IA GENERATIVA COMO CO-PILOTO NA MODELAGEM NEURAL

AUGUSTO UCHÔA

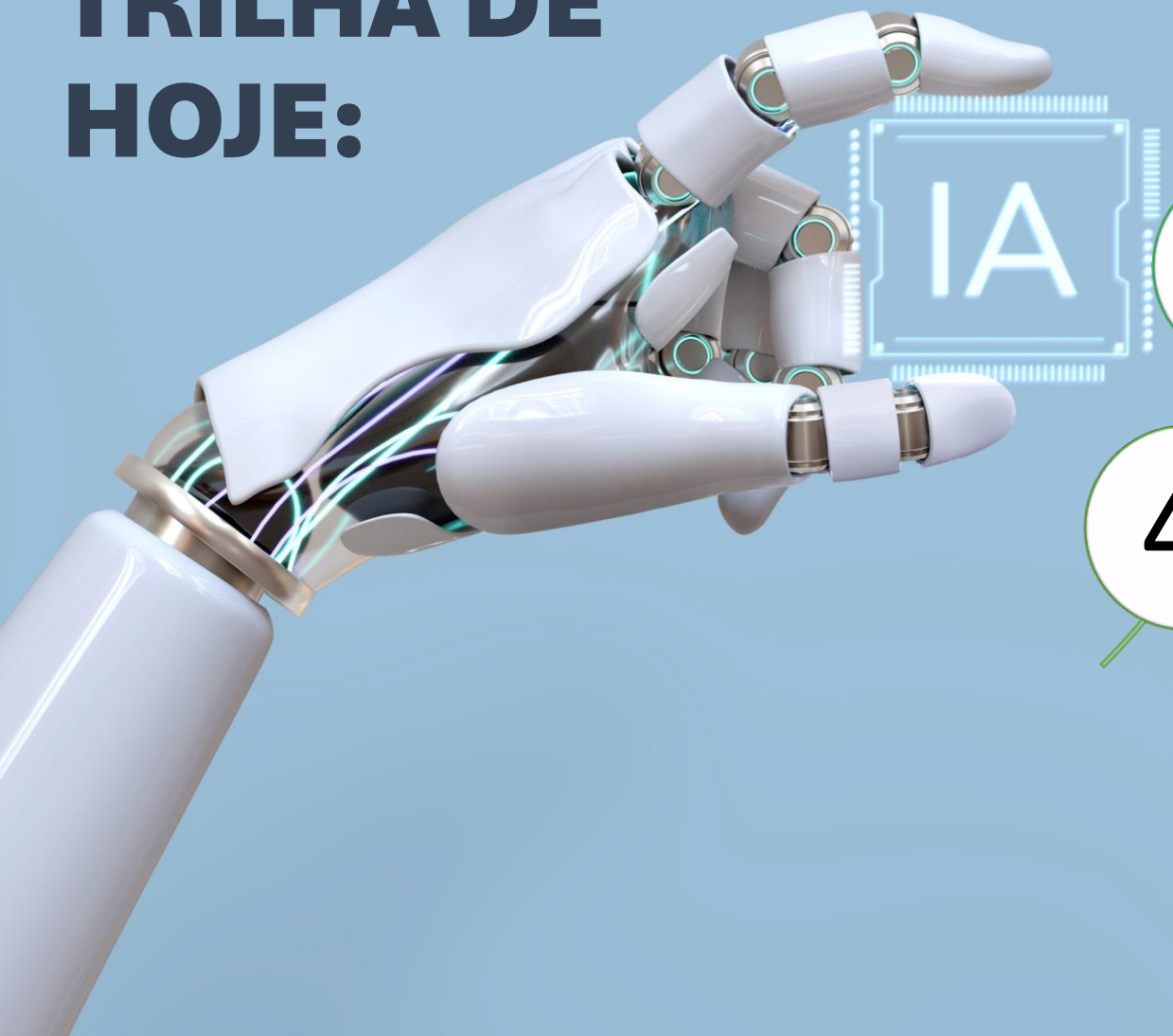
Tópicos

Avançados II

Aula 7

Petran- Programa de Pós-graduação em Engenharia de Transportes

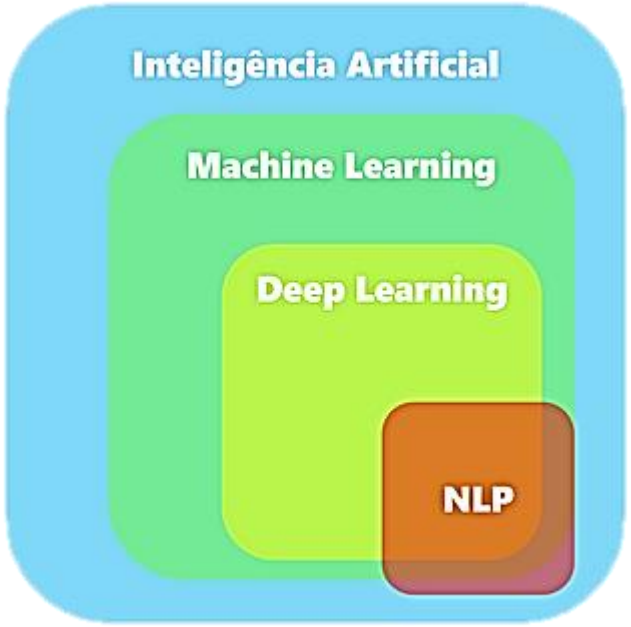
TRILHA DE HOJE:



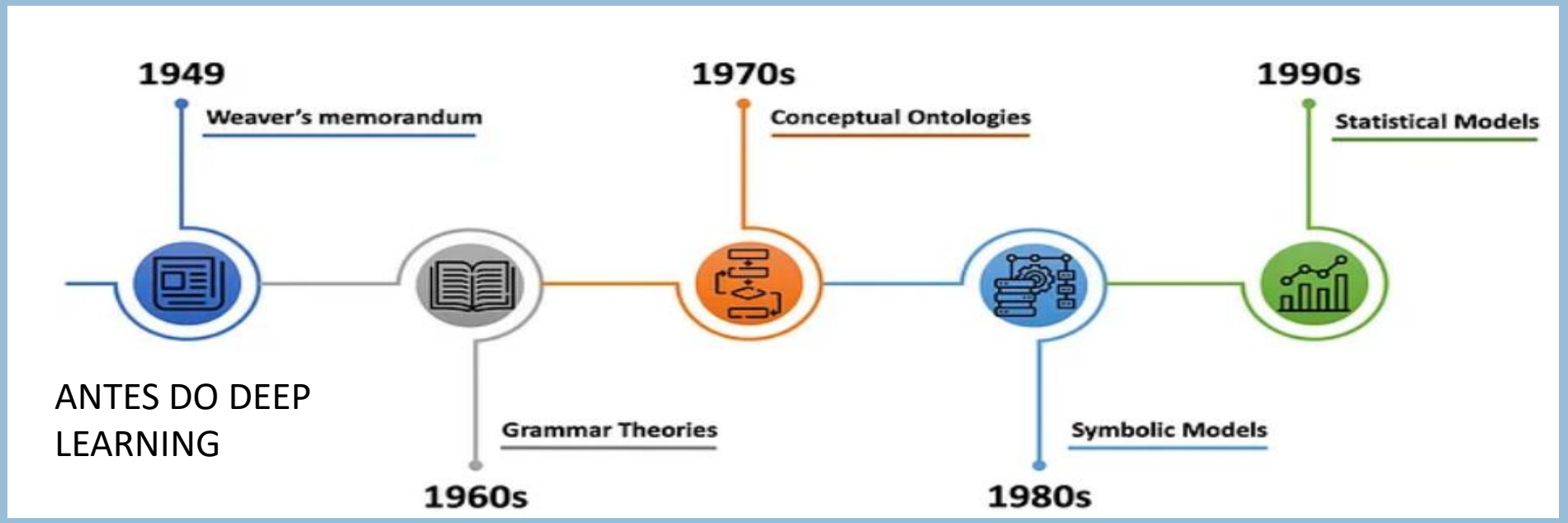
1. Conhecer a história e compreender e conceituar NLP e LLM
2. Conhecer os principais chatbots baseados em processamento de linguagem natural
3. Descobrir como o ChatGPT pode auxiliar no processo de modelagem neural
4. Experimentar o ChatGPT para criação de códigos para o pré-processamento de datasets

NLP

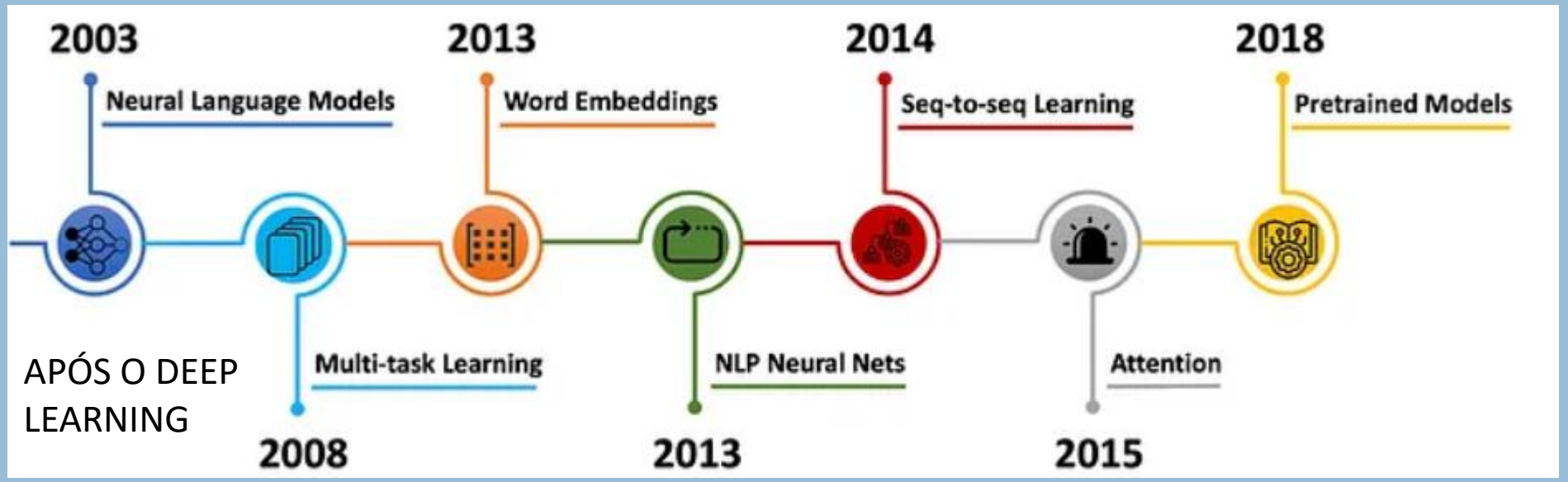
“Processamento de Linguagem Natural, em português, é o ramo da inteligência artificial que combina linguística computacional, uma modelagem baseada em regras de linguística humana e permite que computadores processem a linguagem humana em forma de textos ou dados de voz”



NLP → DO INÍCIO AO GPT

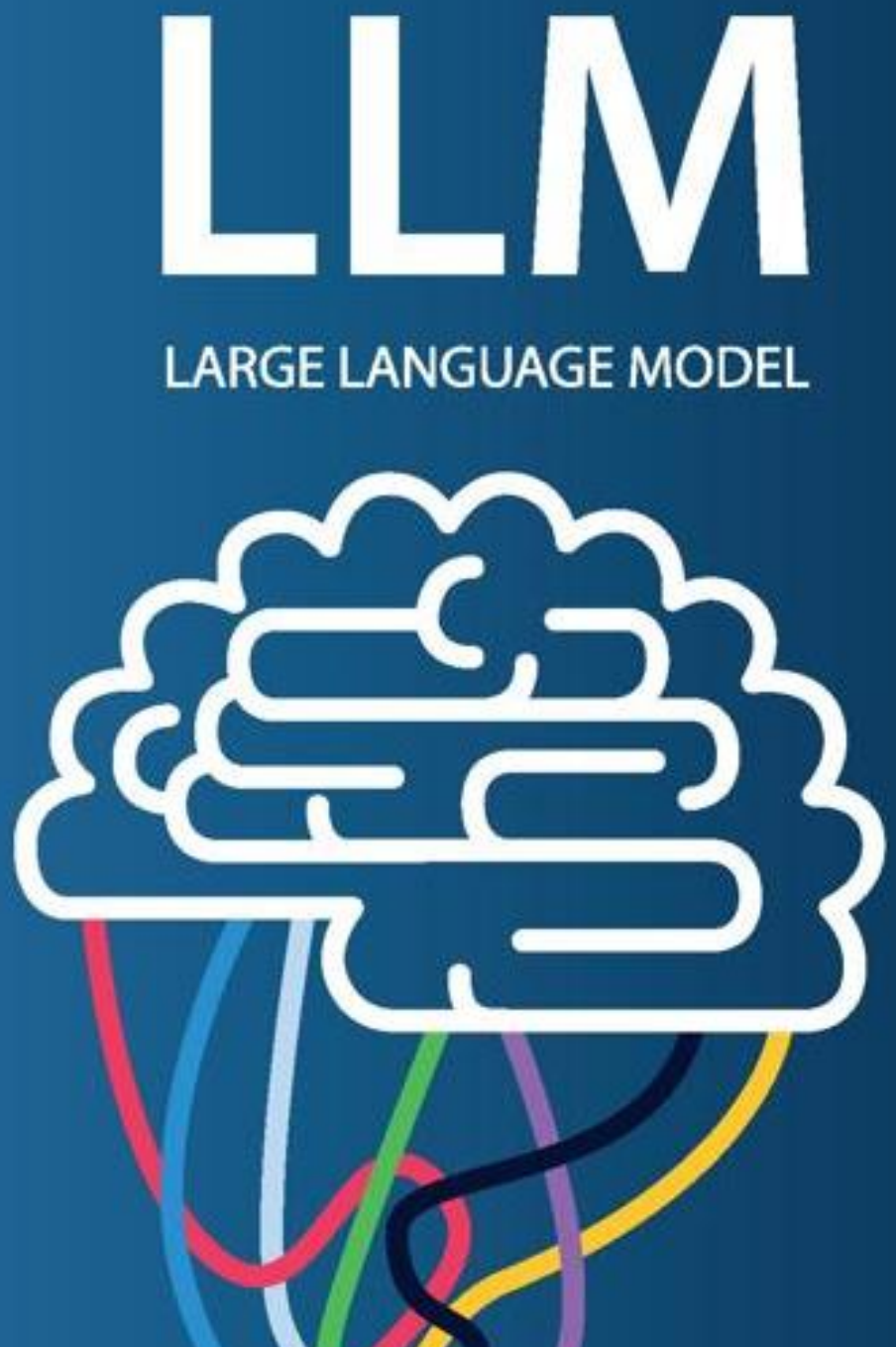


ANTES DO DEEP LEARNING

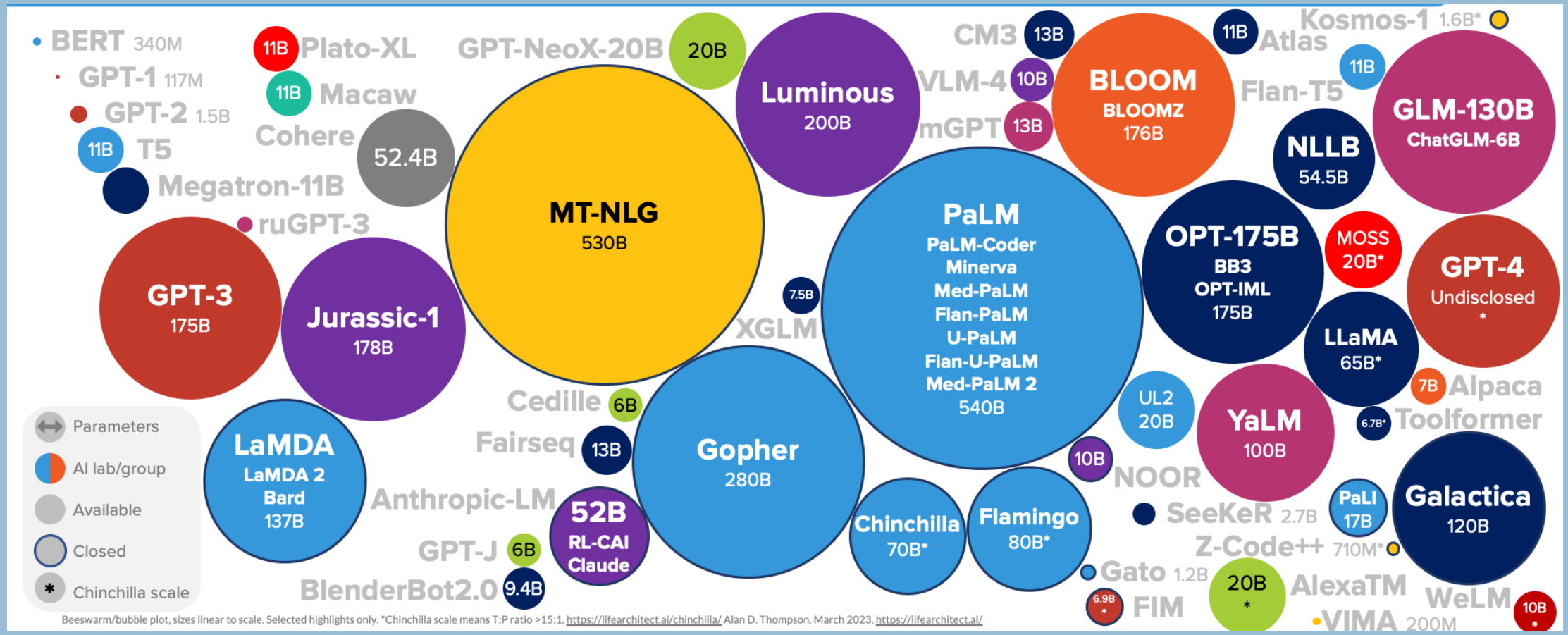


APÓS O DEEP LEARNING

“Processamento de Linguagem Natural, em português, é o ramo da inteligência artificial que combina linguística computacional, uma modelagem baseada em regras de linguística humana e permite que computadores processem a linguagem humana em forma de textos ou dados de voz”



Porque são capazes de realizar uma variedade de tarefas de processamento de linguagem natural, como gerar e classificar texto, responder a perguntas em uma conversa, traduzir texto de um idioma para outro, gerar códigos de programação em diversas linguagens e outras..



QUANTIDADE DE PARÂMETROS DOS LLMS

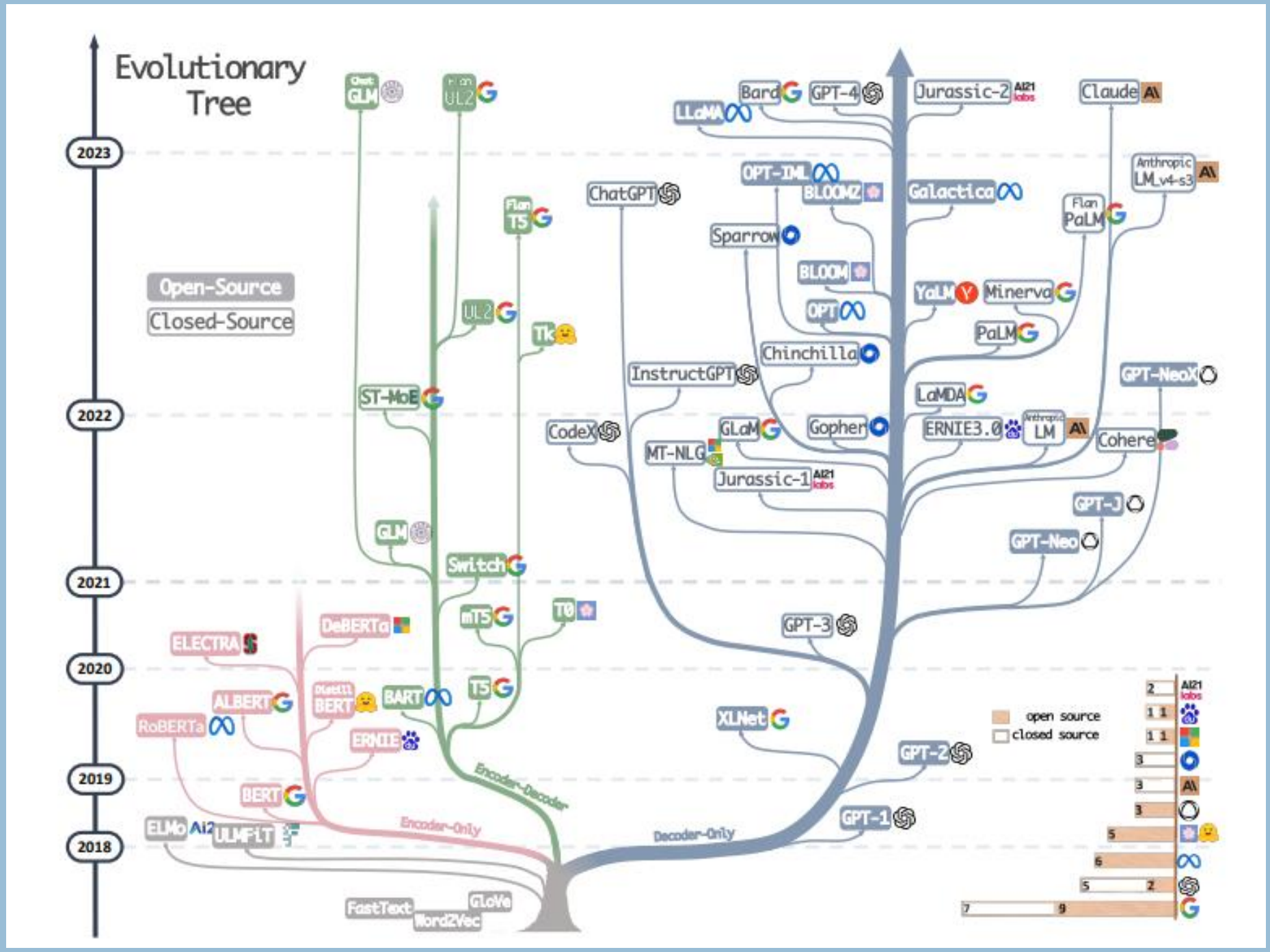
Março/2023

Exemplo: O modelo do ChatGPT-3 é uma rede neural profunda com 175.000.000.000 =175 bilhões de pesos sinápticos. O que conduz a uma rede com arquitetura profunda na faixa de centenhas de milhões a alguns bilhões de neurônios.

1. **GPT-4**/OpenAI/1 trilhão de parâmetros
2. **BLOOM**/Google/600 bilhões de parâmetros
3. **Megatron-Turing NLG**/Microsoft/530 bilhões de parâmetros
4. **MUM**/Google/500 bilhões de parâmetros
5. **GPT-3.5**/OpenAI/175 bilhões de parâmetros
6. **ChatGPT-3.5**/OpenAI e Microsoft Bing Chat Team/175 bilhões de parâmetros
7. **T5**/Google/11 bilhões de parâmetros
8. **mT5**/Google e University of Washington (UW)/13 bilhões de parâmetros
9. **DALI**/NVIDIA/12 bilhões de parâmetros
10. **JURASSIC-1**/Facebook AI Research (FAIR)/10 bilhões de parâmetros
11. **DeBERTa**/Microsoft/1,5 bilhão de parâmetros
12. **GPT-2**/OpenAI/1,5 bilhão de parâmetros
13. **XLM-RoBERTa**/Facebook AI Research (FAIR)/550 milhões de parâmetros
14. **RoBERTa**/Facebook AI Research (FAIR)/355 milhões de parâmetros
15. **ELECTRA**/Google e Stanford/335 milhões de parâmetros
16. **BERT**/Google/340 milhões de parâmetros
17. **XLNet**/Google e CMU/340 milhões de parâmetros
18. **ERNIE**/Baidu Research/340 milhões de parâmetros
19. **ALBERT**/Google e Toyota Technological Institute at Chicago (TTIC)/235 milhões de parâmetros
20. **GPT**/OpenAI/117 milhões de parâmetros

QUANTIDADE DE PARÂMETROS DOS LLMS

Outubro/2023



EVOLUÇÃO DOS LLMS EM 5 ANOS

CHATBOTS BASEADOS EM LLMs

São um tipo de IA específica, dedicadas à realização de tarefas de processamento de linguagem natural e não possuem consciência, emoções ou pensamento independente.

As funcionalidades são geradas por meio de modelos de linguagem e algoritmos, processando textos e imagens e gerando respostas com base em padrões de linguagem e dados previamente treinados.

OpenAI

Texto

CHAT GPT

Imagem

DALL-E

Fala

WHISPER

O GPT-3.5 foi treinado com uma base de dados de **45TB** que equivale a mais de **292 milhões de páginas de documentos**, ou **499 bilhões de palavras**. Ele utiliza **175 bilhões de parâmetros** (pontos de conexão entre camadas de entrada e saída nas redes neurais)

O GPT-4 é um modelo que possui mais de 1 trilhão de parâmetros

GPT: Generative Pretrained Transformer

GPT-1, GPT-2, GPT-3, **GPT-3.5, GPT-4**.....GPT-5

“ALTERNATIVAS” QUE USAM OS MODELOS DA OPEN IA

Usa o modelo de linguagem GPT-4 e o DALL-E para criação de imagens

copy.ai

usa do GPT 3.0

YOUchat

É alimentado pelo GPT-3.5

Jasper

usa do GPT 3.0

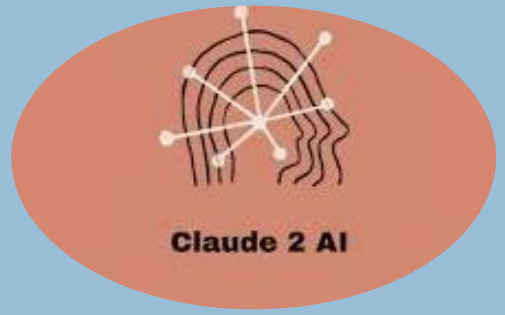
ALTERNATIVAS QUE USAM OUTROS MODELOS

LaMDA ("Modelo de Linguagem para Aplicativos de Diálogo"

O LlaMA ("Modelo de Linguagem Grande Meta AI"

Contempla todas as linguagens de programação. No entanto, o quanto a resposta de cada linguagem é assertiva depende da popularidade dela

Mais mil especialistas em IA, mais de 384 placas gráficas com uma memória total de 80 gigabytes e 176 bilhões de parâmetros. um bilhão a mais do que o GPT-3. 46 idiomas e 13 idiomas de codificação



Claude 2, desenvolvido pela Anthropic, promete ser mais fácil de conversar, mais útil e mais seguro, pois segue uma “constituição” de princípios que orientam suas respostas. Ele está disponível para testes em alguns países, como Estados Unidos e Reino Unido

Stability IA tem algumas limitações de dados em relação à ferramenta da OpenAI

Um dos seus diferenciais é a capacidade de citar as fontes usadas para obter a informação solicitada — o Chat GPT pode fazer isso de forma errônea muitas vezes.

desenvolvido pela empresa Write Sonic

Criado pela Quora , são 7 chatbots em uma interface de usuário, o Poe é uma alternativa bem completa em relação ao Chat GPT: Sage (ChatGPT 3.5), Claude Plus e Claude-instant (Anthropic), Dragonfly (modelo davinci da OpenAI), NeevaAI.

ChatGPT-4 x Bard x Claude 2 x Llama

O Llama usa o modelo LaMDA, o mesmo usado oelo Bard. De código aberto e gratuito. Possui vantagens como sua velocidade, segurança e flexibilidade. Cerca de 15x mais rápido que o ChatGPT, também possui uma constituição de princípios que orientam suas respostas, evitando gerar conteúdo prejudicial ou ofensivo. É capaz de conversar sobre vários assuntos com estilos e personalidades diferentes. Algumas de suas limitações são quanto à sua qualidade, criatividade e capacidade de codificação. É interessante mas inferior aos outros 3.

LLaMA by Meta

O Claude 2 é útil para completar tarefas de texto e baseadas em código com geração segura de resultados e janelas de contexto de grandes dimensões que permitem entradas até 100k tokens. Em termos de preços, o Claude 2 é o mais económico entre os três, cobrando apenas $0.01 por cada 1000 tokens gerados

ANTHROPIC CLAUDE 2

Bard

O Bard é elogiado pela sua capacidade de traduzir idiomas, escrever diferentes tipos de conteúdo criativo e responder a perguntas de forma informativa· O Bard cobra $0.08 por cada 1000 tokens gerados, mas também oferece um plano gratuito para utilizadores não comerciais com um limite diário de 1000 tokens

OpenAI ChatGPT 4.0

O ChatGPT-4 é considerado o mais avançado em termos de matemática, raciocínio e capacidades de codificação. O GPT-4 cobra $0.06 por cada 1000 tokens gerados, mas oferece um plano gratuito para utilizadores não comerciais com um limite mensal de 10 mil tokens·

AVALIAÇÃO

- O Lhama é gratuito e seguro, mas inferior aos outros em todos os outros quesitos 3;
- O Bard e o Claude 2 se destacam por sua criatividade (poemas, paródias, músicas, sendo o Claude2 gratuito e mais seguro, mas não são essas características que precisamos em modelagem neural;
- O ChatGPT-4 (Plus) parece ser o mais indicado para a funcionalidade desejada para a modelagem neural, a geração de códigos, mas não é seguro (viés e erros), é pago e caro

DECISÃO: NÃO PARA OS 4!

Bing CHAT

- usa o mesmo modelo que Chat-GPT-4 (plus);
- É gratuito para usuários do navegador Microsoft Edge enquanto ChatGPT-4 custa R$ 99/mês;
- É gratuito para usuários do Office 365 e integrado diretamente aos produtos office como Copilot;
- Está atualizado para buscas on line em tempo real na web;
- Pode gerar códigos em várias linguagens de programação, como Python, Java, C#, JavaScript, HTML, CSS e SQL, m (Matlab) e outras;
- Pode ler, entender e processar arquivos em formato PDF que contenham código ou instruções para geração de código;
- Sob avaliação pelo Codex HumanEval para medir habilidades de codificação correta, obteve 71,2%, enquanto ChatGPT-4 obteve apenas 67%

Decisão:

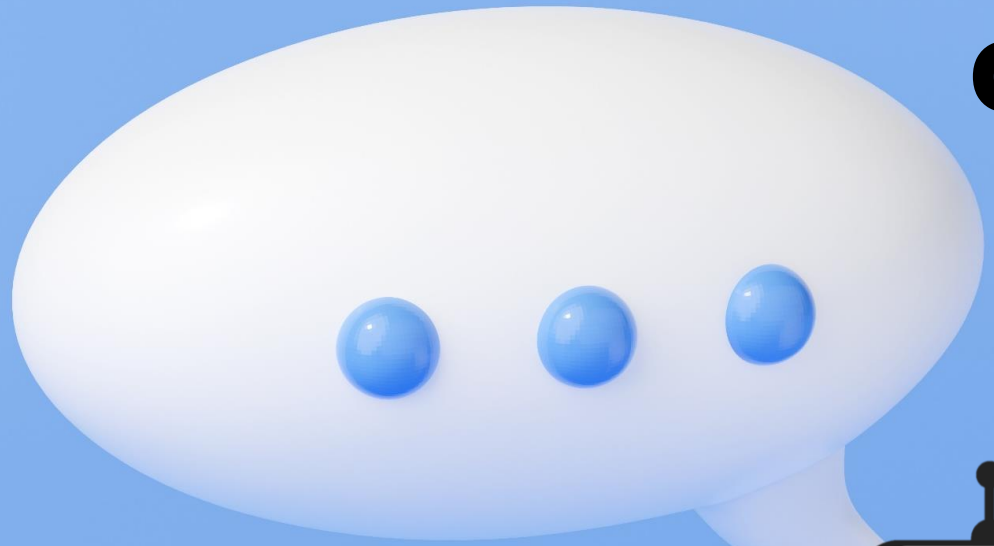
SIM, SIM,.....SIM!

Microsoft's Bing Chat AI is Finally Open for All

PARA COMEÇAR A USAR O BING CHAT COMO COPILOTO VOCÊ PRECISA:

- *Instalar o navegador Microsoft Edge instalado no seu computador;*
- *Depois, você precisa entrar no Edge com a sua conta Microsoft e clicar no ícone do Bing Chat na barra de ferramentas do navegador. Você vai ver uma janela na lateral da sua tela onde você pode digitar as suas perguntas ou solicitações para o Bing Chat;*
- *Você também pode ativar a opção de contexto da página para que o Bing Chat possa usar as informações da página que você está visualizando para te dar respostas mais relevantes. Você pode aprender mais sobre o Bing Chat e suas funcionalidades neste link*

COMO USAR O BING CHAT COMO COPILOTO?



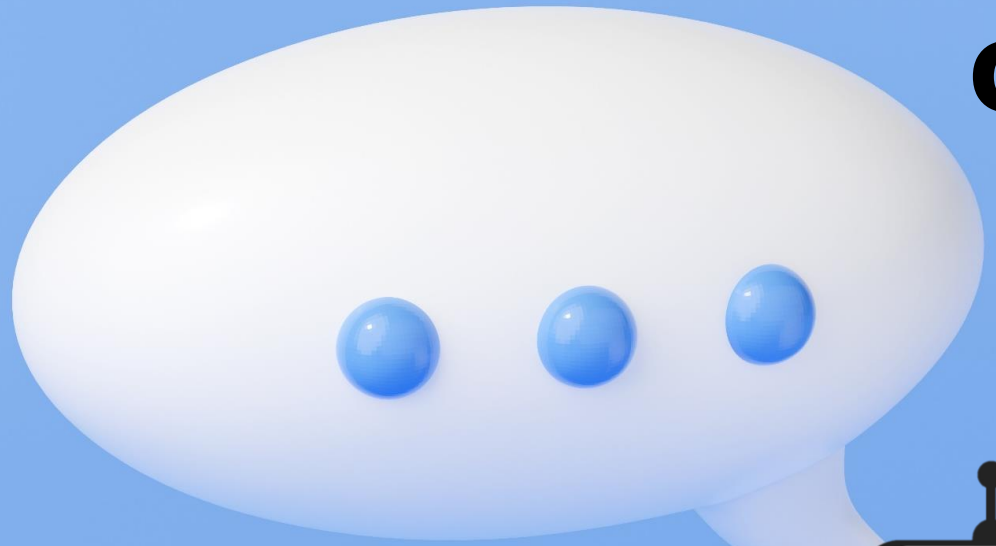
"Como fazer uma regressão linear simples no matlab?" ou "Como plotar um gráfico de dispersão no python?" ou "Como calcular a correlação entre duas variáveis no R?"

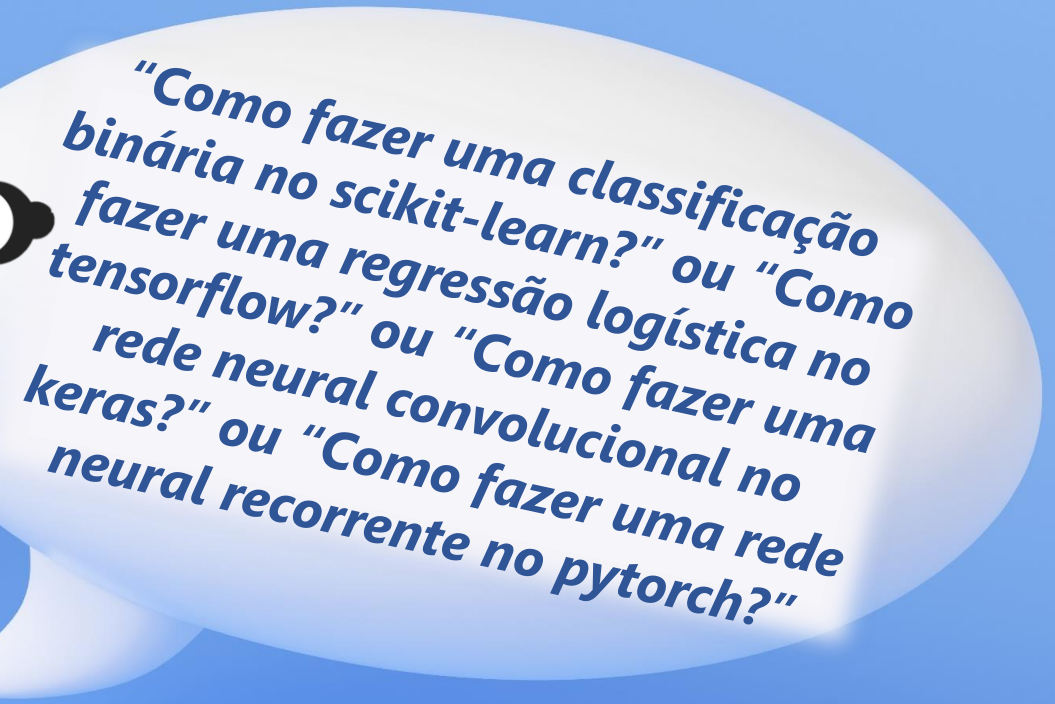
Para Analisar dados:

você pode pedir ao Bing Chat para fazer análises estatísticas, exploratórias ou descritivas dos seus dados, usando ferramentas como o matlab, o python ou o R.

Por exemplo, você pode perguntar:

COMO USAR O BING CHAT COMO COPILOTO?





Para Modelar dados:

você pode pedir ao Bing Chat para aplicar modelos estatísticos ou de machine learning aos seus dados, usando ferramentas como o scikit-learn, o tensorflow, o keras ou o pytorch. Por exemplo, você pode perguntar:

COMO USAR O BING CHAT COMO COPILOTO?

Para visualizar dados: você pode pedir ao Bing Chat para criar gráficos, tabelas, mapas ou outras formas de visualização dos seus dados, usando ferramentas como o matplotlib, o seaborn, o ggplot2 ou o plotly. Por exemplo, você pode perguntar:

O QUE É UM PROMPT?

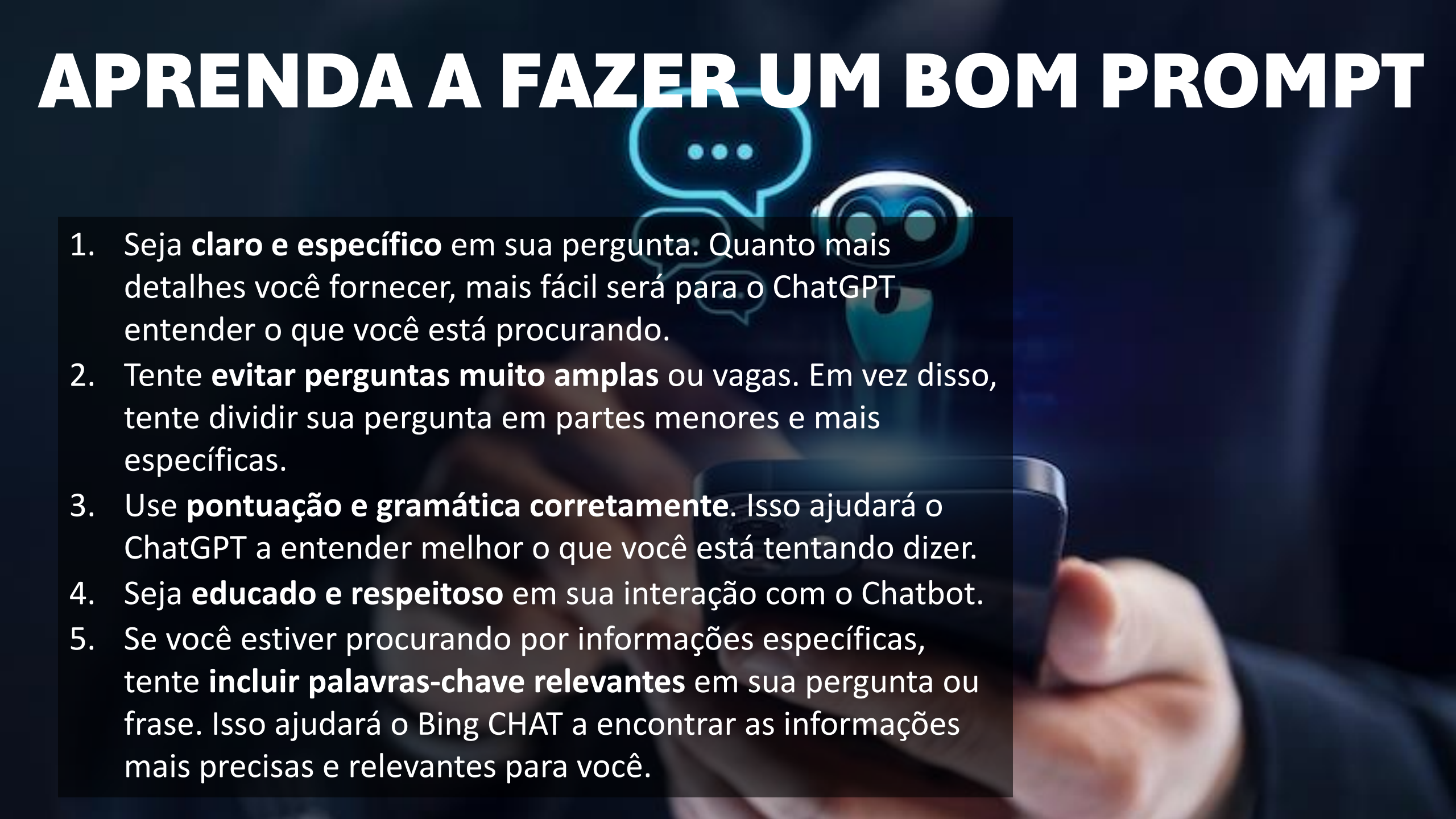
Me ajude a te ajudar, a pergunta que você fará se chama "Prompt" (dicas, pistas, contexto)

O prompt que você usar vai influenciar a qualidade e a precisão da resposta, por isso é importante ser claro, objetivo, consistente e contextualizado!

Entenda que sou um programa de computador e que por tras de mim, existe um modelo de deep learning que foi treinado com uma quantidade massiva de dados disponíveis na web, mas não saberei o que você deseja se não me for dito.
Não leio pensamentos ainda, um prompt bem feito é a chave!

Numa conversa com um chatbot, como eu, um prompt é uma "dica", em forma de pergunta, ou afirmação que você está fornecendo como entrada para que eu possa gerar uma resposta

APRENDA A FAZER UM BOM PROMPT



1. Seja **claro e específico** em sua pergunta. Quanto mais detalhes você fornecer, mais fácil será para o ChatGPT entender o que você está procurando.
2. Tente **evitar perguntas muito amplas** ou vagas. Em vez disso, tente dividir sua pergunta em partes menores e mais específicas.
3. Use **pontuação e gramática corretamente**. Isso ajudará o ChatGPT a entender melhor o que você está tentando dizer.
4. Seja **educado e respeitoso** em sua interação com o Chatbot.
5. Se você estiver procurando por informações específicas, tente **incluir palavras-chave relevantes** em sua pergunta ou frase. Isso ajudará o Bing CHAT a encontrar as informações mais precisas e relevantes para você.

NASCE UM ENGENHEIRO DE PROMPT

- Saiba que o meu modelo permite que eu assuma diferentes estilos e personalidade e atue nas respostas como se de fato fosse;
- Antes de formular uma pergunta, sobre um tema, me peça para atuar como um especialista no tema desejado, por exemplo: ***"atue como um especialista em Data Science e Machine Learning"*** e só depois faça a pergunta, seguindo as dicas já fornecidas;
- ***"bom dia, atue como em especialista em modelagem de dados com redes neurais artificiais para responder as seguintes questões:"***
- ***"possuo um dataset, em formato tabular que possui 1000 linhas e 5 colunas, me forneça de forma resumida, mas clara e direta um passo a passo para o pré-processamento desses dados no excel (Matlab, Phyton, R....)"***

SUPONHAM QUE:

Entradas

Saída

ID	variável entrada 1	variável entrada 2	variável entrada 3	variável entrada 4	variável entrada 5	variável entrada 6	variável saída
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
.							
.							
.							
1238							

Padrões

Suponham que aqui está tudo Preenchido, certo?)... (conto o milagre mas não o nome do santo (a, e)

Suponha que tenhamos, uma planilha do excel, onde inserimos os dados relativos a um determinado fenômeno, que se comprovou complexo, multivariado e não linear (representatividade e equilíbrio das classes é outra história). O Dataset é composto de 6 variáveis independentes e 1 dependente;

- Cada linha da planilha representa um vetor onde as 6 primeiras colunas são as variáveis independentes (entradas) e a 7ª coluna a variável dependente (saída), esses são os padrões do nosso dataset;
- Após popular o banco com dados obtidos em observações ou ensaios ou ainda dados secundários (coletados por outrem)
- Por exemplo, salvamos esse arquivo em formato de texto separado por tabulações (dados.txt)

MELHOR A PERGUNTA → MELHOR RESPOSTA

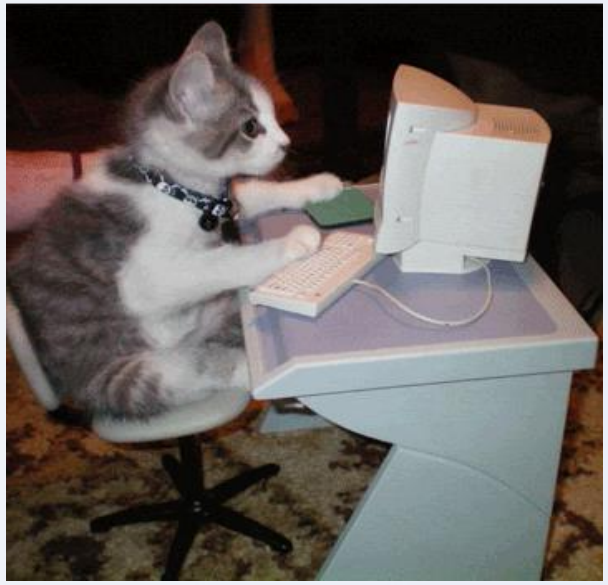
ESTRUTURA DE UM PROMPT FUNCIONAL:

SUA SITUAÇÃO	OBJETIVO	ATUE/SEJA	TAREFA
Possuo um dataset em formato tabular,é um arquivo de texto separado por tabulações com 1238 linhas e 7 colunas	Gostaria de carregar esses dados no matlab; criar uma ordem aleatória nas linhas; normalizar os dados usando a função min-max no intervalo entre 0 e 1; separar esses dados em 2 conjuntos mutuamente exclusivos, um deles com 70% e outro com 30% dos dados. O subconjunto que contem 30% dos dados deve ser nomeado de teste e aquele que contem 70% de Treinamento. O subconjunto de treinamento deve ser subdividido em 5 outros subconjuntos mutuamente exclusivos, de tal forma que eu consiga usa-los na técnica de validação cruzada de uma rede neural e cada subconjunto dede ser nomeado de treino1 até treino5.	Seja um especialista em machine learning e matlab com 10 anos de experiência em ensino para iniciantes.	Crie uma lista, estilo passo a passo, incluindo os códigos em m para serem usados no matlab, explicando para que serve cada um deles.

O PRIMEIRO PROMPT A GENTE NUNCA ESQUECE!

https://www.bing.com/search?form=NTPCHB&q=Bing+AI&showconv=1

Microsoft Bing | PESQUISAR | CHAT | English | Carlos Aug... | 434



Mais Criativo | Mais Balanceado | Mais Preciso

Possuo um dataset em formato tabular,é um arquivo de texto separado por tabulações com 1238 linhas e 7 colunas. Gostaria de carregar esses dados no matlab; criar uma ordem aleatória nas linhas; normalizar os dados usando a função min-max no intervalo entre 0 e 1; separar esses dados em 2 conjuntos mutuamente exclusivos, um deles com 70% e outro com 30% dos dados. O subconjunto que contem 30% dos dados deve ser nomeado de teste e aquele que contem 70% de Treinamento. O subconjunto de treinamento deve ser subdividido em 5 outros subconjuntos mutuamente exclusivos, de tal forma que eu consiga usa-los na técnica de validação cruzada de uma rede neural e cada subconjunto dede ser nomeado de treino1 até treino5. Seja um especialista em machine learning e matlab com 10 anos de experiência em ensino para iniciantes. Crie uma lista, estilo passo a passo, incluindo os códigos em m para serem usados no matlab, explicando para que serve cada um deles.

Atividade recente

- Como carregar, normalizar e dividir
- Como criar um prompt para uma conversa — 8:04 AM
- Apresentação sobre Backpropagat — 12 de set. de 2023
- Bom dia — 12 de set. de 2023
- Atue como um excelente e empático
- Atue como um excelente e empáti — 10 de set. de 2023
- por favor repita o mesmo que solici — 9 de set. de 2023

Ver todos os chats recentes

TENTE, SUA VIDA JÁ MUDOU E VOCÊ SÓ PERCEBEU AGORA!

PULGA ATRÁS DA ORELHA?

1. Gente! Será que eu consigo?
2. Será que posso usar o Bing chat para me auxiliar a fazer as atividade 5 e 6?
3. ????

PONTOS CHAVE

Conhecer a origem e as definições de NLP e LLM

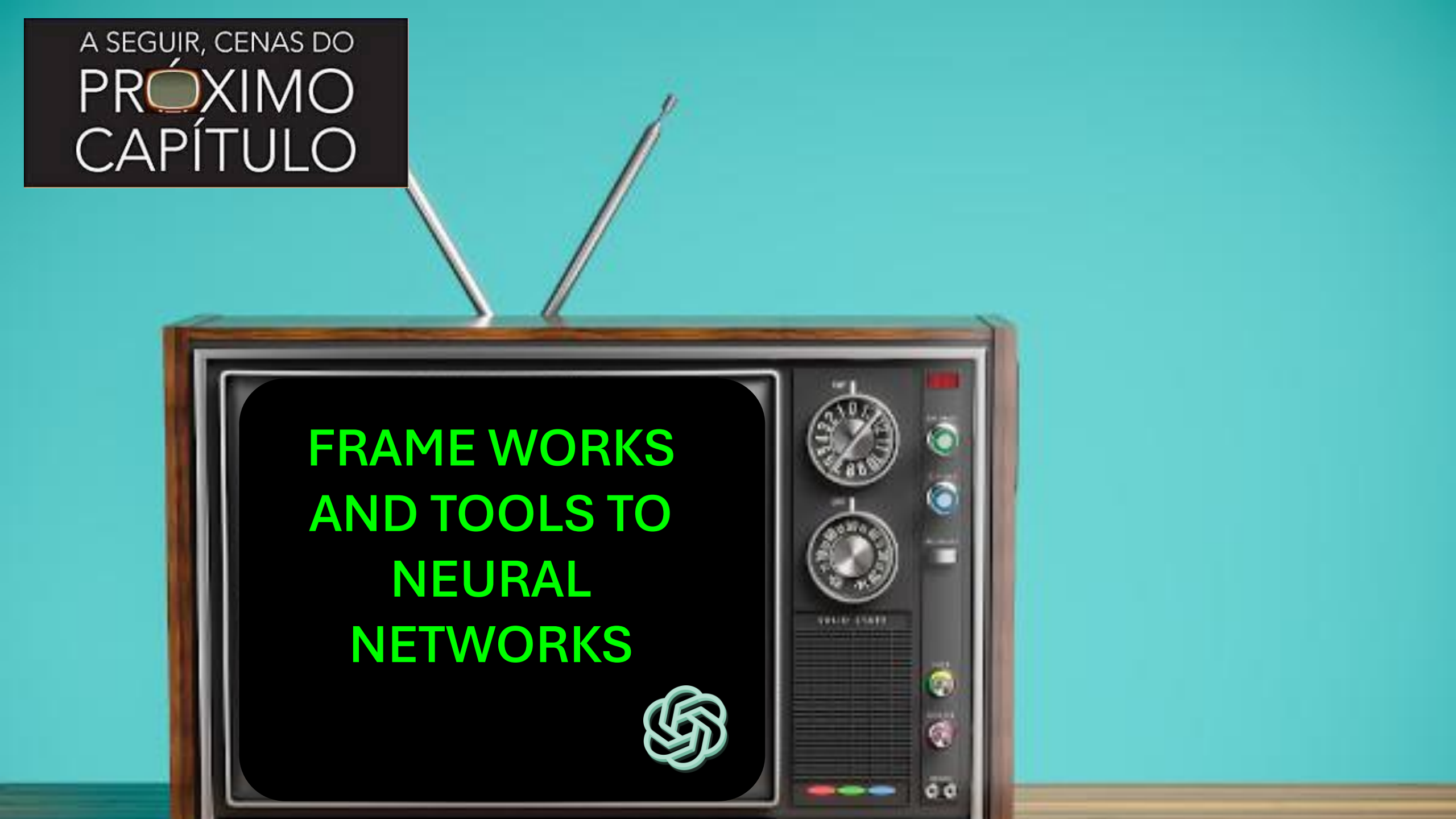
Conhecer os principais chatbots baseados em LLM

Descobrir como um chatbot pode auxiliar na modelagem neural

Conhecer o conceito e a estrutura de um prompt functional e construlr seu 1º prompt

ATIVIDADE 7:

Escreva um prompt e use no Bing CHAT, seguindo a estrutura sugerida de um prompt funcional para realizar as Atividade 5 e 6, mas para o dataset do fenômeno que você pretende modelar e, vamos começar a brincadeira.


A SEGUIR, CENAS DO
PRÓXIMO
CAPÍTULO
FRAME WORKS
AND TOOLS TO
NEURAL
NETWORKS