Fundamentos do Processamento de Linguagem Natural

Flávio Belizário da Silva Mota
Isabela Neves Drummond

Créditos da imagem: MIT Technology Review

O que é a linguagem natural?

Linguagem que usamos para nos expressar, baseada em um conjunto de protocolos acordados mutuamente envolvendo palavras e sons que usamos para nos comunicarmos

A CHEGADA. Direção: Denis Villeneuve. Produção de Paramount Pictures. Canadá: Sony Pictures, 2016. (116 min.)

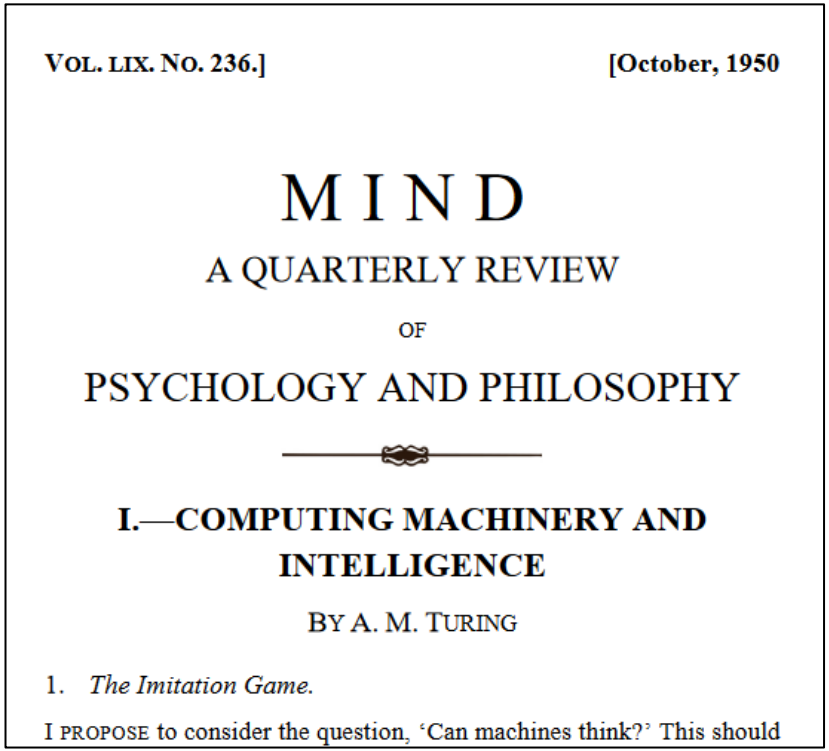
Processamento de Linguagem Natural (PLN)

"- Abra as portas da nave, HAL.

- Sinto muito, Dave, mas temo não ser capaz de abri-las."

2001: UMA ODISSEIA NO ESPAÇO. Direção: Stanley Kubrick. Produção de Metro-Goldwyn-Mayer. Estados Unidos: Metro-Goldwyn-Mayer, 1968. (148 min.)

O teste de Turing



VOL. LIX. NO. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

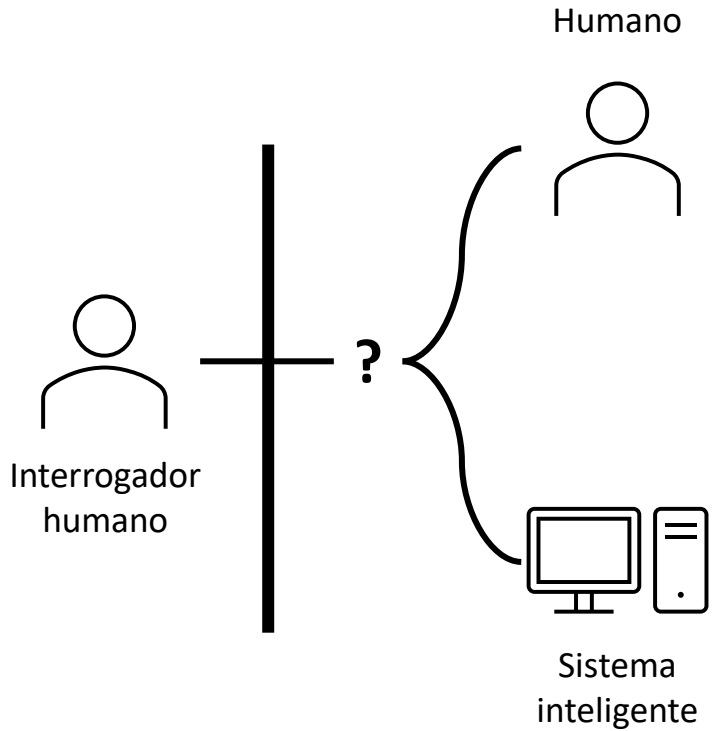
PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

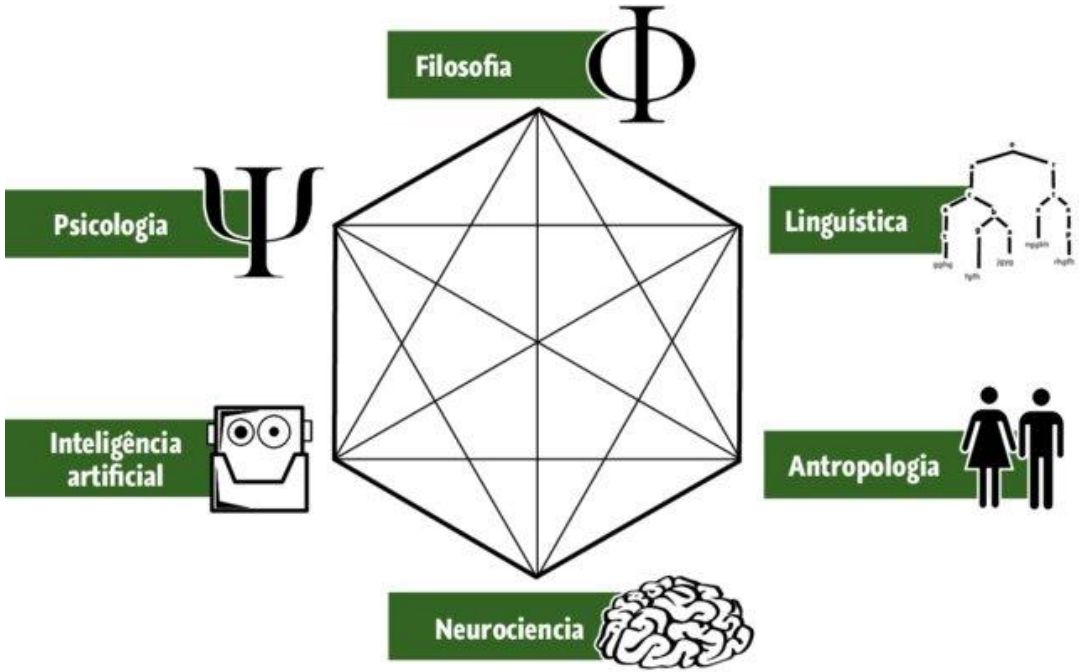
1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should



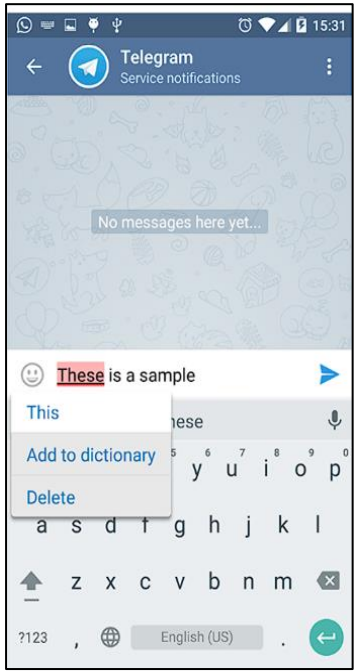
Computing Machinery and Intelligence

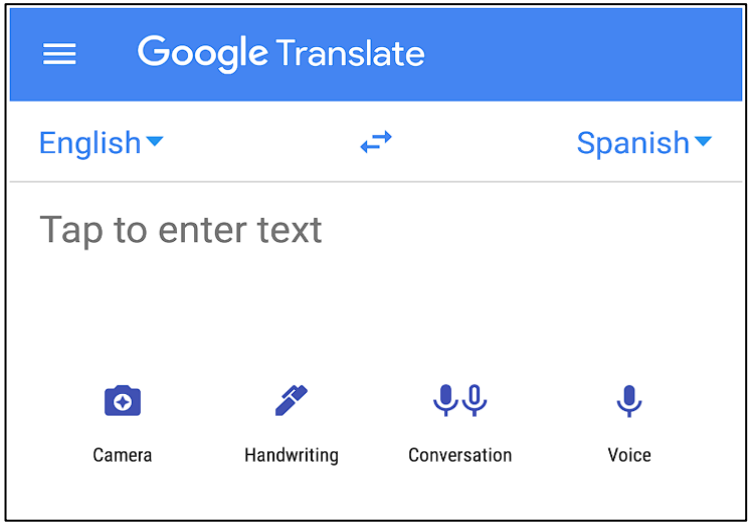
Uma ciência cognitiva

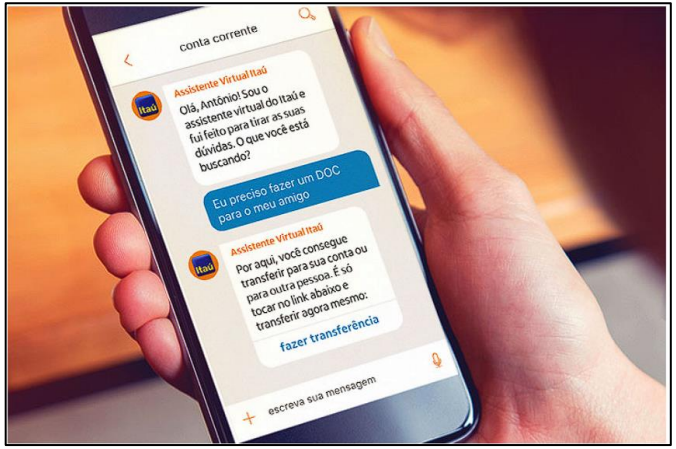


FERREIRA, M.; LOPES, M. Linguística Computacional.
1. ed. Contexto, 2020.

Aplicações


15:31
Telegram
Service notifications
No messages here yet...
These is a sample
This
Add to dictionary
Delete
English (US)


Google Translate
English
Spanish
Tap to enter text
Camera
Handwriting
Conversation
Voice


conta corrente
Assistente Virtual Itaú
Olá, Antônio! Sou o assistente virtual do Itaú e fui feito para tirar as suas dúvidas. O que você está buscando?
Eu preciso fazer um DOC para o meu amigo
Assistente Virtual Itaú
Por aqui, você consegue transferir para sua conta ou para outra pessoa. É só tocar no link abaixo e transferir agora mesmo:
fazer transferência
escreva sua mensagem


"Alexa, quando é a minha próxima reunião?"

Aplicações

ChatGPT

INTELIGÊNCIA ARTIFICIAL

Por que a GPT-3 é o melhor e o pior da IA atualmente

A linguagem de Inteligência Artificial da Open AI impressionou o público com seu aparente domínio do inglês — mas será tudo uma ilusão?

by **MIT Technology Review** Abril 1, 2021

https://mittechreview.com.br/por-que-a-gpt-3-e-o-melhor-e-o-pior-da-ia-atualmente/

Aplicações

HUMANOS E TECNOLOGIA

IA e ChatGPT são criações inevitáveis do nosso tempo, assim como os benefícios da sua democratização

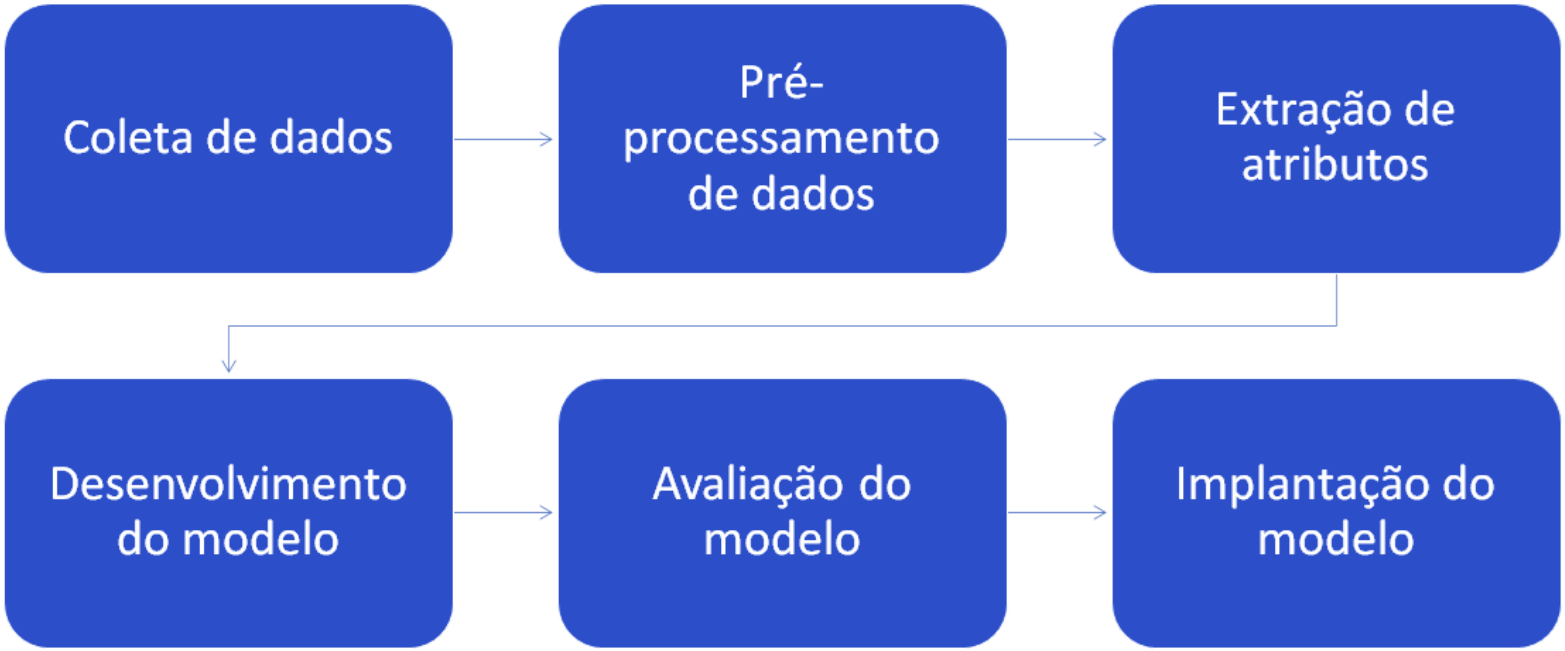
Notícias sobre IA e chats costumam acender discussões sobre o futuro do emprego, assim como aconteceu diversas vezes na sociedade: da invenção da lâmpada elétrica à revolução industrial. IA é a revolução do nosso tempo, e seus benefícios para a sociedade podem ser muito maiores do que os riscos apontados por alarmistas.

by **Fernando Teixeira** Junho 22, 2023

Sabendo que a IA e suas formas generativas são a inevitável revolução do nosso tempo, precisamos focar em como extrair valor, incluir pessoas, criar profissões e, assim, beneficiar a sociedade. O mundo está cheio de problemas para resolver e não faltarão algoritmos para ajudar.

https://mittechreview.com.br/ia-e-chatgpt-sao-criacoes-inevitaveis-do-nosso-tempo-assim-como-os-beneficios-da-sua-democratizacao/

Fases de um projeto de PLN

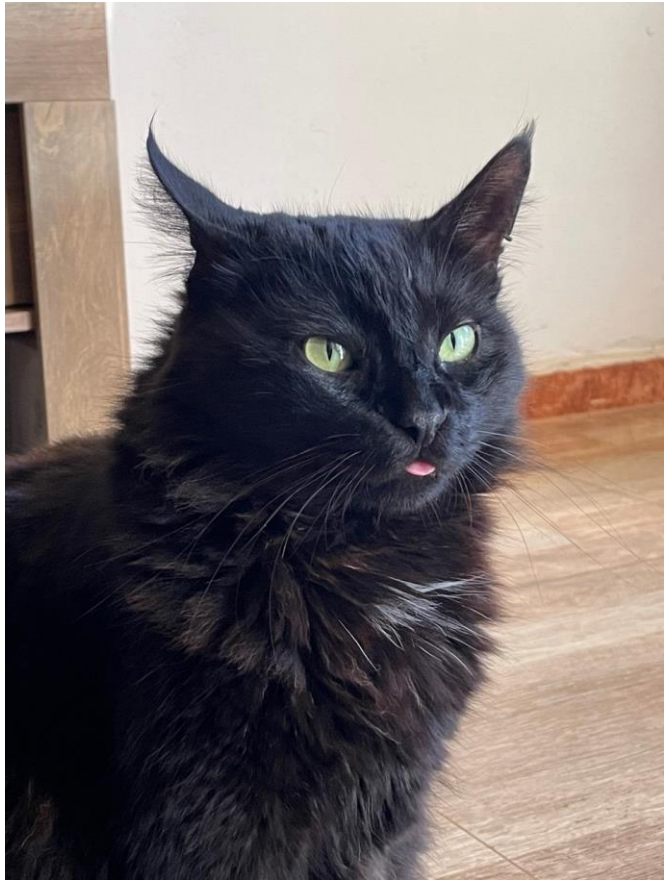


Prática - Google Colab

Introdução ao Processamento de Linguagem Natural

Extração de características

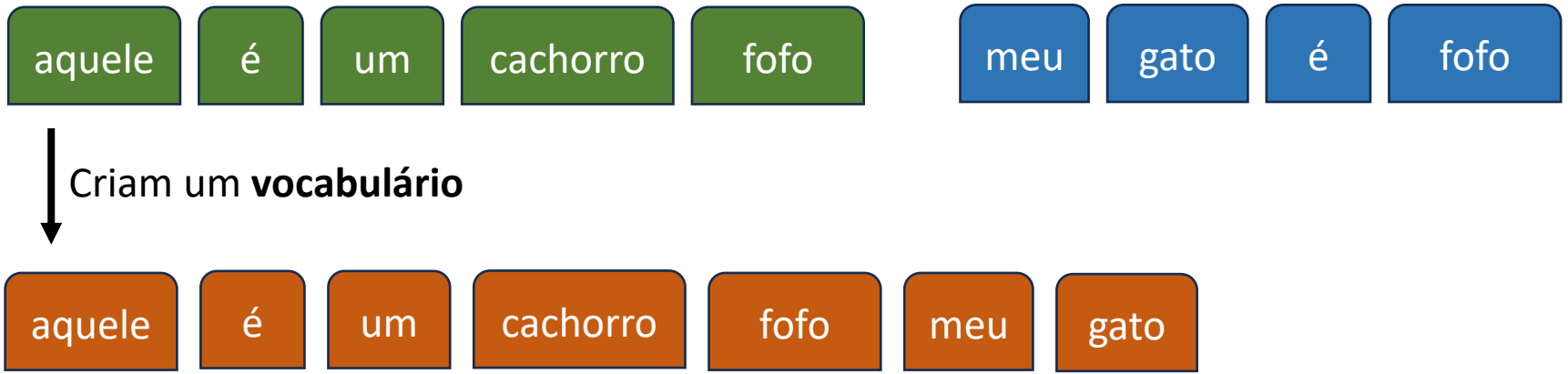
Bag of Words
Tem como objetivo construir uma matriz onde cada coluna representa uma palavra do vocabulário de um documento e cada linha um texto desse documento



Extração de características

Bag of Words

Textos **tokenizados**



Extração de características

Bag of Words

Extração de características

TF-IDF - Term Frequency - Inverse Document Frequency

- **TF = Frequência do Termo:**
  - Se um documento tem 100 palavras e a palavra "cachorro" aparece 5 vezes, então o TF para "cachorro" é 5/100 = 0.05.
- **IDF = Frequência Inversa em Documentos**
  - Se há um total de 1.000.000 de documentos e a palavra "cachorro" aparece em 1.000 documentos, então o IDF para "cachorro" é log(1.000.000 / 1.000) = log(1000) ≈ 3.
- **TF-IDF = TF x IDF**
  - TF-IDF para a palavra "cachorro" no exemplo acima seria

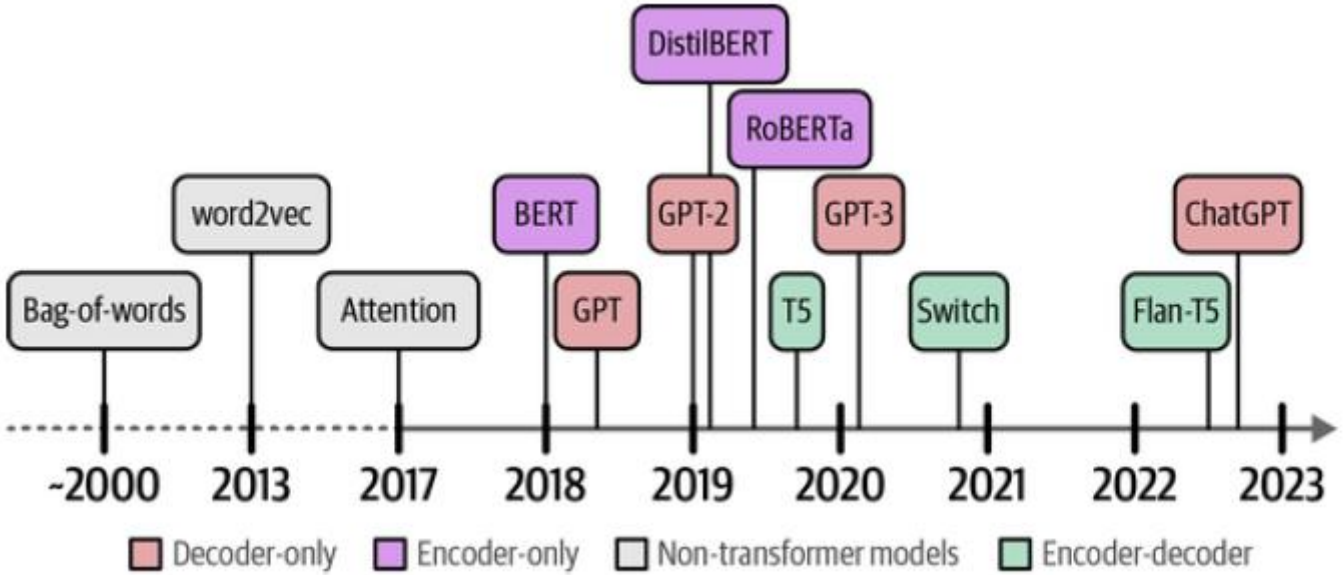
    0.05 (TF) * 3 (IDF) ≈ 0.15

Extração de características

TF-IDF - Term Frequency - Inverse Document Frequency

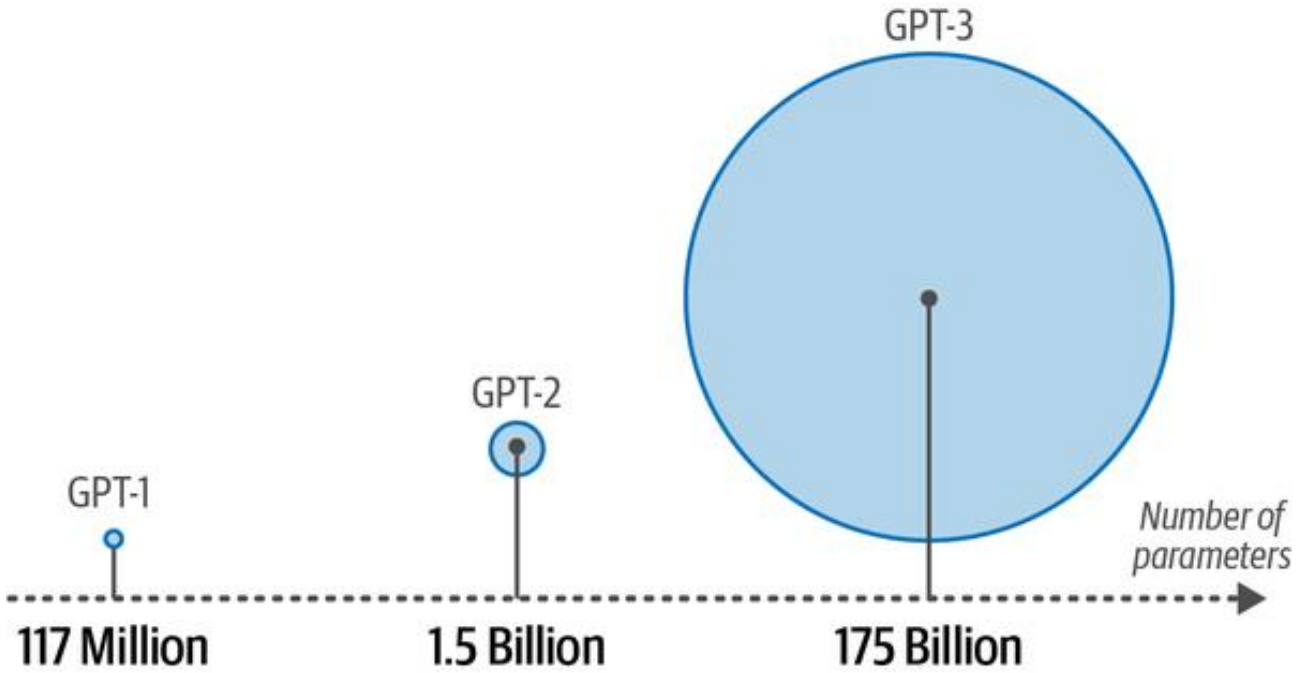
	aquele	é	um	cachorro	fofo	meu	gato
Aquele é um cachorro fofo	0,5	0,3	0,5	0,5	0,3	0	0
Meu gato é fofo	0	0,4	0	0	0,4	0,6	0,6

E os LLMs (Large Language Models)?



ALAMMAR, J.; GROOTENDORST, M. **Hands-On Large Language Models**. O'Reilly, 2024.

E os LLMs (Large Language Models)?



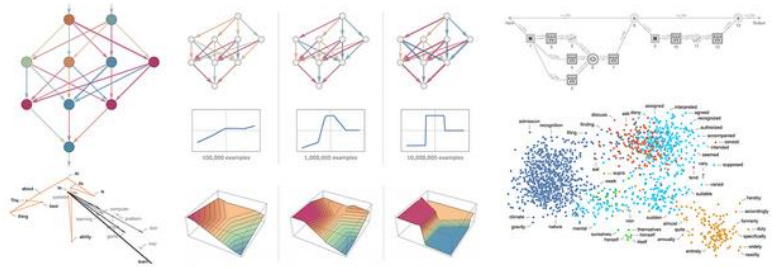
ALAMMAR, J.; GROOTENDORST, M. **Hands-On Large Language Models**. O'Reilly, 2024.

E os LLMs (Large Language Models)?

Stephen Wolfram

What Is ChatGPT Doing … and Why Does It Work?

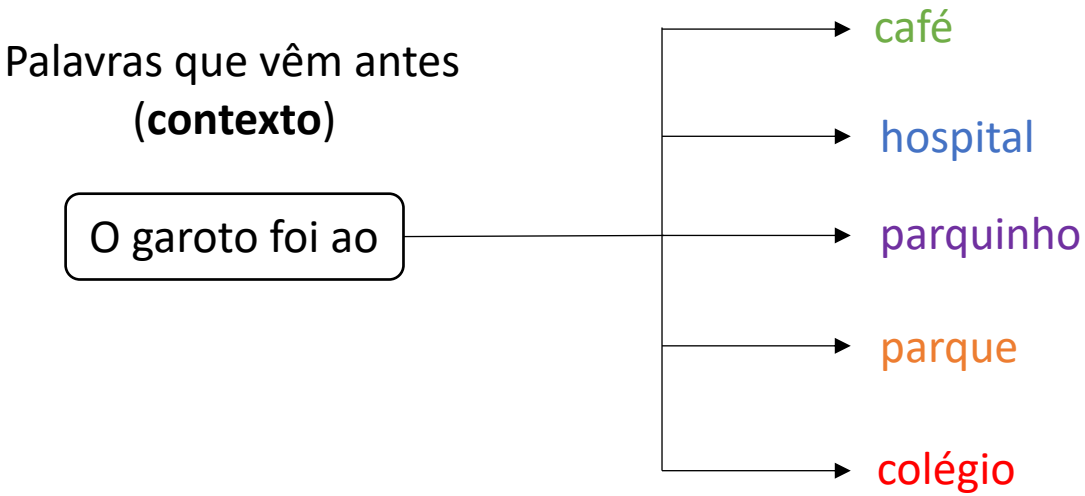
February 14, 2023



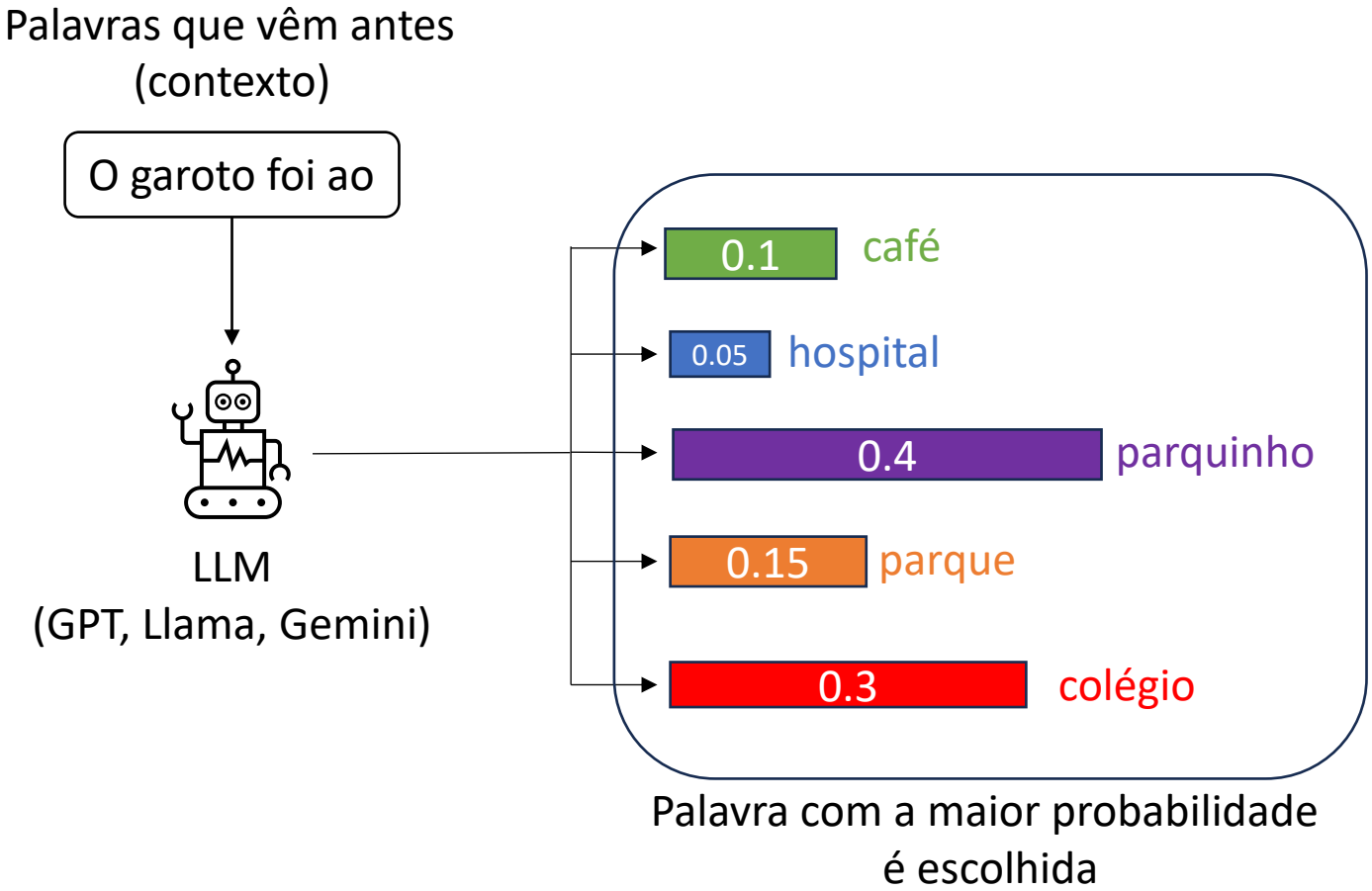
It's Just Adding One Word at a Time

O conceito básico do ChatGPT é, em certo nível, bastante simples. Comece com uma grande amostra de textos criados por humanos, retirados da web, livros, etc. Em seguida, treine uma rede neural para gerar textos que sejam "como esses". E, em particular, faça com que ela seja capaz de começar a partir de um "prompt" e depois continuar com um texto que seja "como o que foi treinado".

E os LLMs (Large Language Models)?



E os LLMs (Large Language Models)?



Indicações e Referências