



# Henrique Santana

Sou um profissional de dados e admirador de personalidades como Cathy O'Neil, Carl Sagan e Hans Rosling.

Atualmente trabalho como Principal Data Engineer na Dell, e ministro aulas na Tera. Tenho bastante interesse por Eletrônica e Computação, e de vez em quando escrevo uns códigos de automação e de inteligência artificial.

EMPRESAS POR ONDE PASSEI



stone

cielo

# IA GENERATIVA NO CONTEXTO DE PRODUTO

# AGENDA

**01**

Como chegamos até aqui?

**02**

Interfaces

**03**

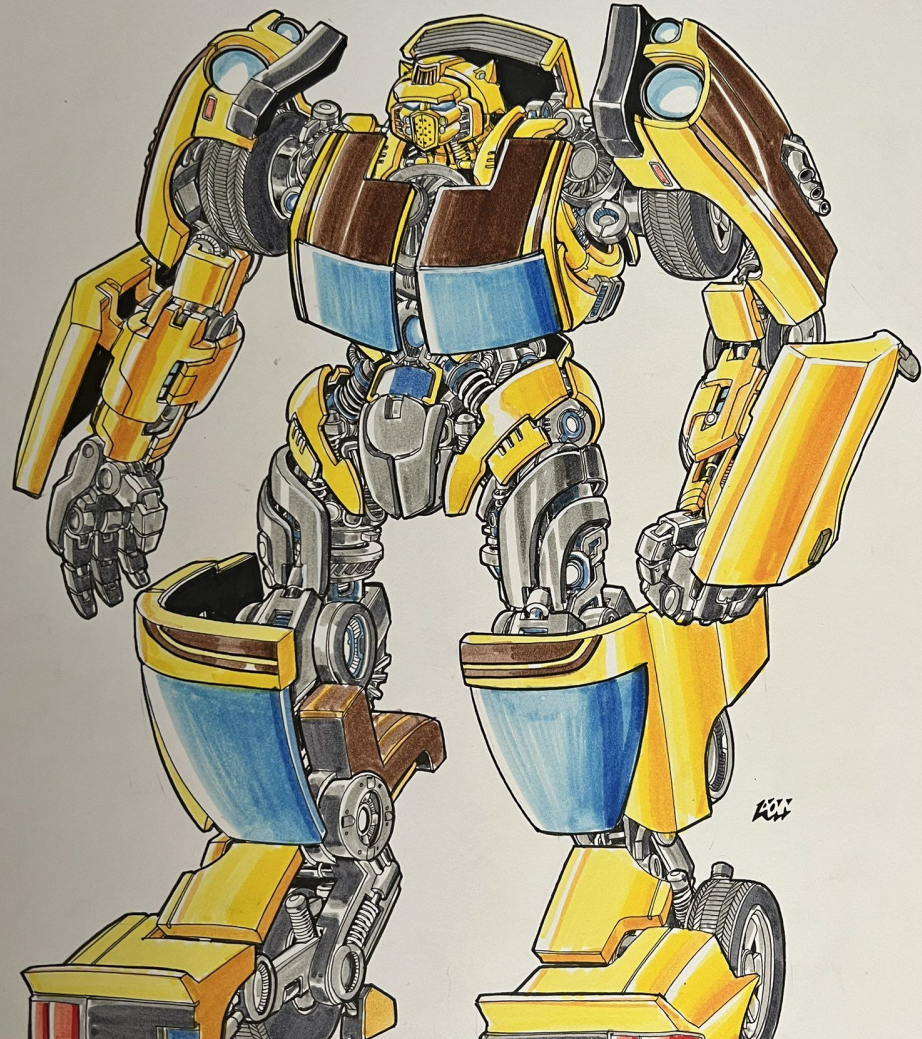
Ferramentas

# Como chegamos até aqui?

# Capítulo 1

Como chegamos  
até aqui?

E onde é exatamente aqui?



# Evolução dos Modelos, Ferramentas e Aplicações

**Eliza**, criada em 1966, foi um dos primeiros Chatbots.

Welcome to

```
EEEEEE LL      IIII  ZZZZZZ  AAAAA
EE      LL      II    ZZ      AA  AA
EEEEEE LL      II    ZZZ      AAAAAA
EE      LL      II    ZZ      AA  AA
EEEEEE LLLLLL  IIII  ZZZZZZ  AA  AA
```

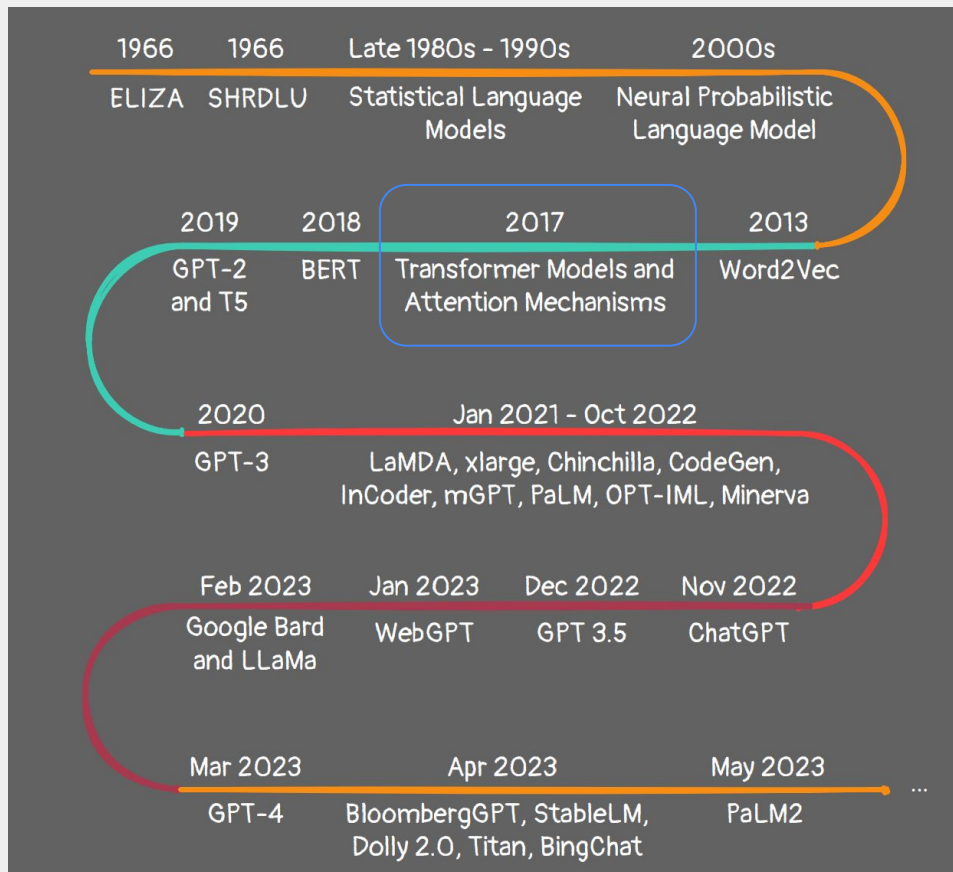
Eliza is a mock Rogerian psychotherapist.  
The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.

```
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```



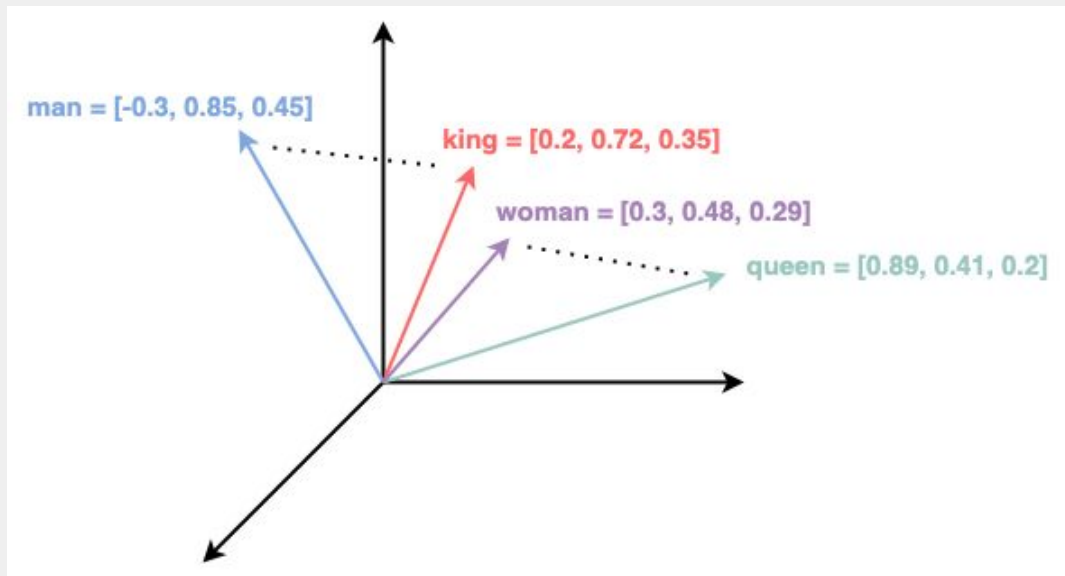
# Evolução dos Modelos, Ferramentas e Aplicações

Em 2017, um grupo de pesquisa do Google lança um paper chamado “**Attention is all you need**”



# NLP

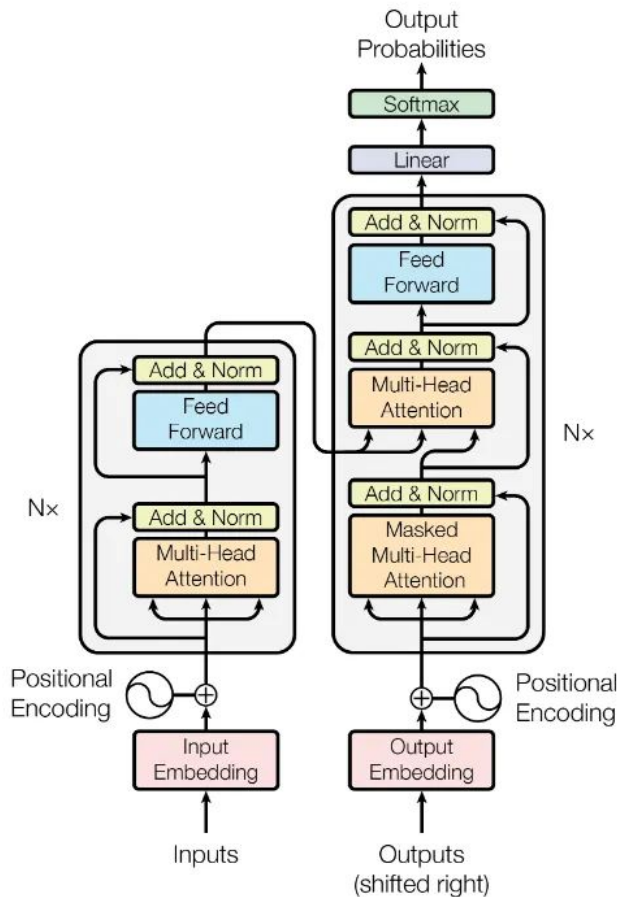
Um dos grandes desafios da computação era fazer uma máquina entender **linguagem natural** (humana).





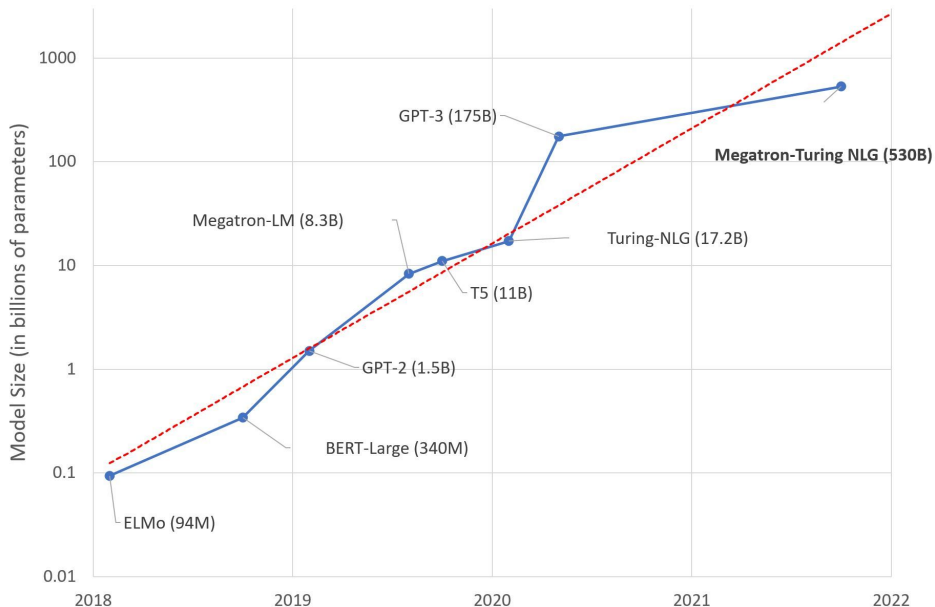
# Transformers

O campo de Deep Learning já conhecia modelos notáveis, como RNN, LSTM e GANs. Mas a arquitetura Transformers foi o que **contribuiu para o salto** que demos em **IA Generativa**.



# LLMs

Os modelos de IA Generativa passaram a ser treinados em **conjunto de dados cada vez maiores**, crescendo o número de **parâmetros**. Daí o nome LLM (Large Language Models).




# Custo com Infraestrutura


O custo das GPUs, por si só, pode chegar a milhões de dólares. Ex: Cada execução de treinamento do GPT-3 da OpenAI exigiu pelo menos **US\$ 5 milhões em GPUs**.

**Forbes**



FORBES > INNOVATION > AI




## What Large Models Cost You – There Is No Free AI Lunch


**Craig S. Smith** Contributor   
*Craig S. Smith is a former correspondent and executive at The New York Times. He is host of the...*

 1

Sep 8, 2023, 10:43am EDT

 Listen to article 6 minutes 

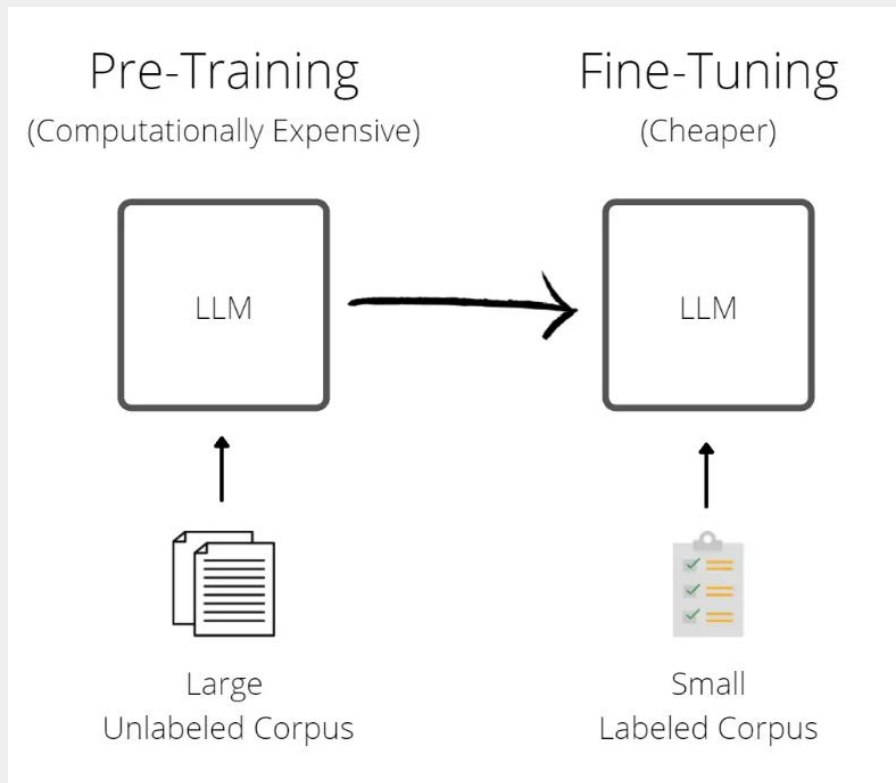


Burning US five and one dollar bills. (Photo by Tom Stoddart/Getty Images) GETTY IMAGES

GPT-4, PaLM, Claude, Bard, LaMDA, LLaMA, Chinchilla, Sparrow – the list of large-language models on the market continues to grow. But behind their remarkable capabilities, users are discovering substantial costs.

# Cortando caminho

As LLMs também popularizaram o conceito de **pré-treinamento** e **ajuste-fino**. Afinal, você não precisa retreinar tudo toda vez.



# Limitações da IA Gerativa

Principais:

- Alucinações
- Imprevisibilidade
- Vieses



# Cuidados com seus Dados

Fique de olho na política de privacidade da empresa que está oferecendo IA Generativa. Muitas delas usam seus dados novas rodadas de treinamento.

## Samsung bane uso de ChatGPT e outras IAs por colaboradores

A justificativa é que usar IA em suas redes internas e dispositivos de propriedade representa um risco à segurança



Por Startupi 14 02/05/2023 14 Atualizado em 02/05/2023 ⌚ Leia em 3 Mins

**Se você usa algum serviço na internet  
sem pagar, então provavelmente você é o  
produto!**



# INTERFACES

# Capítulo 2

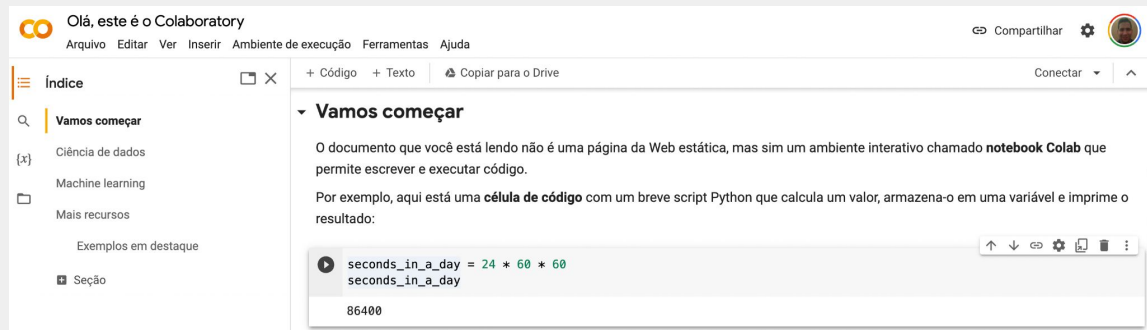
## Interfaces

Meios para interagir com os  
modelos de IA



# Notebooks

Notebooks são ferramentas que apresentam **células** de **processamento de dados** usando uma linguagem de programação (Python) e células de **documentação** (Markdown)



A vertical bar on the left side of the slide with a color gradient from green at the top to blue at the bottom.

colab



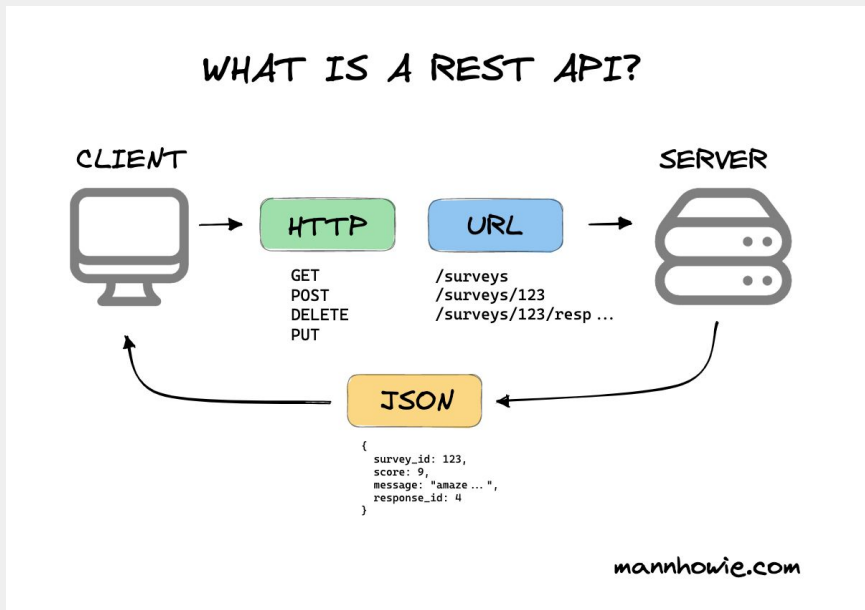
# DEMO

---



# APIs

API (Application Programming Interface) é uma forma de um **sistema** cliente solicitar ou passar uma **informação** sem que o mesmo tenha acesso a todo o banco de dados do sistema servidor.







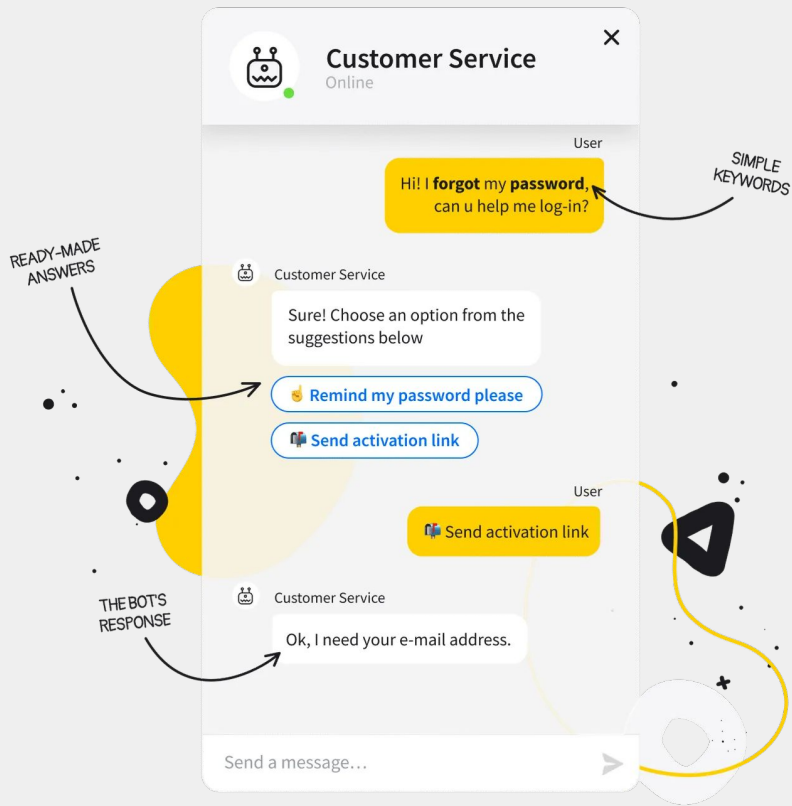
# DEMO

---



# ChatBots

ChatBots são **aplicações** como qualquer outra. A grande vantagem é que ao invés de navegar por páginas através de cliques, você entra em um **modo conversacional** para trocar informações e desempenhar tarefas.





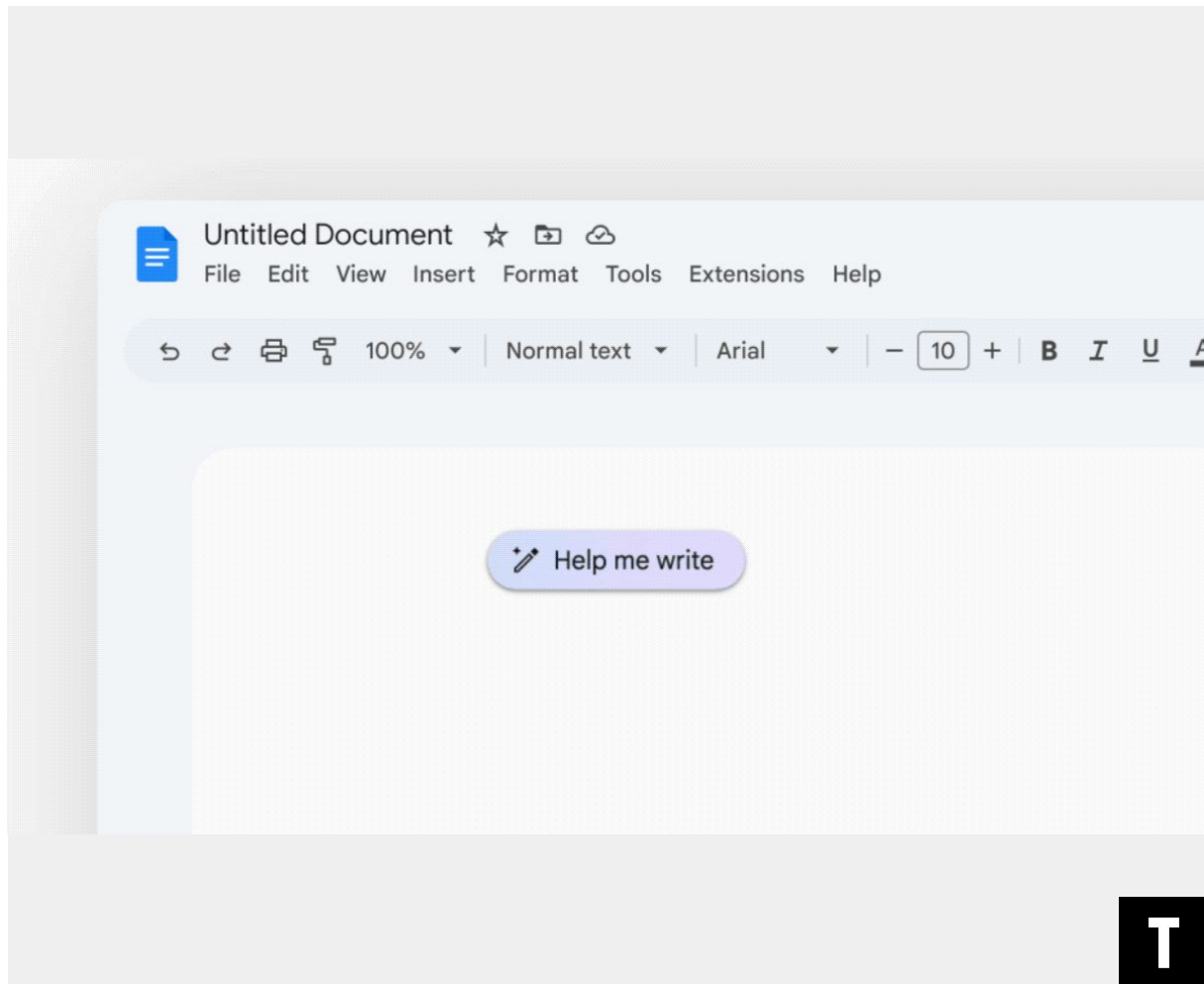
# DEMO

---



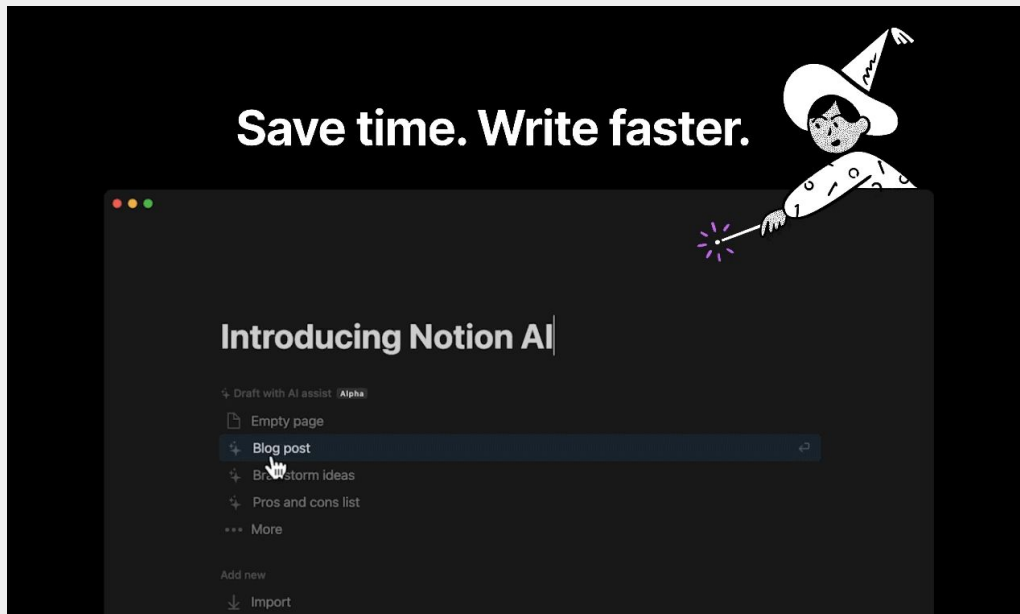
# Outras interfaces conversacionais

Já imaginou usar IA para te ajudar a **redigir um email** ou construir uma **documentação**? Cada vez mais, as gigantes da produtividade estão adicionando esses recursos aos seus produtos.



# Outras interfaces conversacionais

Já imaginou usar IA para te ajudar a **redigir um email** ou construir uma **documentação**? Cada vez mais, as gigantes da produtividade estão adicionando esses recursos aos seus produtos.



# FERRAMENTAS



# Capítulo 3

## Ferramentas

Acessórios para manter em seu  
cinto de utilidades



# Modelos Pré-Treinados na HuggingFace

A plataforma onde a comunidade de Machine Learning colabora em modelos, conjuntos de dados e aplicativos.



## Transformers

build

passing

license

Apache-2.0

website

online

release

v2.0.0



# Hugging Face

# DEMO

---



# Conversação e Plugins no ChatGPT

Plugins são **ferramentas** projetadas especificamente para modelos de linguagem tendo a segurança como princípio fundamental e ajudam o **ChatGPT** a acessar **informações atualizadas**, executar cálculos ou usar **serviços de terceiros**.





Crie um csv com 50 linhas de dados aleatórios de pessoas com as seguintes colunas: nome, email, cidade, estado. As cidades e estados precisam ser do Brasil.



Claro, vou criar uma estrutura para os dados que você solicitou. Primeiro, vou mostrar uma prévia da estrutura do CSV com algumas linhas de dados. Se estiver tudo certo, gerarei o arquivo CSV completo com 50 linhas de dados aleatórios. Aqui está a estrutura de dados proposta:



Nome	Email	Cidade	Estado
João Silva	<a href="mailto:joao.silva@email.com">joao.silva@email.com</a>	São Paulo	SP
Maria Luiza	<a href="mailto:maria.luiza@email.com">maria.luiza@email.com</a>	Rio de Janeiro	RJ
Fernando C.	<a href="mailto:fernando.c@email.com">fernando.c@email.com</a>	Curitiba	PR
...	...	...	...

Esses dados são apenas exemplos para ilustrar a estrutura. Posso proceder com essa estrutura e gerar dados aleatórios para as 50 linhas? Você gostaria de fazer

☐ Parar de gerar

Envie uma mensagem

..







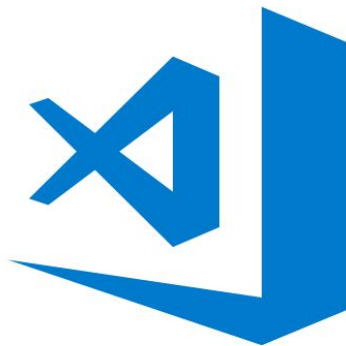
# DEMO



# Pair-Programming com o GitHub Copilot

Gaste menos tempo com códigos repetitivos. Escreva um comentário descrevendo a lógica desejada e o GitHub Copilot vai **sugerir** imediatamente o **código** para implementar a solução.





Visual Studio Code

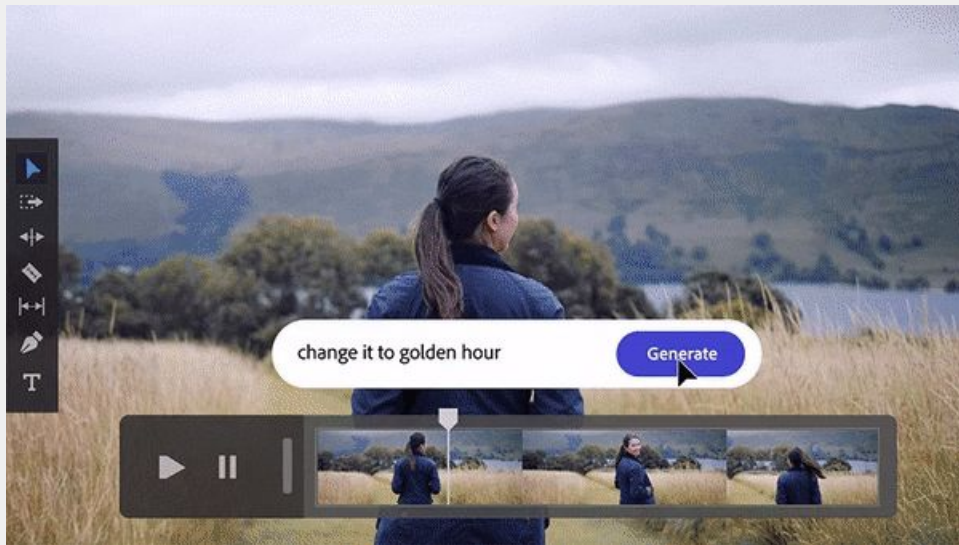


# DEMO



# IAs Generativas para Imagens

A Adobe não ficou de fora, e também está ampliando os recursos de IA Generativa em seus produtos, como o Firefly.



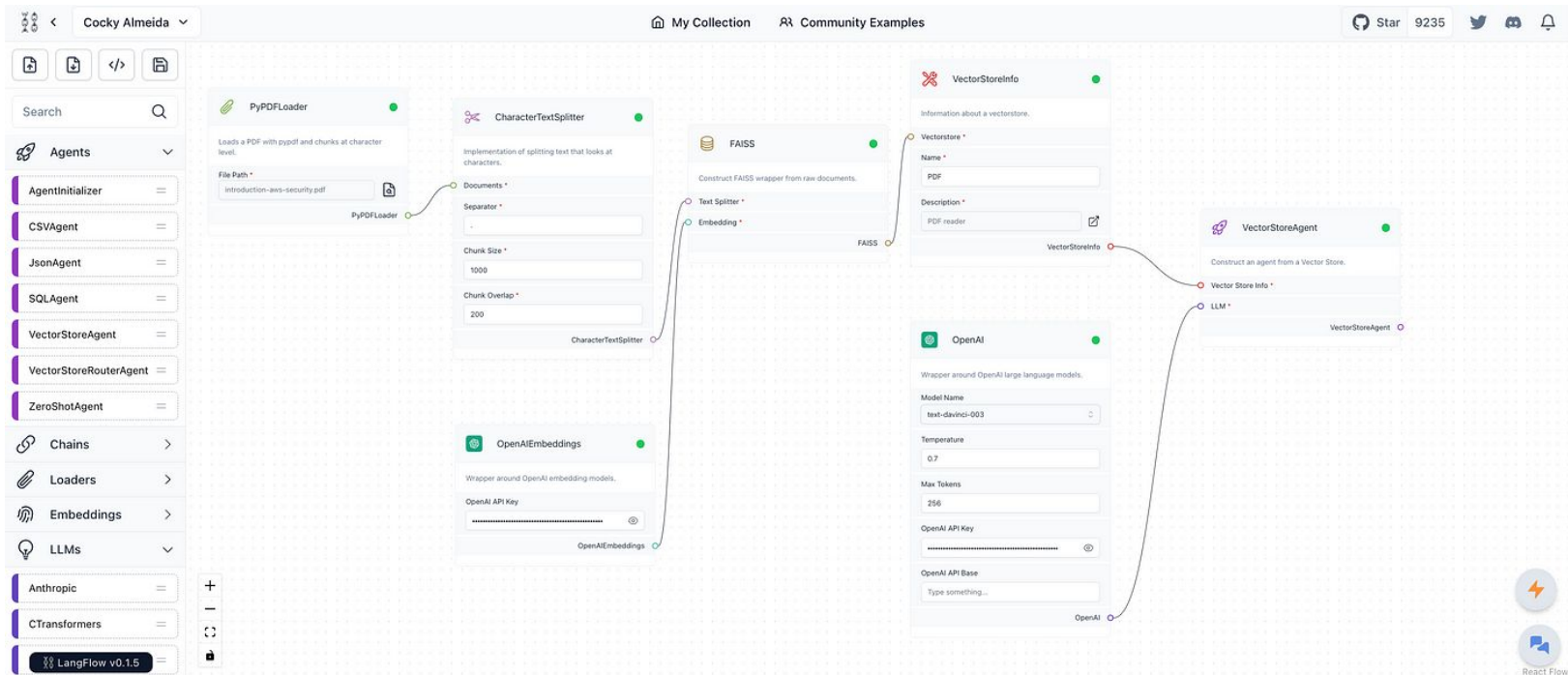
# Extra: LangChain

É um framework que vai além das chamadas de API padrão por ser ciente de dados e agente, permitindo conexões com várias fontes de dados para experiências mais ricas e personalizadas.





# Extra: LangChain





# RESUMO

# TAKEAWAY



**Lembre-se: se você usa algum serviço na internet sem pagar, então provavelmente você é o produto!**

# TAKEAWAY



**Não existe bala de prata! Use a ferramenta mais adequada para resolver um problema**

# TAKEAWAY



**Esteja sempre alerta, a tecnologia muda muito rápido.**

