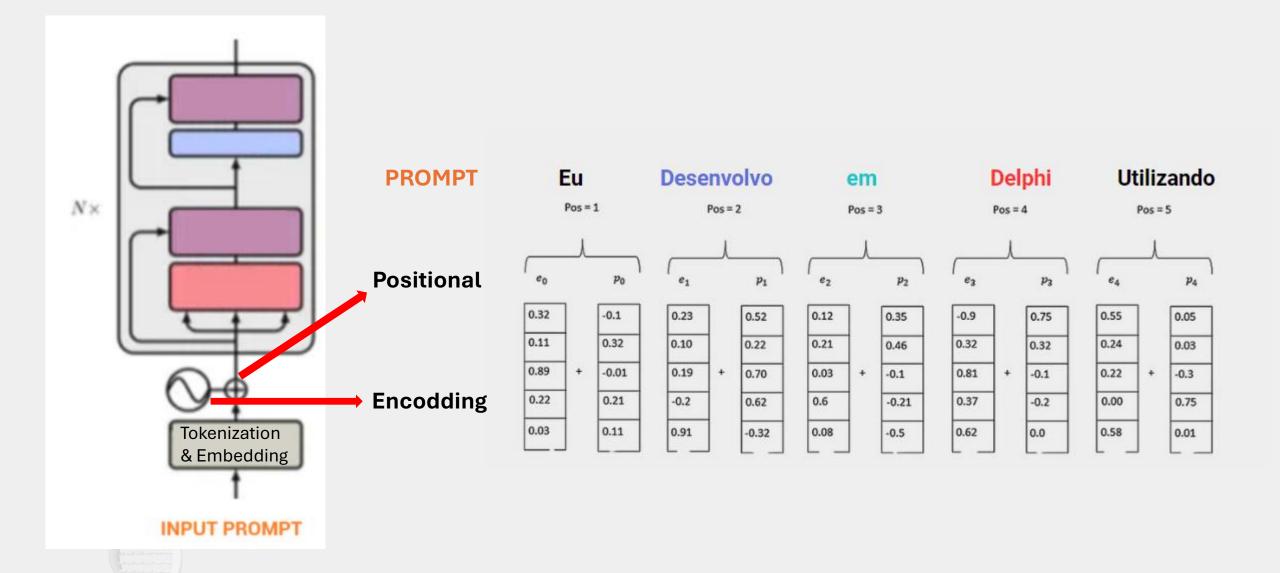
Gere método em Delphi para geração de um gráfico, utilizando o componente TChart, considerando os seguintes <valores>...</valores>

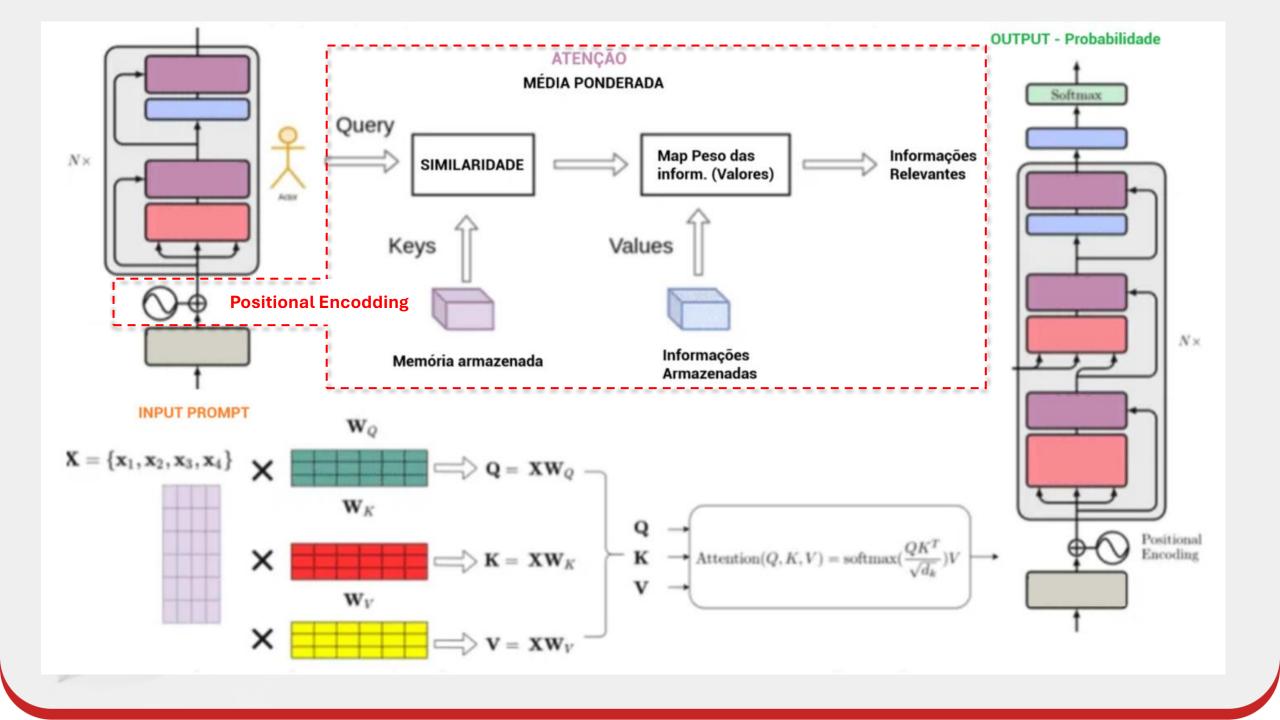


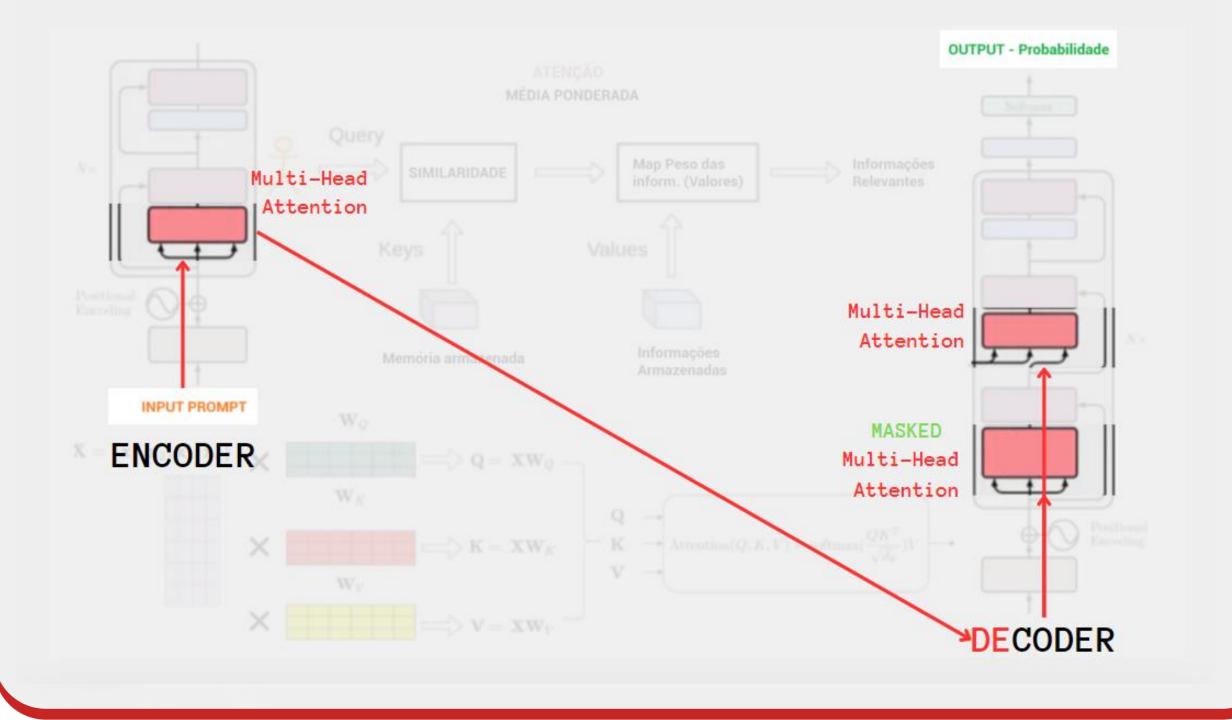
https://www.tribalscale.com/our-services/llm-calculator

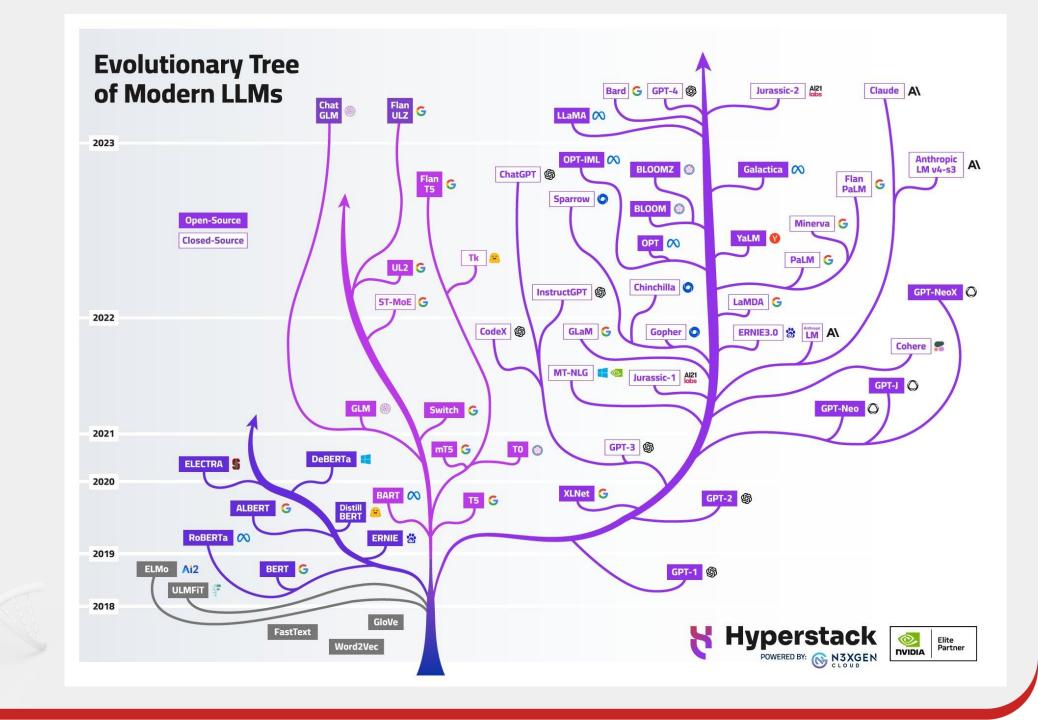


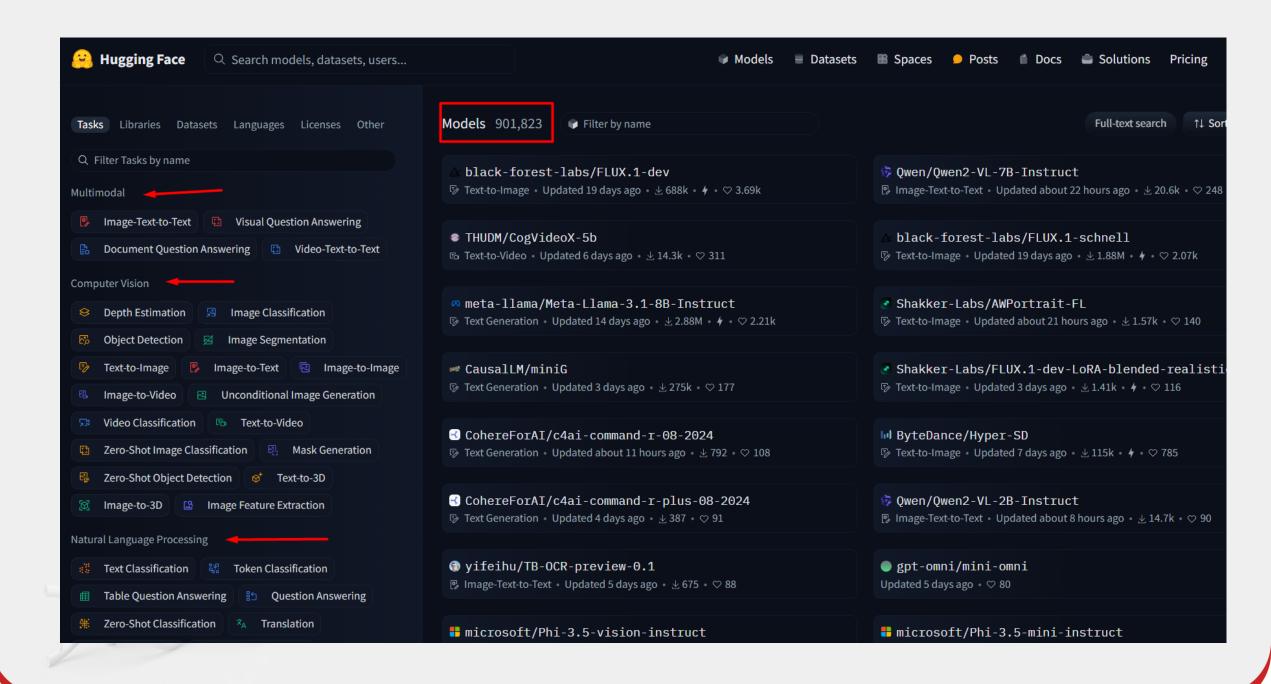












O que é Engenharia de PROMPT

É a arte de **formular instruções claras** para **obter respostas precisas** de modelos de IA, combinando habilidades linguísticas e técnicas específicas.

A chave é fornecer o máximo de contexto possível em um prompt bem estruturado.



Porque eu preciso de engenharia de PROMPT?







como calibrar pneu de trem?



Calibrar os pneus de um trem é um processo importante para garantir a segurança e a eficiência do veículo ferroviário. O processo de calibração dos pneus de um trem é semelhante ao de um veículo comum, mas existem algumas diferenças importantes a serem observadas.

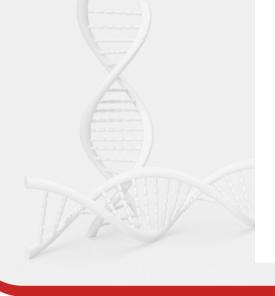
- Verifique a pressão recomendada: Antes de começar a calibrar os pneus, verifique a pressão recomendada pelo fabricante do trem. Isso geralmente pode ser encontrado no manual do proprietário ou nas especificações técnicas do trem.
- Prepare os equipamentos de calibração: Você precisará de um medidor de pressão de pneus e um compressor de ar para calibrar os pneus do trem.
 Certifique-se de que o medidor de pressão esteja calibrado corretamente antes de começar a medição.







- 3. Remova as tampas de válvula: Remova as tampas de válvula dos pneus do trem. Essas tampas são geralmente encontradas na parte superior do pneu e devem ser removidas com uma chave de válvula.
- 4. <u>Conecte o medidor de pressão</u>: Conecte o medidor de pressão à válvula do pneu e pressione-o firmemente para garantir uma conexão segura.
- Meça a pressão dos pneus: Verifique a pressão dos pneus do trem e anote a leitura. Repita o processo em cada um dos pneus para garantir que todos estejam calibrados corretamente.
- 6. Adicione ou libere ar: Se a pressão estiver abaixo do recomendado, adicione ar aos pneus usando um compressor de ar. Se a pressão estiver acima do recomendado, libere ar dos pneus usando a válvula de liberação.
- 7. Verifique novamente a pressão: Após adicionar ou liberar ar, verifique novamente a pressão dos pneus e ajuste conforme necessário.
- 8. Coloque as tampas de válvula de volta: Depois de calibrar todos os pneus do trem, coloque as tampas de válv 🗸 de volta.



É importante lembrar que a calibração dos pneus deve ser feita regularmente para garantir que o trem esteja operando de maneira segura e eficiente.





Principais Técnicas de Engenharia de Prompt

Zero Shot

Essa técnica envolve fornecer ao modelo uma tarefa que ele não viu durante seu treinamento.

Ela testa a capacidade do modelo de generalizar e produzir saídas relevantes sem depender de exemplos anteriores.

Zero Shot - Exemplo

##PROMPT

Dado o seguinte texto:

"Hoje foi um dia incrível! Acordei cedo, tomei um café e fui programar. Logo na primeira tarefa, resolvi um BUG milenar! Finalmente o cliente vai me dar paz rs. Estou me sentindo o ninja da programação!"

##RESPOSTA

Classifique o sentimento do texto em uma escala de 1 a 5, sendo 1 muito negativo e 5 muito positivo, sem usar nenhum modelo de linguagem treinado para essa tarefa específica.



Zero Shot - Exemplo



Para classificar o sentimento do texto apresentado, vamos considerar os elementos emocionais e as expressões utilizadas.

Elementos Positivos

- Entusiasmo: O autor expressa um sentimento de alegria ao descrever seu dia como "incrível".
- Realização: A resolução de um "BUG milenar" é mencionada com um tom de conquista, o que indica satisfação e orgulho.
- 3. Humor: A frase "Finalmente o cliente vai me dar paz rs" sugere um alívio e uma leveza na situação, mostrando que o autor está lidando bem com a pressão.
- Confiança: A expressão "me sentindo o ninja da programação" indica uma autoconfiança elevada e um sentimento de competência.

Elementos Negativos

Não há elementos claramente negativos no texto. A única menção que poderia ser interpretada como negativa é a referência ao "BUG milenar", mas isso é contrabalançado pela alegria de ter resolvido o problema.



Zero Shot - Exemplo

Classificação Final

Com base na análise acima, o sentimento do texto pode ser classificado como **5** na escala de 1 a 5, onde 1 é muito negativo e 5 é muito positivo. O texto reflete um estado de espírito otimista e satisfeito, típico de um dia produtivo e gratificante.



Principais Técnicas de Engenharia de Prompt

Few-shot

A técnica de Few-shot prompting, ou aprendizado em contexto, envolve fornecer exemplos que ajudam a guiar a resposta do modelo.

Few-shot - Exemplo

Prompt

Instruções: Abaixo estão alguns exemplos de sentimentos expressos em textos. Depois dos exemplos, forneça um novo texto e classifique o sentimento em uma escala de 1 a 5, onde 1 é muito negativo e 5 é muito positivo.

Exemplo 1:

 Texto: "Hoje foi um dia muito difícil. Enfrentei muitos problemas no trabalho e me senti sobrecarregado."

Sentimento: 2

Exemplo 2:

 Texto: "Tive um ótimo final de semana! Fui ao parque com amigos e nos divertimos muito."

Sentimento: 5

Exemplo 3:

 Texto: "Estou me sentindo um pouco ansioso com a apresentação de amanhã, mas acredito que tudo vai dar certo."

Sentimento: 3



Few-shot - Exemplo

Novo Texto: "Finalmente consegui terminar meu projeto! Estou tão feliz e aliviado, mal posso esperar para mostrar ao meu chefe."

Sentimento: 5



Principais Técnicas de Engenharia de Prompt

Chain-of-Thought

A técnica de Chain-of-Thought (Cadeia de Pensamento) envolve guiar o modelo a pensar passo a passo, detalhando seu raciocínio antes de chegar a uma conclusão.

Essa técnica é útil para resolver problemas complexos ou para classificar sentimentos de forma mais precisa.

Chain-of-Thought - Exemplo

Prompt

Instruções: Leia o texto abaixo e, em seguida, siga os passos para analisar o sentimento. Explique seu raciocínio antes de fornecer a classificação final em uma escala de 1 a 5, onde 1 é muito negativo e 5 é muito positivo.

Texto: "Hoje foi um dia desafiador. Tive que lidar com várias tarefas ao mesmo tempo e, no final, me senti exausto. No entanto, consegui completar tudo e isso me deixou um pouco satisfeito."

Passo a Passo

- Identificação de Emoções: Comece identificando as emoções expressas no texto. O autor menciona um "dia desafiador" e "me senti exausto", o que indica um sentimento negativo inicial.
- Análise de Conflito: Apesar da exaustão, o autor também menciona que "consegui completar tudo", o que sugere um sentimento de realização. Essa é uma emoção positiva que contrabalança a exaustão.
- 3. Peso das Emoções: Avalie a intensidade das emoções. A exaustão parece ser forte, mas a satisfação por ter completado as tarefas também é significativa. O autor se sente "um pouco satisfeito", o que é um sentimento positivo, mas não tão forte quanto o negativo.



Chain-of-Thought - Exemplo

Classificação Final: Considerando que o texto expressa tanto um sentimento negativo (exaustão) quanto um positivo (satisfação), mas que a exaustão parece dominar, eu classificaria o sentimento como 3. Isso indica um estado neutro, onde as emoções se equilibram.

Sentimento: 3



MARKDOWN

https://www.markdownguide.org/cheat-sheet/

LLM's Aprenderam com base em textos e replicam isso.

<XML>

O uso de XML em prompts não apenas melhora a organização e a clareza das instruções, mas também aumenta a precisão e a flexibilidade na interação com modelos de linguagem.

Esses benefícios tornam o XML uma escolha ideal para estruturar prompts complexos e garantir que as saídas sejam de alta qualidade e facilmente manipuláveis.

{{Variáveis}}

Utilizar variáveis facilita a sobreposição de respostas.



``` CÓDIGO ```

Um prompt com instruções de código caracteriza o formato de saída.

```
##Exemplo de código
unit uCliente;
interface
Uses System.SysUtils;
type
TCliente = class
...
```

Como criar meus prompts?

grogcloud

https://console.groq.com/playground

ANTHROP\C

https://console.anthropic.com

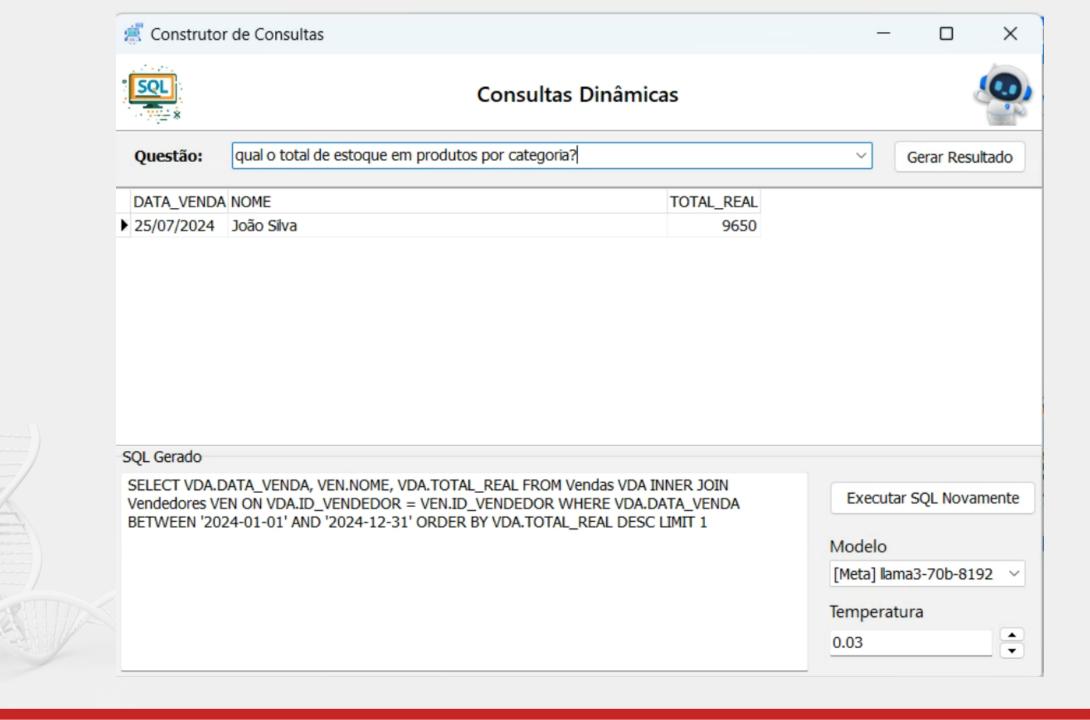


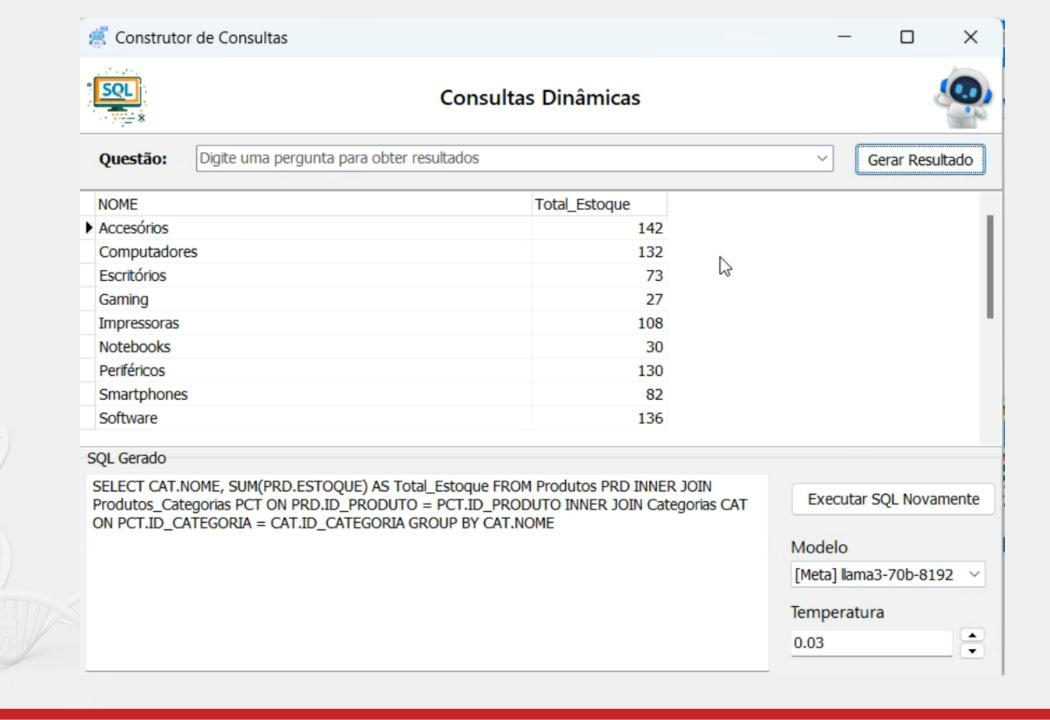
https://platform.openai.com/playground/

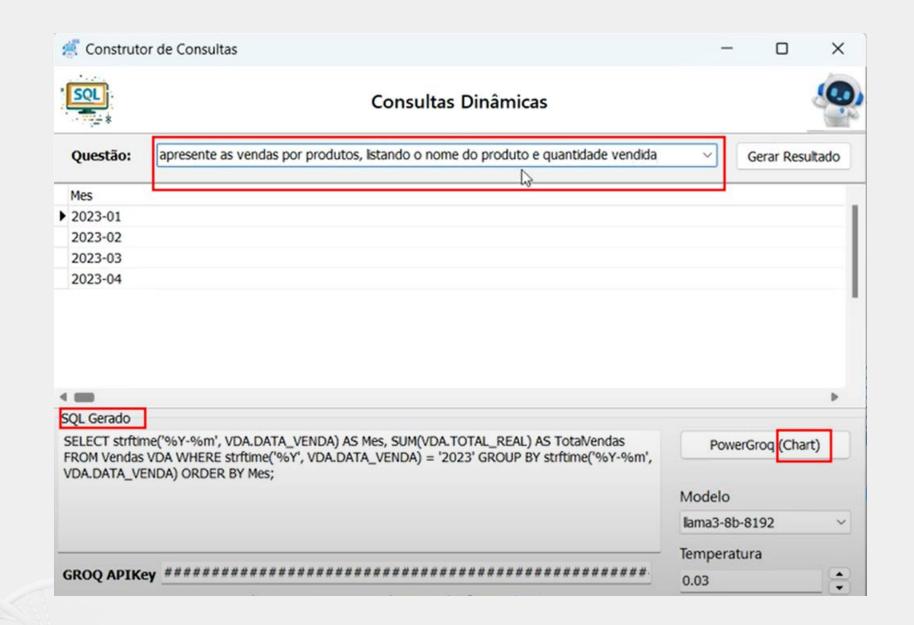
Exemplos de PROMPT

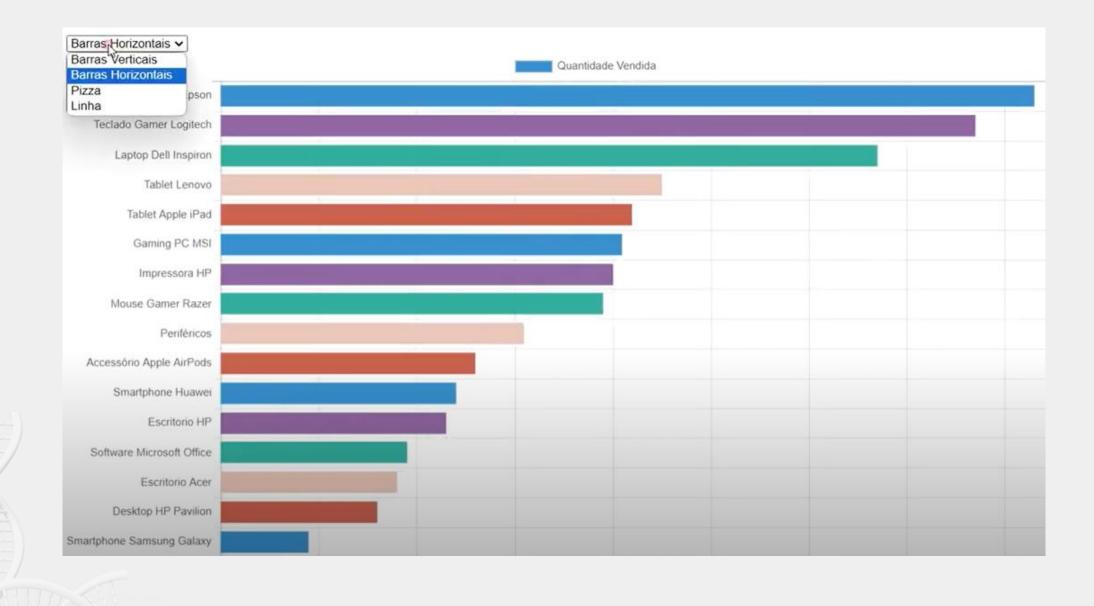












https://lmarena.ai



Embarcadero Conference 2024

Inovação faz parte do nosso DNA!



Quer me ver na
#ECON25?
Acesse o QRCode
e avalie minha palestra!



Carlos Eduardo Paulino

- @DelphiCleanCode
- in <u>linkedindopalestrante</u>
- delphicleancode@gmail.com
- <u>(14) 9 9195 4605</u>