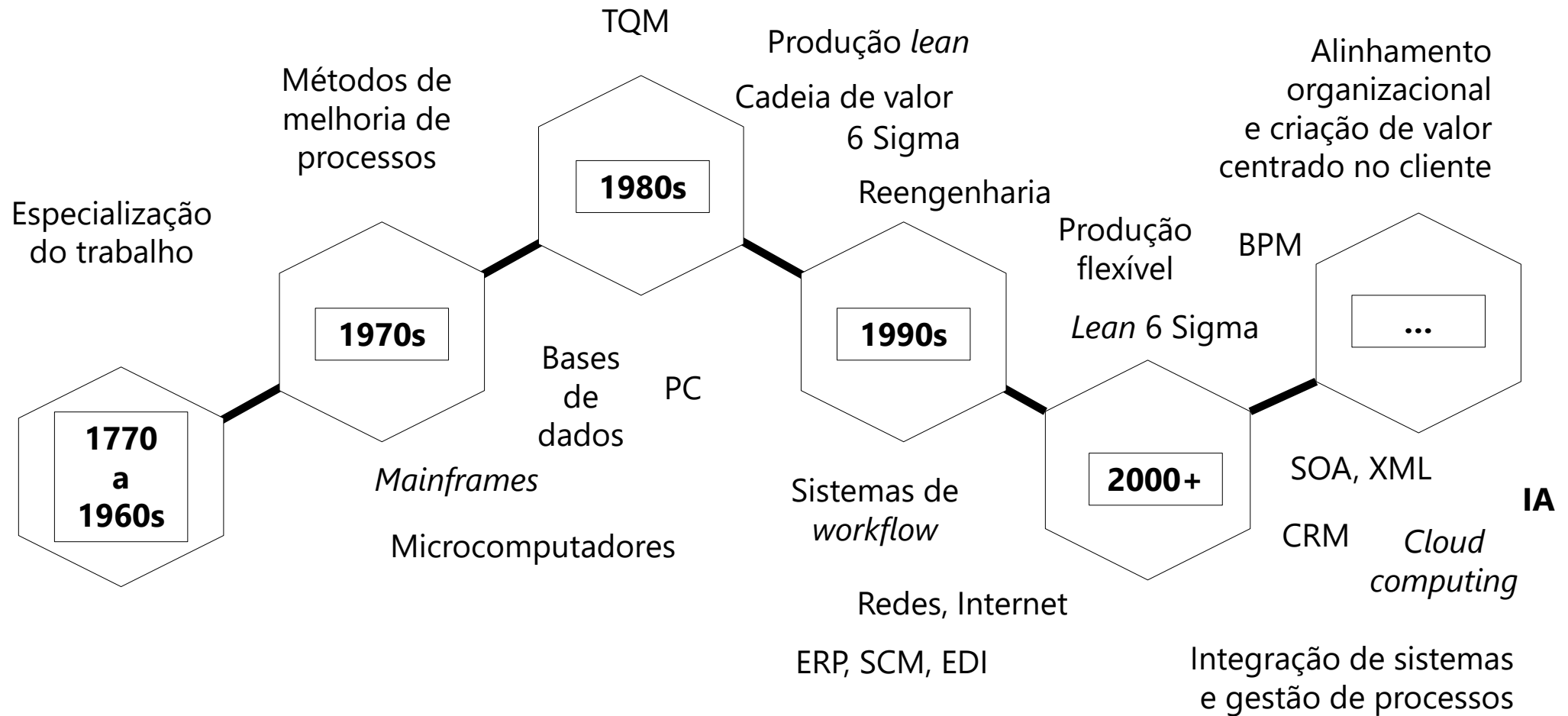


# IA em ambiente empresarial

Gestão de Informação e dos Processos de Negócio

# Enquadramento



## ➔ **Diversidade de informação para melhores decisões**

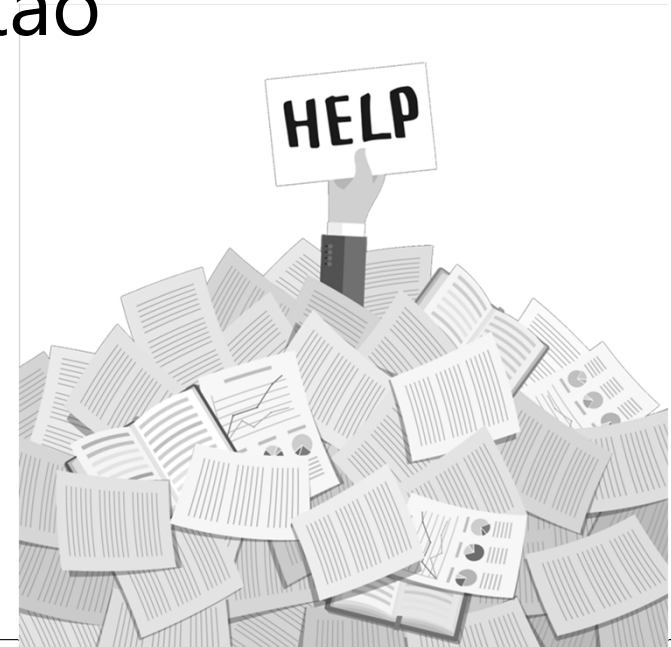
A extração e interpretação de informação a partir de diferentes fontes de informação para **potenciar melhorias nos processos de decisão** é, ainda hoje, uma preocupação permanente na gestão



## → Quantidade para melhores decisões

Não só a diversidade da informação é importante, mas também a questão da **quantidade de informação** é também relevante.

A **IA** tem aqui um papel a desempenhar





# Enquadramento

## → Cada vez mais informação!



Notícias online



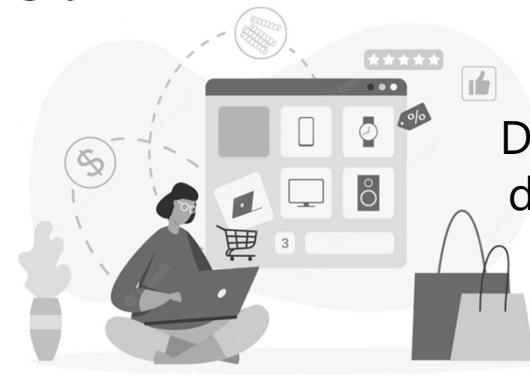
Relatórios empresariais



Informação online



Discurso nas redes sociais



Dados da interação de compras online

Atividades de CRM



## Enquadramento

### → Tempo para melhores decisões

Necessidade de tomar decisões em tempo adequado/oportuno.

BETTER THREE HOURS TOO  
SOON THAN A MINUTE TOO  
LATE. WILLIAM SHAKESPEARE



## → Decisores para tomar decisões

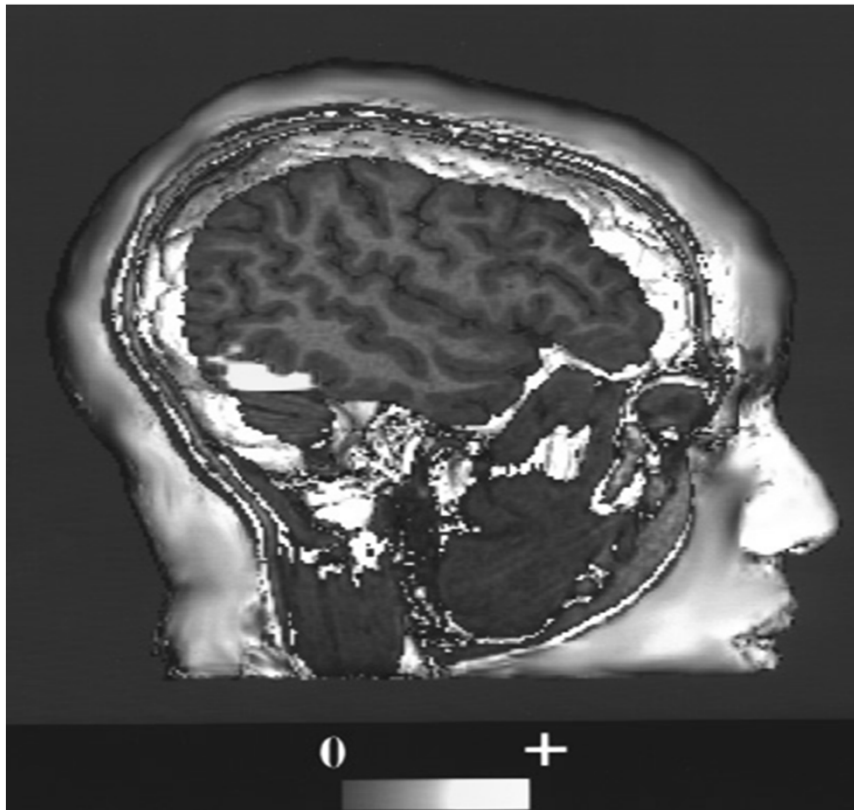
Um dos fatores mais importantes para um processo de tomada de decisão é o próprio **agente de decisão**.



## Curiosidade #01

---

# Áreas de reconhecimento facial



# Procura por mais informação

## → Cada vez mais informação



Notícias online



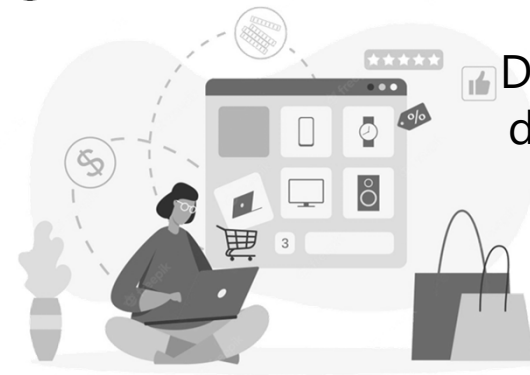
Relatórios empresariais



Informação online



Discurso nas redes sociais

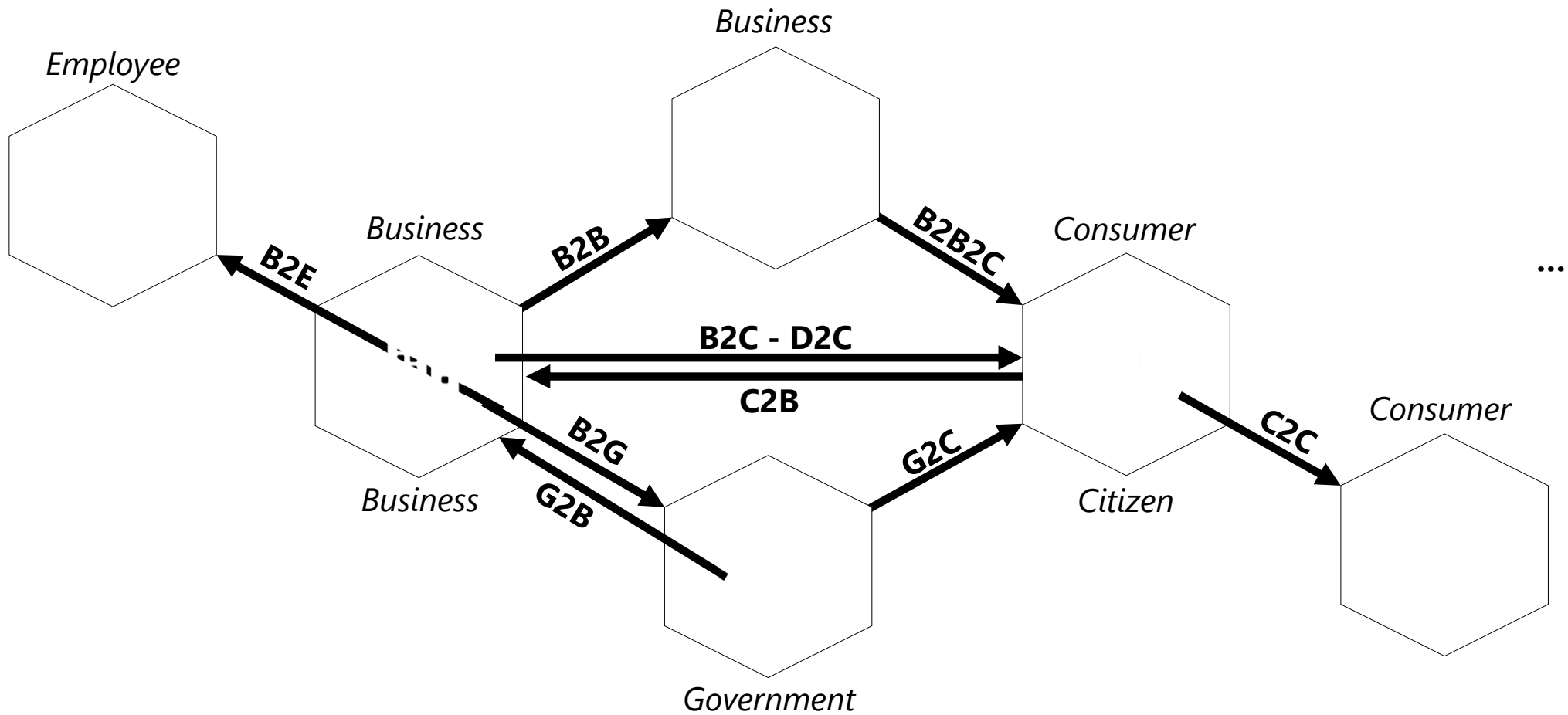


Dados da interação de compras online

Atividades de CRM



# Modelos de negócio eletrônico



## Procura por mais informação

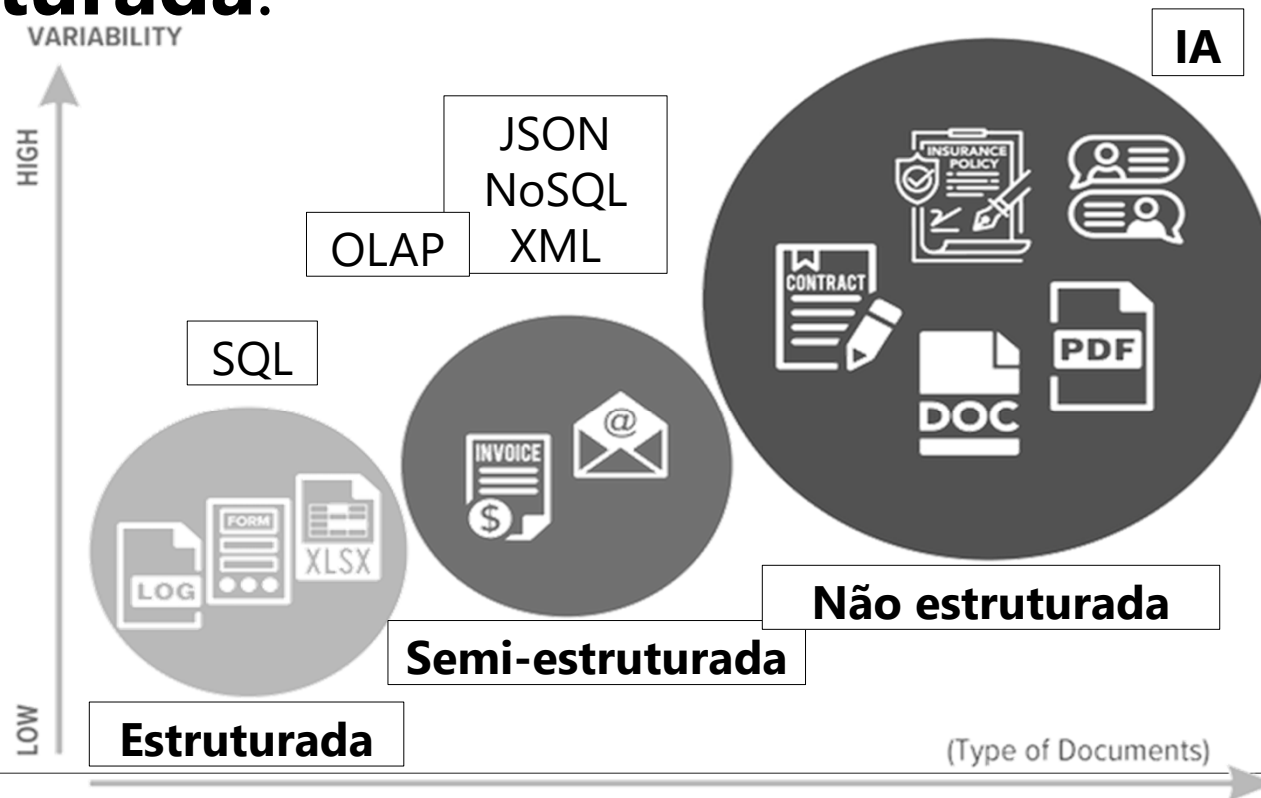
---

É possível obter **mais informação** para tomada de decisão?



## Procura por mais informação

Maior capacidade para lidar com informação **não estruturada**.





Procura por mais informação



LiDAR (Light Detection And Ranging)

## Pré-processamento do texto

Os dados obtidos a partir de *posts*, avaliações, comentários, observações e críticas em textos, nomeadamente RSO, são **altamente não estruturados**.

Assim o pré-processamento é uma etapa crítica, que impacta significativamente a qualidade/validade do trabalho seguinte.

## Extração de características (*feature extraction*)

Processo de **ML** no qual informações relevantes ou características são selecionadas e extraídas dos dados brutos.

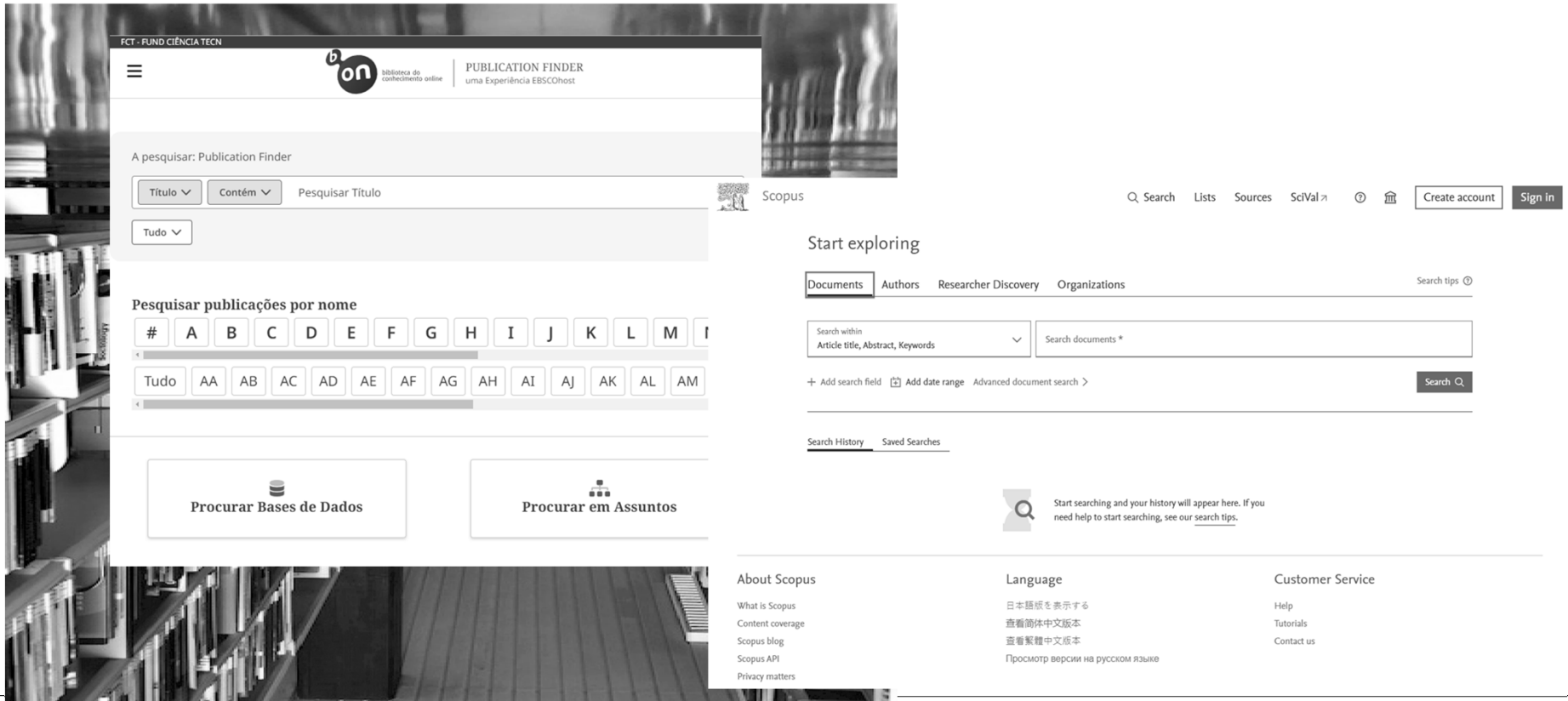
Na análise de sentimentos/emoções, as características podem incluir frequências de palavras, sequências de palavras adjacentes ou outros atributos linguísticos.

## → ***Machine learning – Vector databases***

Uma base de dados vetorial é um tipo de base de dados que armazena e gere dados não estruturados, como texto, imagens ou áudio, em ***embeddings*** vetoriais (vetores de alta dimensionalidade) para facilitar a busca e recuperação rápida de objetos semelhantes.

# Conceitos adicionais

## → *BD relacionais vs vetoriais*



The screenshot displays the Scopus database interface. At the top, it shows the logo for 'FCT - FUNDO CIÊNCIA TECN' and 'PUBLICATION FINDER'. Below this, there's a search bar with a dropdown menu for 'Pesquisar Título' and a 'Pesquisar' button. To the right, there are links for 'Search', 'Lists', 'Sources', 'SciVal', and 'Create account'. Below the search bar, there's a section titled 'Start exploring' with tabs for 'Documents', 'Authors', 'Researcher Discovery', and 'Organizations'. A search box is provided with a dropdown menu for 'Search within' and a 'Search' button. At the bottom, there are links for 'About Scopus', 'Language', and 'Customer Service'.

**Search Interface Details:**

- Search Bar:** Includes a dropdown menu for 'Pesquisar Título' and a 'Pesquisar' button.
- Search Options:** Includes a 'Tudo' dropdown menu and a 'Pesquisar publicações por nome' section with a grid of letters (A-M, AA-AM).
- Search Results:** Includes a 'Start exploring' section with tabs for 'Documents', 'Authors', 'Researcher Discovery', and 'Organizations'.
- Search Box:** Includes a dropdown menu for 'Search within' and a 'Search' button.
- Footer:** Includes links for 'About Scopus', 'Language', and 'Customer Service'.

Conceitos adicionais



BD relacionais vs vetoriais

amazon

Deliver to Portugal

All

book where the main character is a wizard girl

EN

Hello Account

All

Today's Deals

Customer Service

Registry

Gift Cards

Sell

4 results for "book where the main character is a wizard girl"

Department

Books

Fantasy

Literature & Fiction

Teen & Young Adult Literature & Fiction

Teen & Young Adult Science Fiction & Fantasy

Manga Comics & Graphic Novels

Kindle Store

Science Fiction & Fantasy

Teen & Young Adult eBooks

Literature & Fiction

Results

Editors' pick

WHERE THE CRAWDADS SING

DELIA OWENS

THE WORLDWIDE SENSATION

NEW YORK TIMES BESTSELLER

Where the Crawdads Sing

by Delia Owens | Aug 14, 2018

★★★★★ ~ 627,584

Kindle

\$12.99 Print List Price: \$18.00

Available instantly

Paperback

\$9.50 List: \$18.00

Delivery Tue, Jun 25

Ships to Portugal

More Buying Choices

\$1.22 (62+ used & new offers)

Audible Audiobook

\$0.00 \$20.25

Free with Audible trial

Available instantly

Other formats: Hardcover , Audio CD

Delia Owens

The Wonderful Wizard of Oz: A Classic Novel for All Ages (Annotated)

by L. Baum and Storytime Cottage | May 5, 2023

Kindle

\$2.99 Print List Price: \$8.99

"She laughed for his sake, something she'd never done. Giving away another piece of herself just to have someone else."  
Highlighted by 24,334 Kindle readers

https://www.facebook.com/photo/?fbid=737424471737789&set=pcb.737446831735553

## Conceitos adicionais

## → Machine learning - Vector embedding (VE)

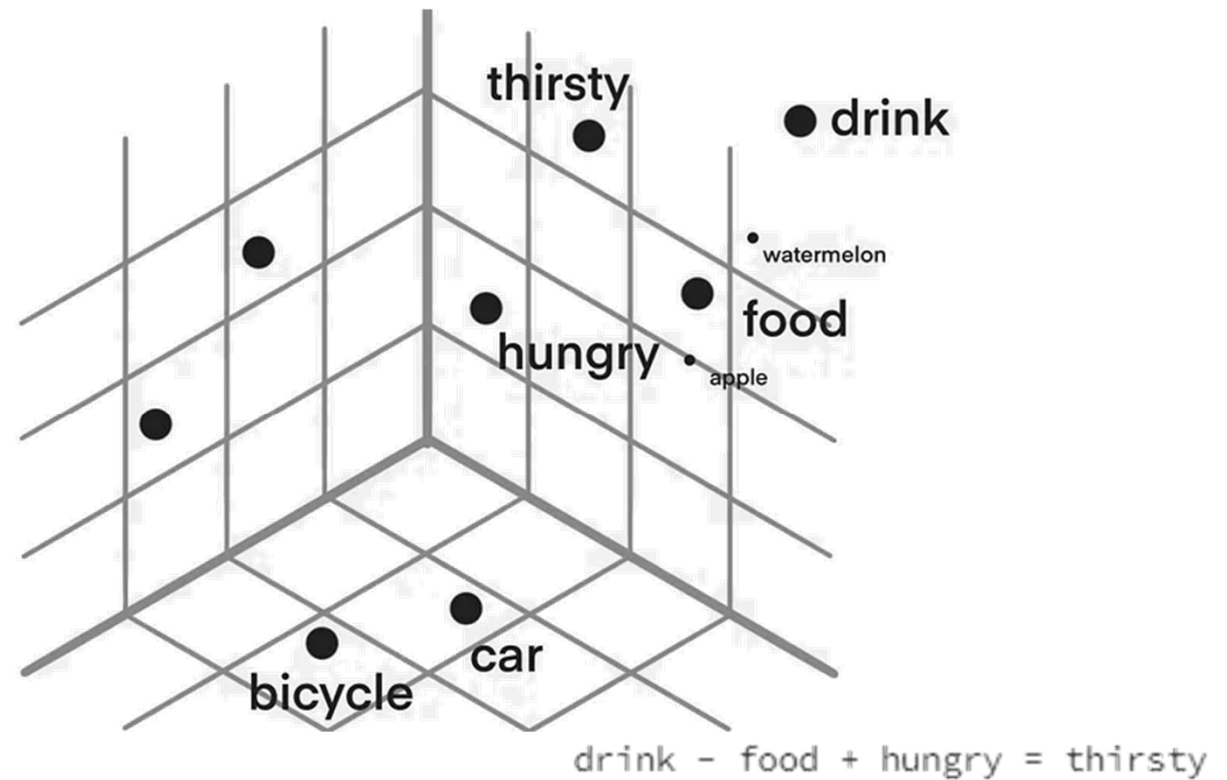


**Grape** ➡ [0.2, 1.1, 1.2, 3.1...]

**Castle** ➡ [0.1, 2.8, 3.9, 4.2...]

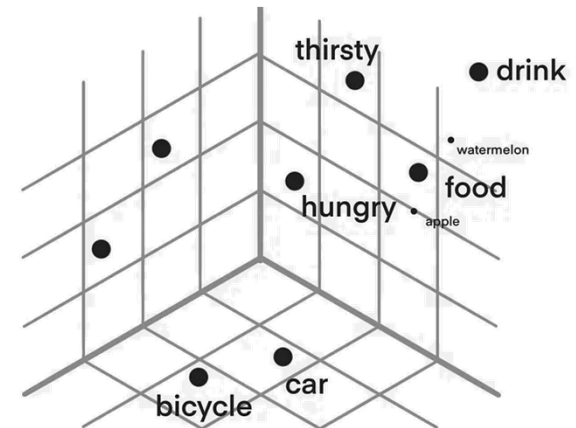
**Orange** ➡ [2.3, 0.9, 3.1, 5.3...]

**Person** ➡ [5.0, 0.8, 0.1, 6.4...]



## Conceitos adicionais

As pesquisas em VD pretendem encontrar documentos, imagens ou outros dados que sejam semelhantes ou "**aproximados**" de alguma maneira (por exemplo características ou contexto).

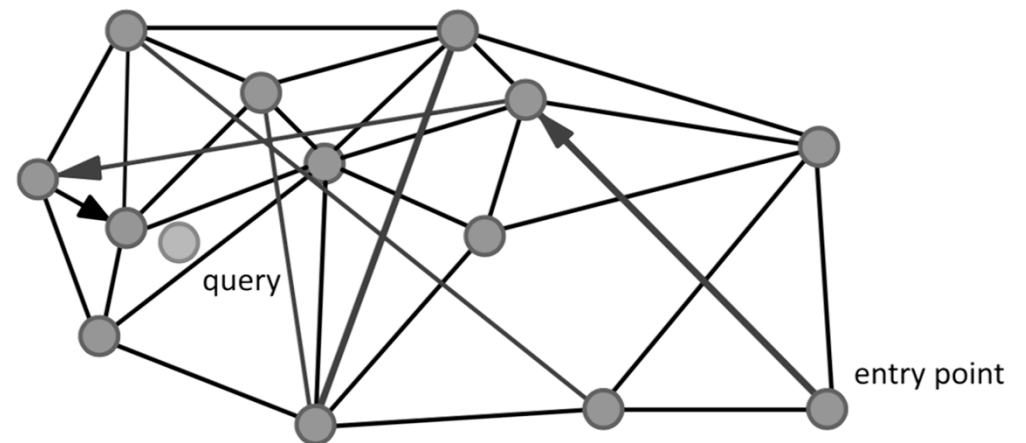




## Conceitos adicionais

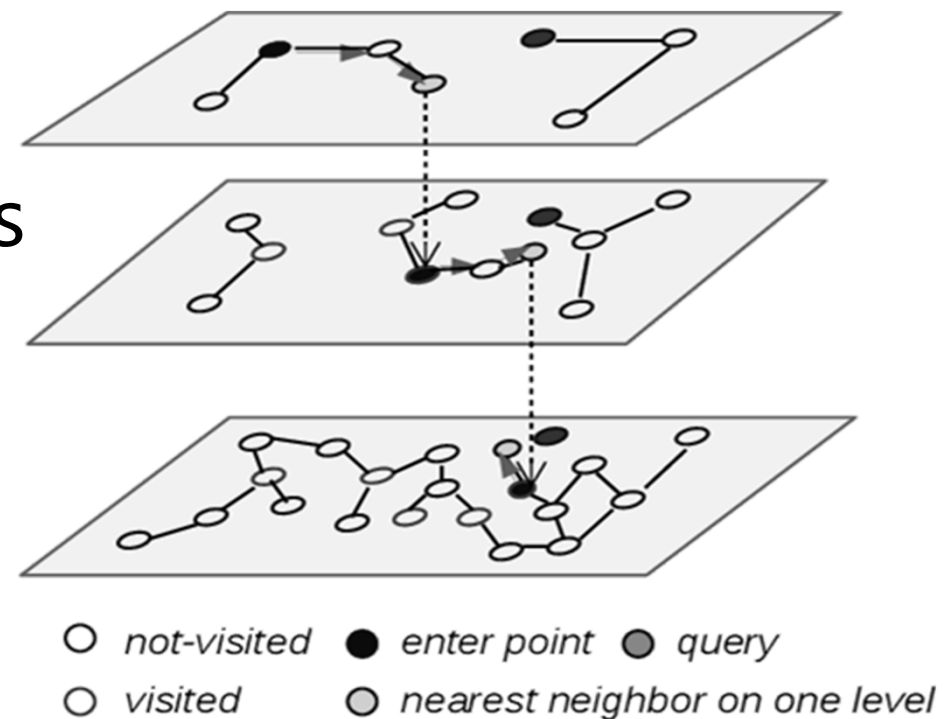
### **Exact Nearest Neighbor (kNN)**

O resultado de uma pesquisa é calculado comparando o **vetor da query** com **todos** os outros vetores da VD.



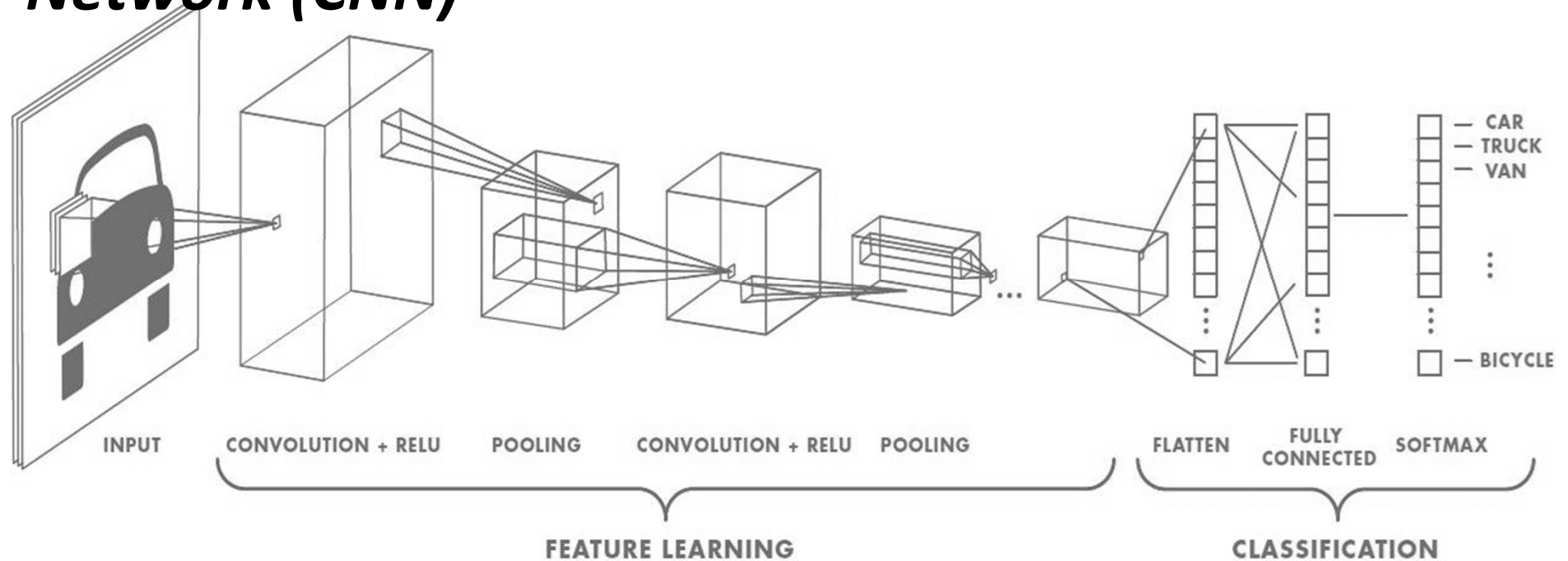
## **Approximate Nearest Neighbor (ANN)**

Com o **ANN**, troca-se alguma precisão por velocidade para recuperar os objetos aproximadamente mais semelhantes a uma consulta.



## Conceitos adicionais

## → *Deep Learning - Convolution Neural Network (CNN)*



## Conceitos adicionais

→ ***Deep Learning - Convolution Neural Network (CNN)***

|    |    |
|----|----|
| -1 | -1 |
| 1  | 1  |

Filter

|    |    |    |
|----|----|----|
| -1 | -2 | -1 |
| 0  | 0  | 0  |
| 1  | 2  | 1  |

Detect

|    |    |   |
|----|----|---|
| -2 | -1 | 0 |
| -1 | 1  | 1 |
| 0  | 1  | 2 |

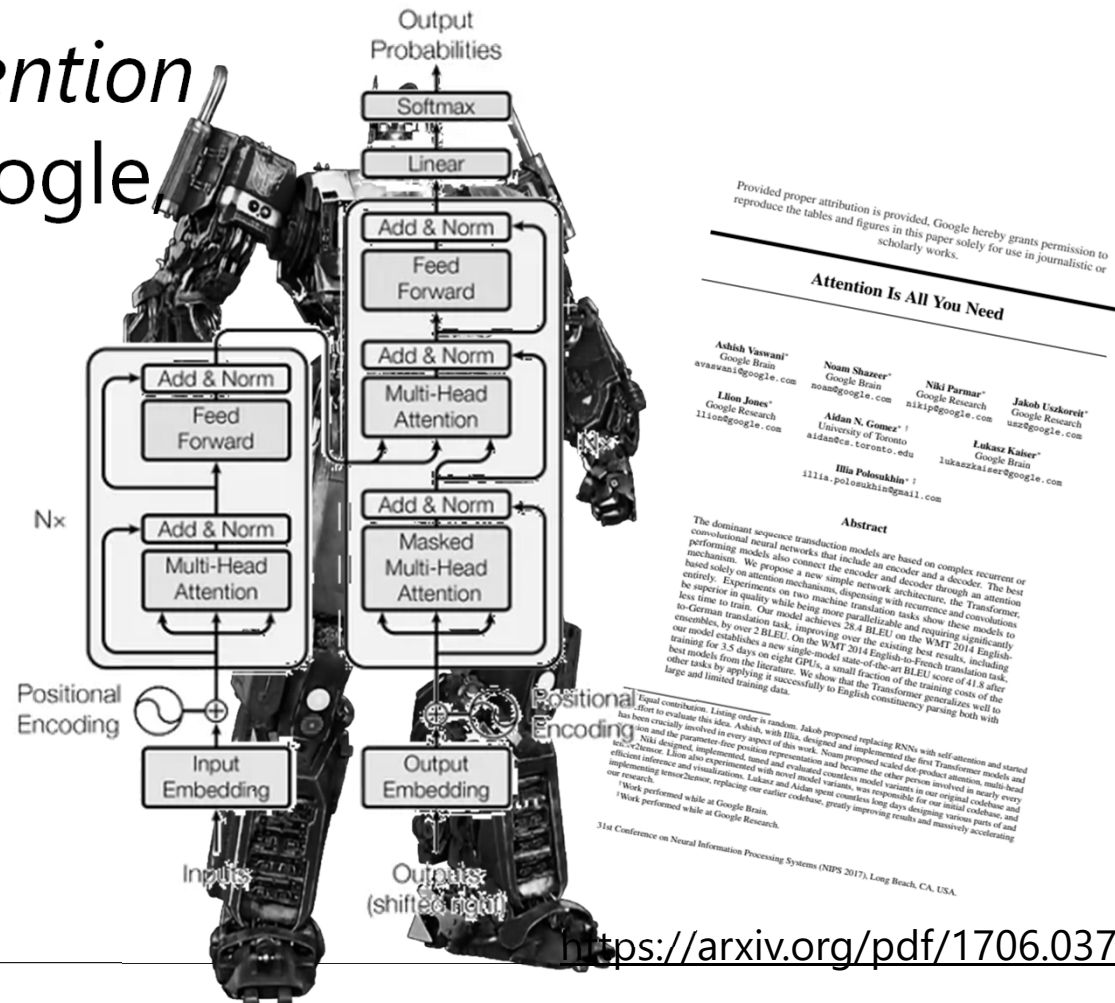
Condense

Input



2017: Ano da mudança

O Research Paper "Attention Is All You Need" da Google apresentou uma **nova** arquitetura de **deep learning** conhecida como o: **Transformer**.



<https://arxiv.org/pdf/1706.03762.pdf>

## Arquitetura "Transformer"

---

### → ***Deep Learning – Transfer Learning***

Bidirectional Encoder Representations from Transformers (BERT)



## Arquitetura "Transformer"

---

### → ***Deep Learning – Long-short term memory (LSTM), Bidirectional Long-short term memory (BI-LSTM)***

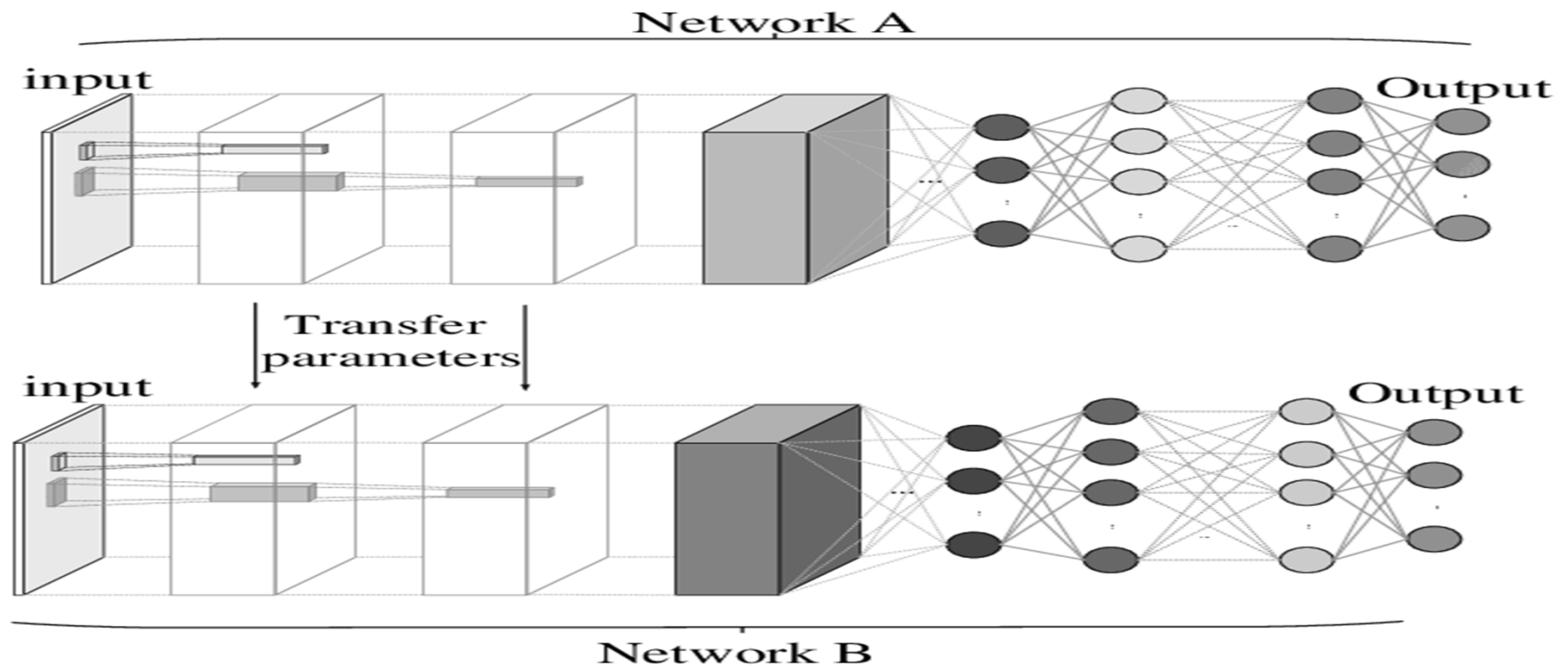
| Apple is something that competitors simply cannot reproduce.

and

| Apple is something that I like to eat.

# Arquitetura "Transformer"

## → *Deep Learning – Transfer Learning*





## Desafios do processamento de linguagem natural

---

“Está um dia maravilhoso!”

“Adoro o meu iPhone, mas a bateria podia ser melhor.”

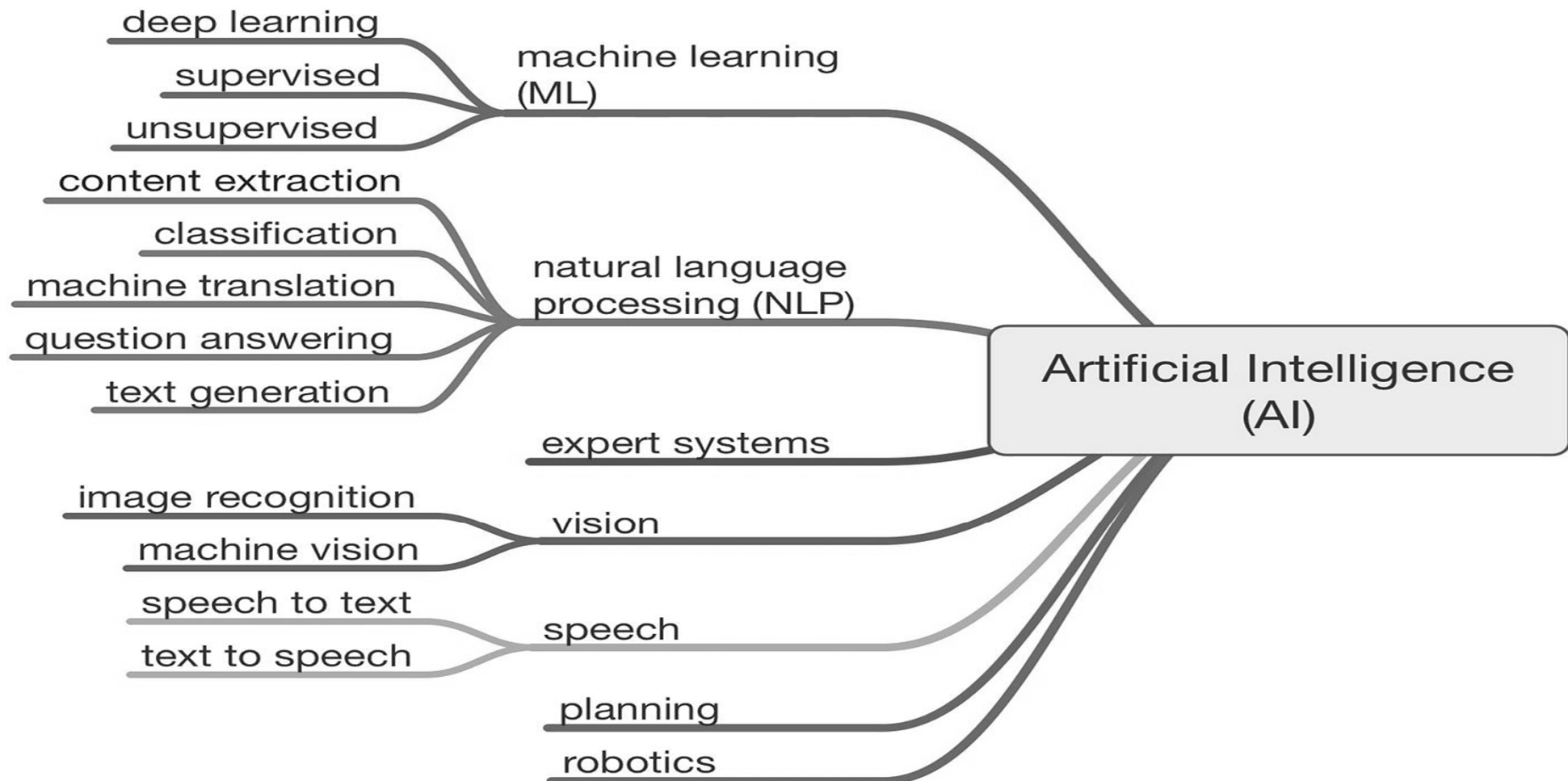
“Tenho saudades de ti!”

“Este meu carro dá-me saudades do anterior e desejo pelo próximo.”

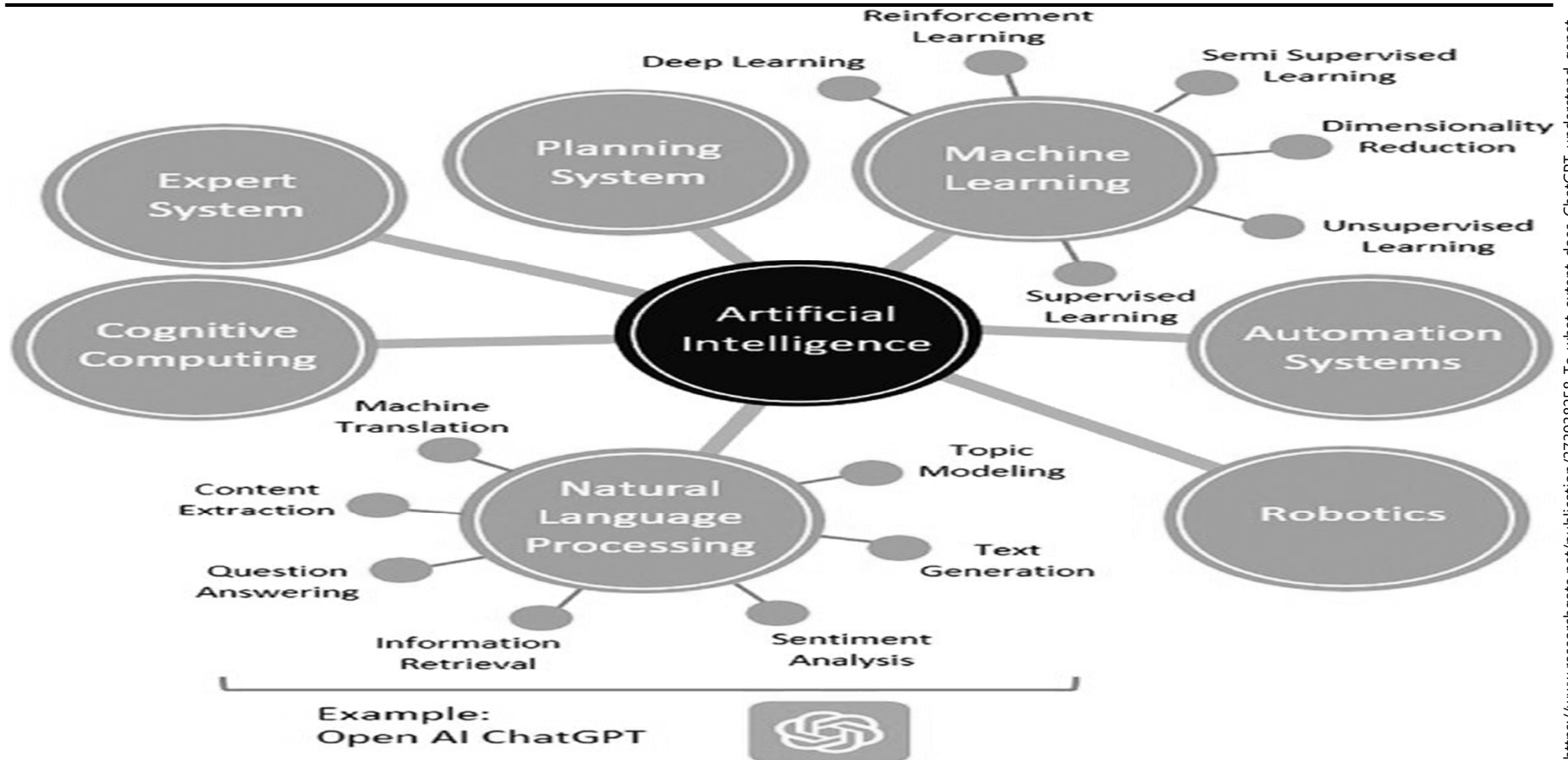
# Desafios do processamento de linguagem natural



# Tipos de inteligência artificial



# Tipos de inteligência artificial



# Tipos de inteligência artificial

## TYPES OF AI



### Reactive AI

Can only respond to the current state of the world. They do not have any memory of past events, and they cannot plan for the future.



### Limited-memory

Can remember past events and use this information to make decisions. Cannot reason about the future or understand the intentions of other agents.



### Theory of Mind

Can understand the thoughts and intentions of other agents. Cooperate with other agents and to achieve goals that would be impossible for a single agent.



### Self Aware

## Curiosidade #02

É anterior a 201

Google

translate

Tudo

Imagens

Videos

Noticias

Livros

Português

Doutoramento em  
Gestão

2006 (18 anos)

18:12

Cancel



Favo

Set

This hel  
wl

Continue

Novo separador

x

+

G ubi licenciatura em g|

Q ubi licenciatura em g - Google Pesquisar

... gestão

... gestao

... geografia

... geologia

... gestão industrial

... gespa

... grp

... ginasio

... graduação

2004 (20 anos)

## IA Generativa

É um tipo de inteligência artificial (IA) que usa técnicas de aprendizagem profunda e conjuntos de dados massivamente grandes para entender, resumir, gerar e prever **novo conteúdo**.

### Artificial Intelligence

Is the field of study

### Machine Learning

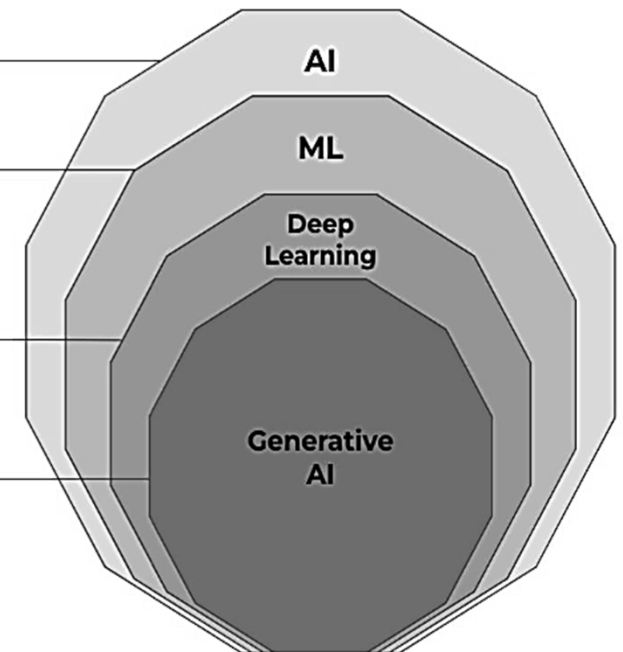
Is a branch of AI that focus on the creation of intelligent machines that learn from data. Another very well know branch inside AI is **Optimization**.

### Deep Learning

Is a subset of Machine Learning methods, based on **Artificial Neural Networks**.  
Examples: CNNs, RNNs

### Generative AI

A type of ANNs that generate data that is similar to the data it was trained on.  
Examples: GANs, LLMs



IA: Geração | Previsão | Co

A IA **preditiva**, em  
artificial **generativ**  
dados históricos p  
eventos, ou mesm  
Organizações util  
tomada de decisã  
orientadas por da

SICNOTÍCIAS

PAÍS

## Investigadora portuguesa cria sistema que deteta ataque cardíaco com recurso à Inteligência Artificial

A doença coronária é considerada a principal causa de morte no mundo, sendo a dor torácica um dos motivos mais comuns de admissão nos serviços de urgência.

Rui Carlos Teixeira, Ana Veloso e Miguel Castro  
17:25, 30 jan.2024



em  
car

norar a

<https://sicnoticias.pt/pais/2024-01-30-Investigadora-portuguesa-cria-sistema-que-deteta-ataque-cardiaco-com-recurso-a-Inteligencia-Artificial-ac7cf8e8>



## IA: Geração | Previsão | Conversação

Através de técnicas de processamento de  
linguagem natural (NLP) e ML, a IA pode ser usada para  
**conversação**

assistentes  
aterro  
linguagem  
de forma

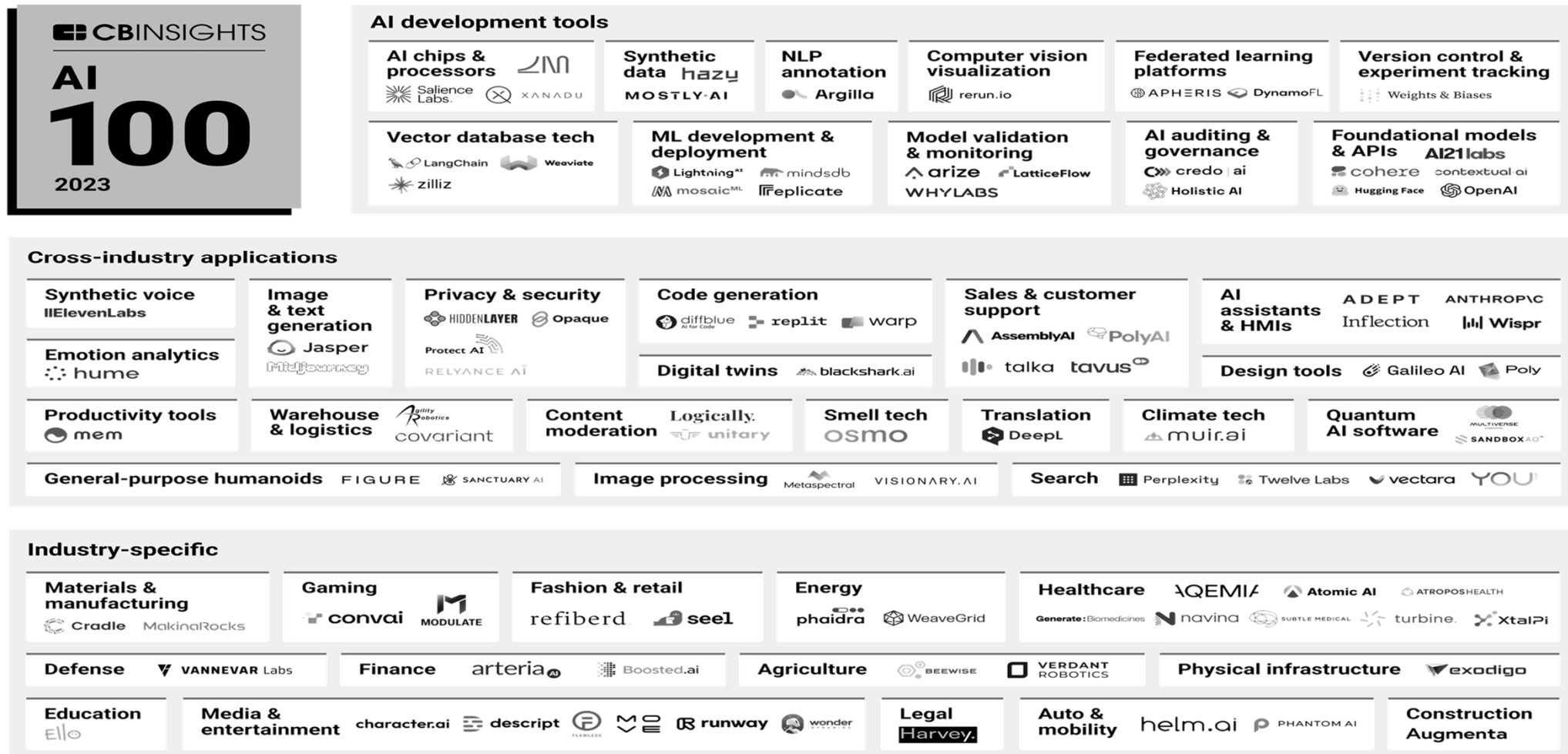


texto ou

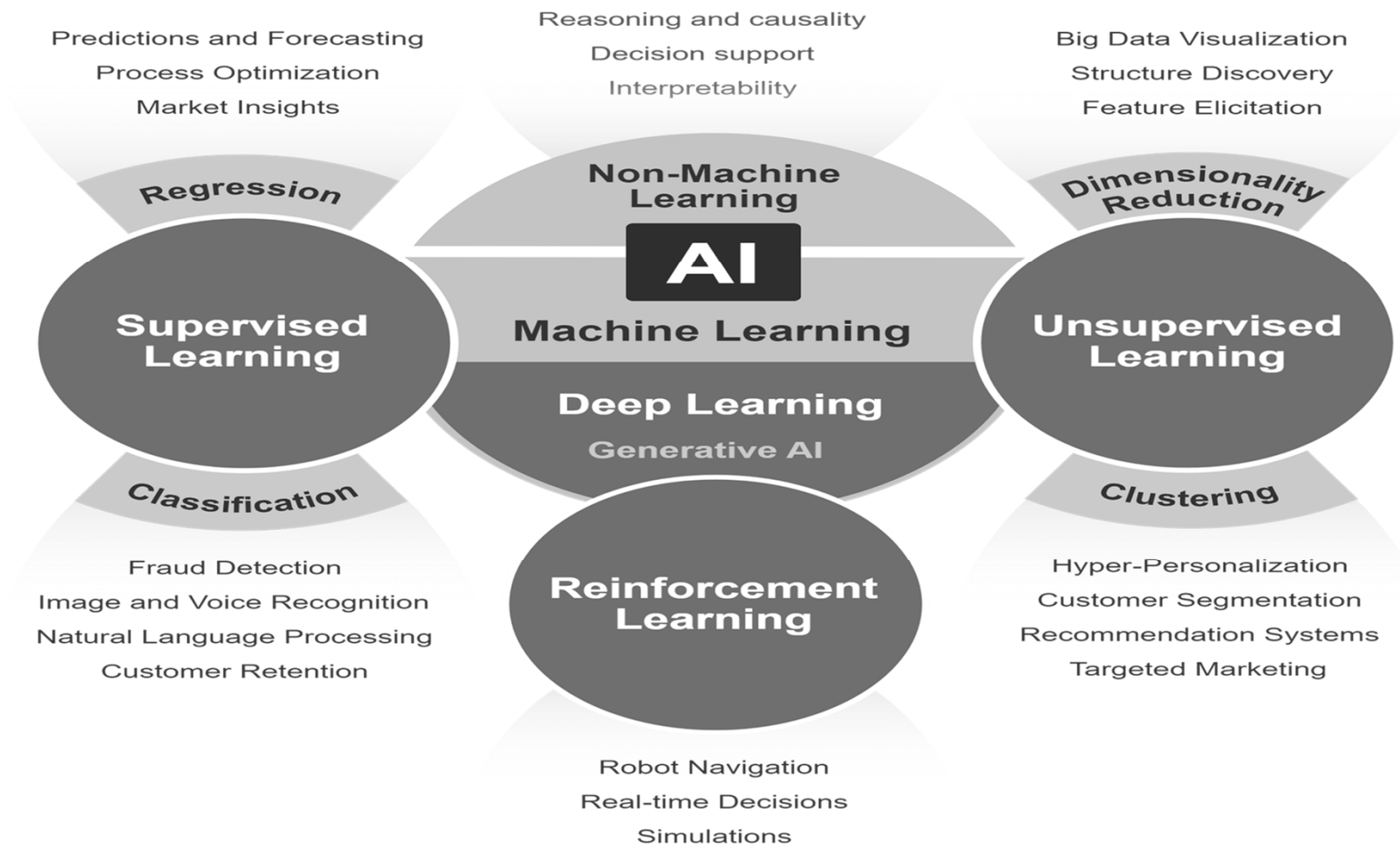
36.



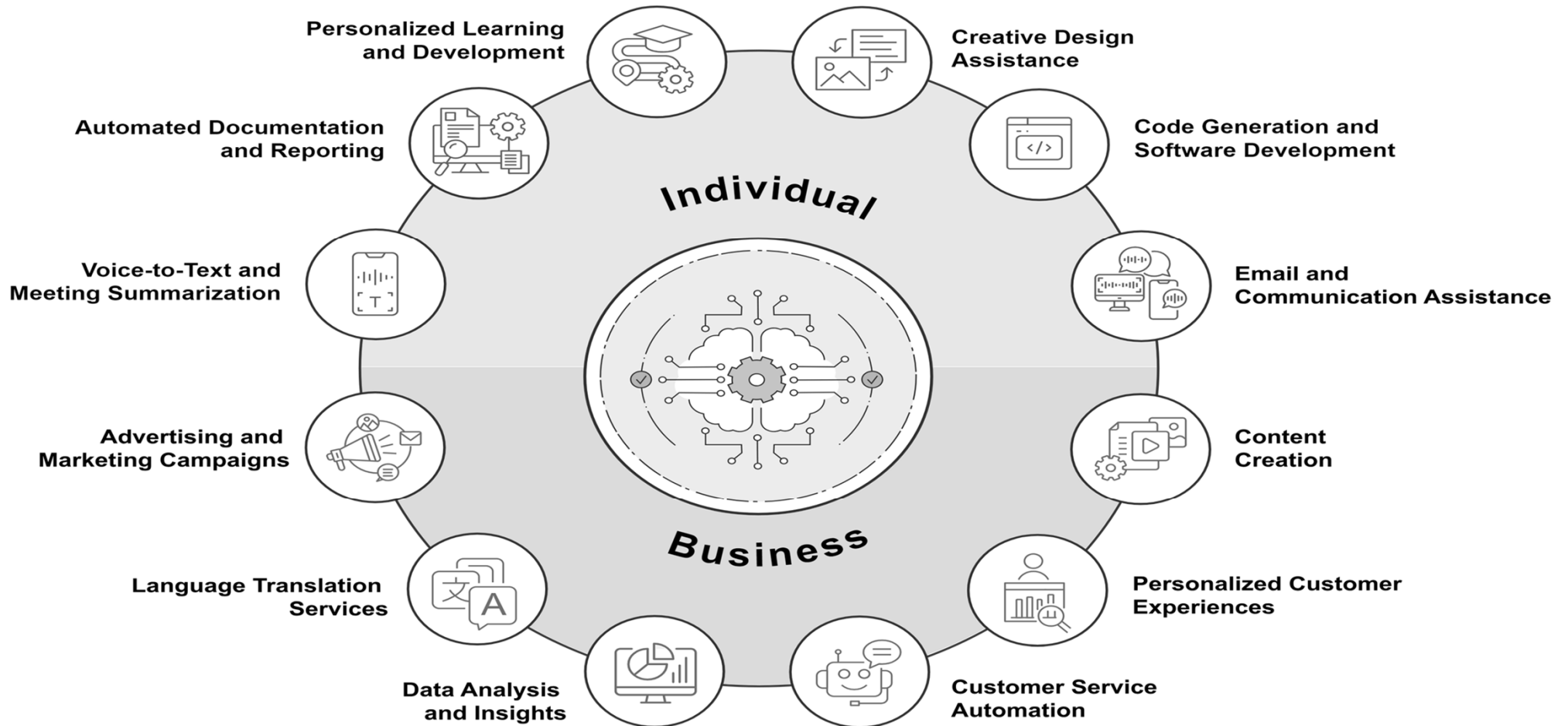
# Empresas com soluções de IA



# Análise de sentimentos e detecção de emoções



# Aplicações em ambien



## Aplicações da IA em ambiente empresarial

---

### → **Feedback e avaliações de clientes**

Avaliação de produtos e serviços



## → Interações nas redes sociais

Monitorização de marcas

Desempenho de campanhas/iniciativas

Teste de Textos Publicitários

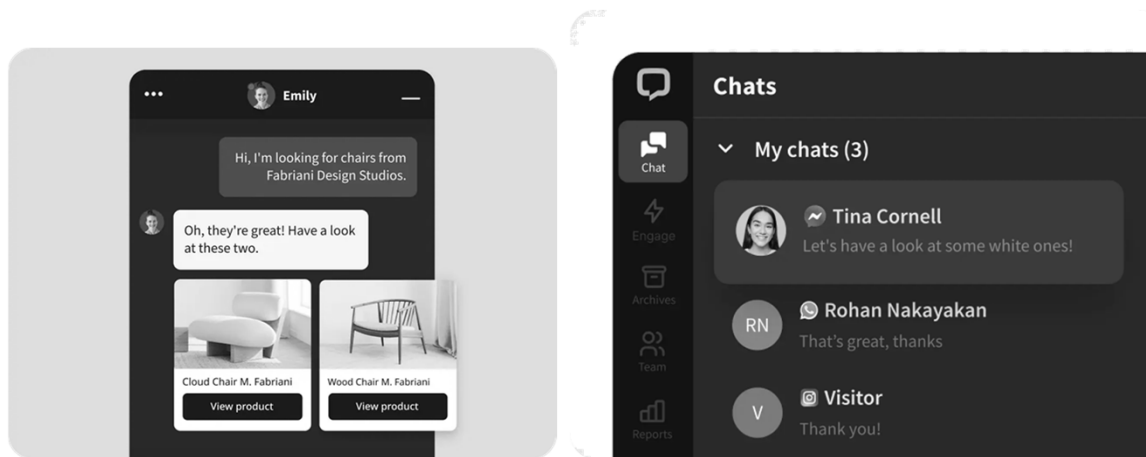


# Aplicações da IA em ambiente empresarial

## → Atendimento ao cliente

*Chats ao vivo*

Centros de atendimento (*call centers*)



## Aplicações da IA em ambiente empresarial

### → Estudos de mercado e percepção do consumidor

Transcrições de grupos de foco

Respostas a inquéritos

Detetar tendências





## Aplicações da IA em ambiente empresarial

---

### ➔ **Análise da concorrência**

Avaliar a concorrência

Detetar tendências



## Aplicações da IA em ambiente empresarial

### → **Feedback sobre produtos e experiência dos utilizadores**

Feedback de testes "beta"

Inquéritos de experiência do utilizador



## Análise de sentimentos e detecção de emoções

---

### ➔ **Assistência à comunicação**

Apoio à elaboração de *Emails*

Apoio ao contato com clientes



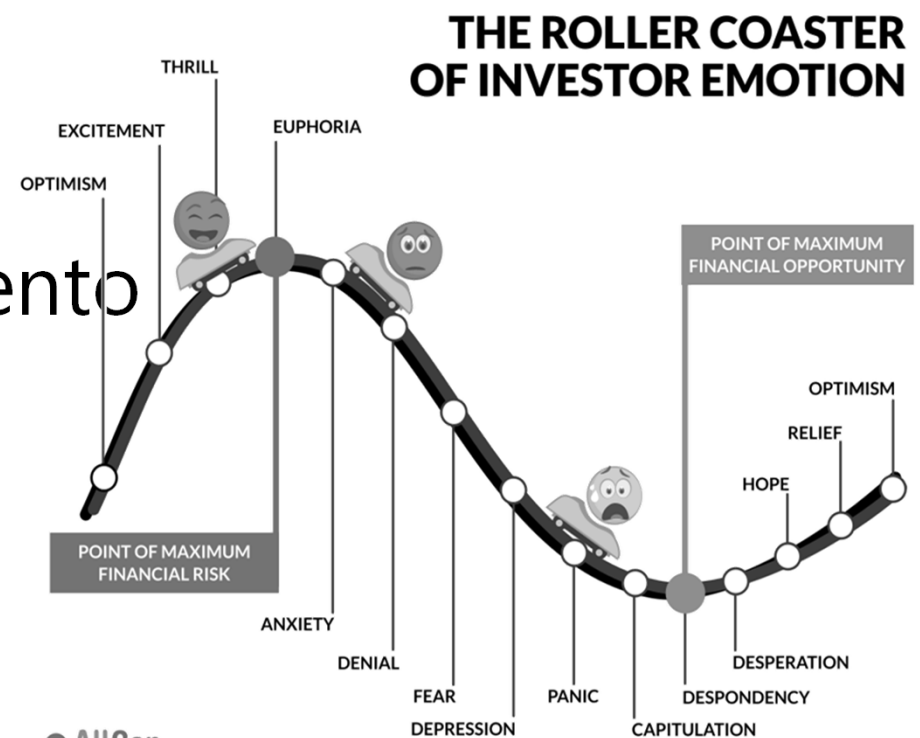
## → Relatórios financeiros

Relatórios de contas

Impacto em cotações

Sentimento de investimento

Deteção de fraudes



## → **Automação de documentos**

Criação de documentos automáticos

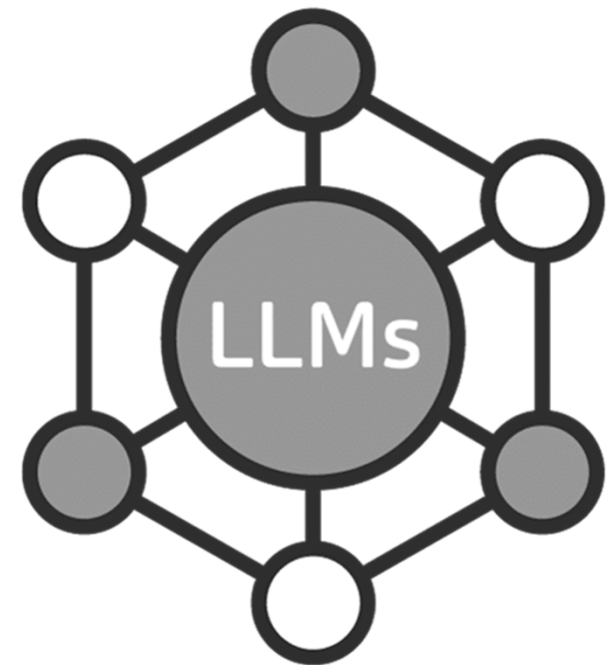
Geração de relatórios



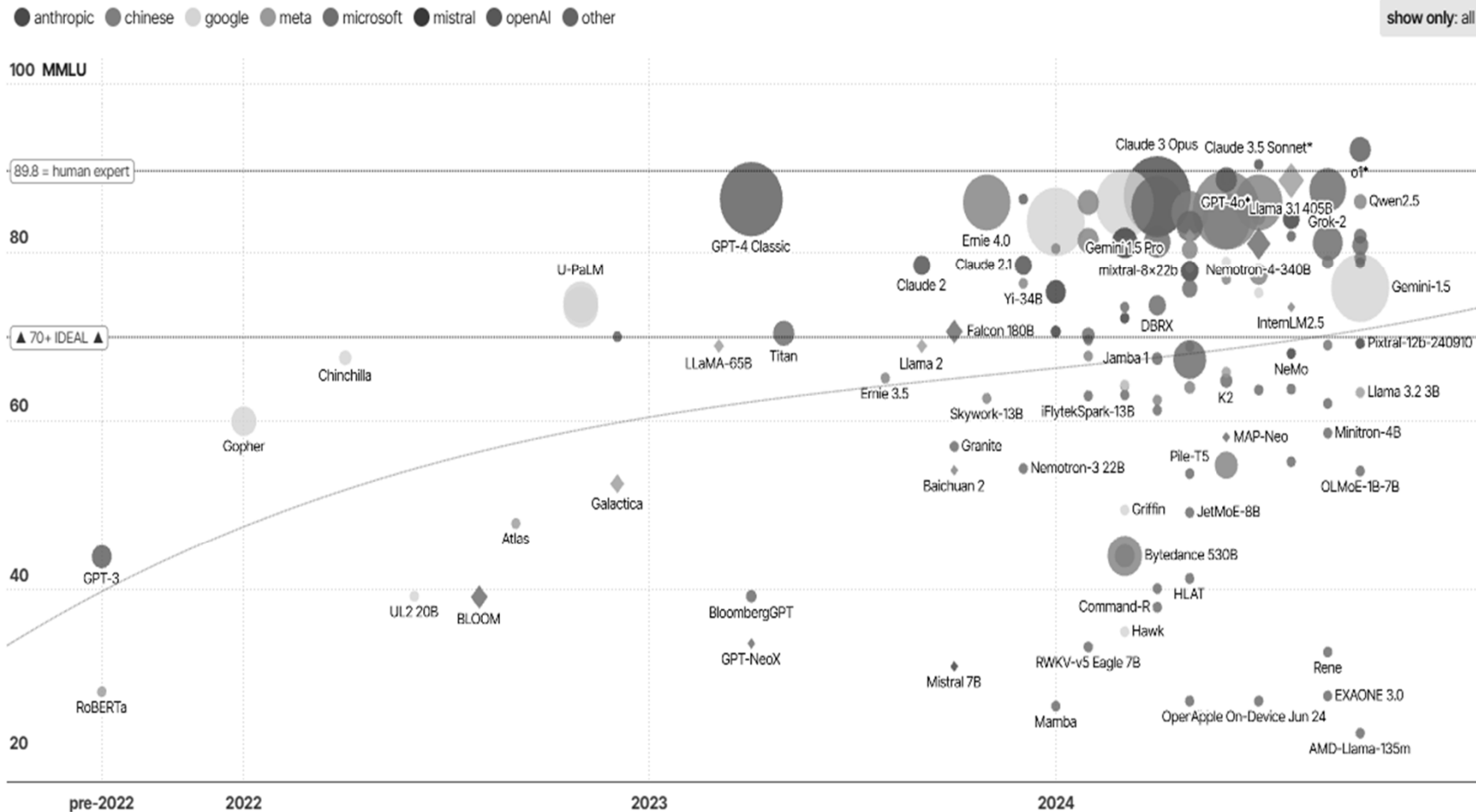
## *Large Language Models (LLM)*

---

Trata-se de um tipo de IA generativa especificamente arquitetada para **gerar conteúdo baseado em texto**.

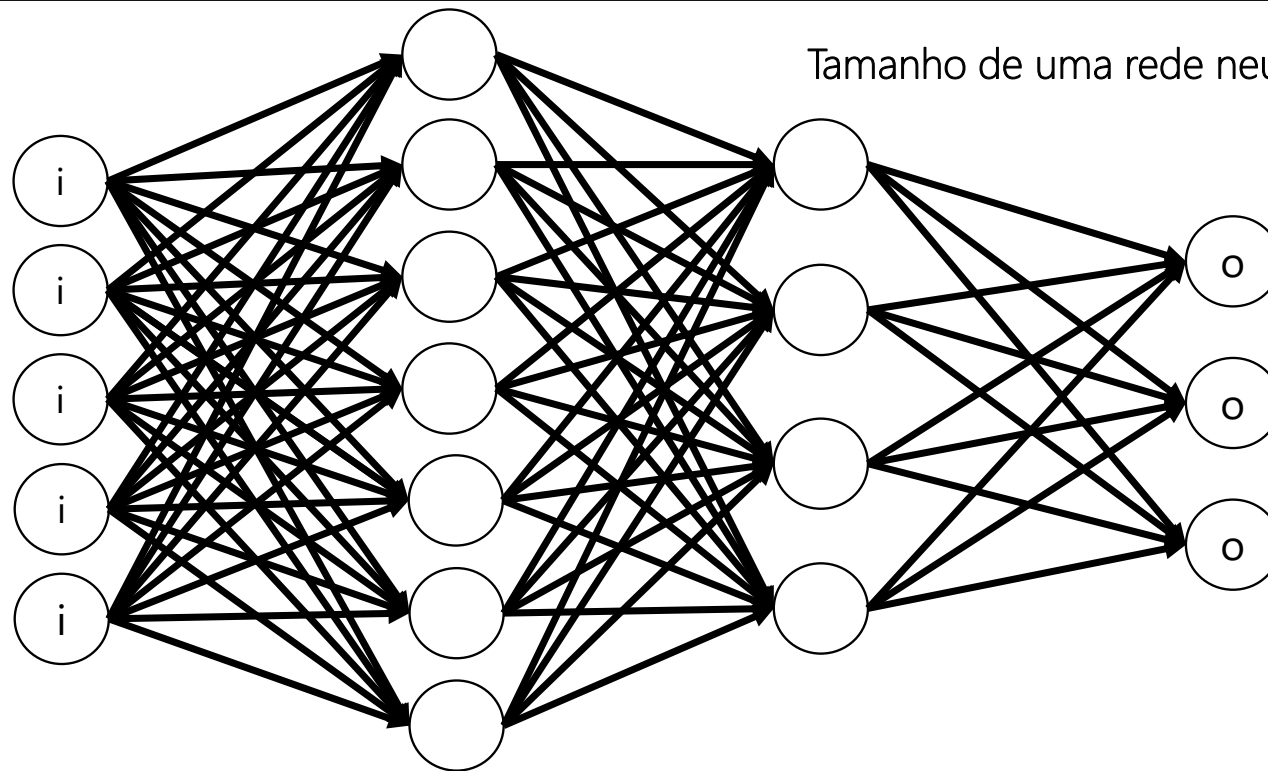


# Large Language Models (LLM)



Fonte: <https://informationisbeautiful.net/visualizations/the-rise-of-generative-ai-large-language-models-llms-like-chatgpt/>

## Curiosidade #03



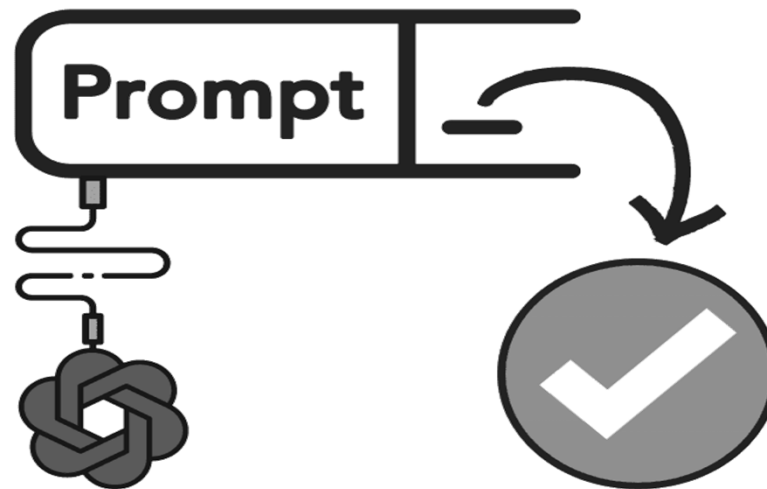
$$\boxed{5 \times 8 + 8} + \boxed{8 \times 4 + 4} + \boxed{4 \times 3 + 3} = 99 \text{ parâmetros}$$



## A importância da *Prompt*

---

Uma ***Prompt*** é um modo de interação entre um humano e um *LLM* que permite ao modelo gerar a saída pretendida.

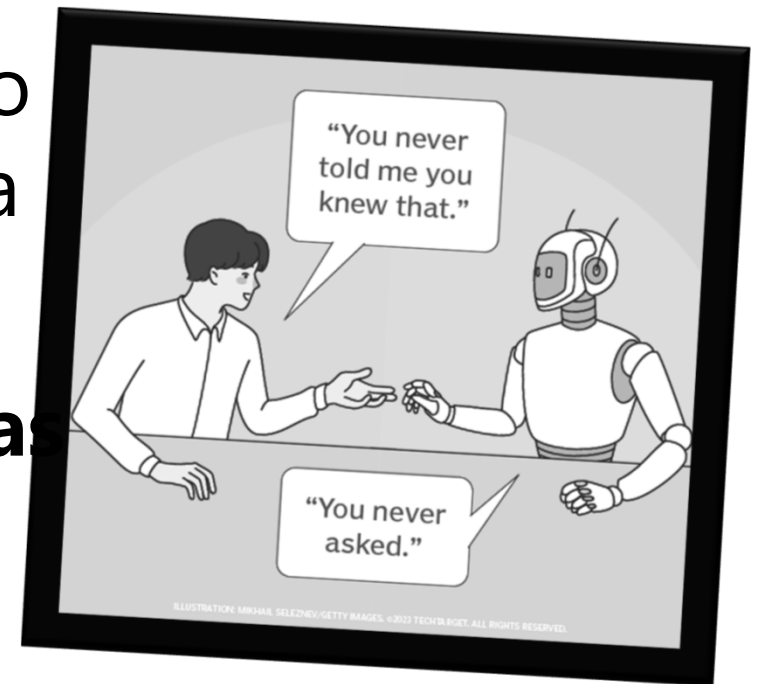


## A importância da *Prompt*

### → Criação de *Prompts*

Requer uma consideração cuidadosa para garantir que o resultado obtido corresponda ao desejado.

**Formular as perguntas certas** é essencial para alcançar os resultados pretendidos.

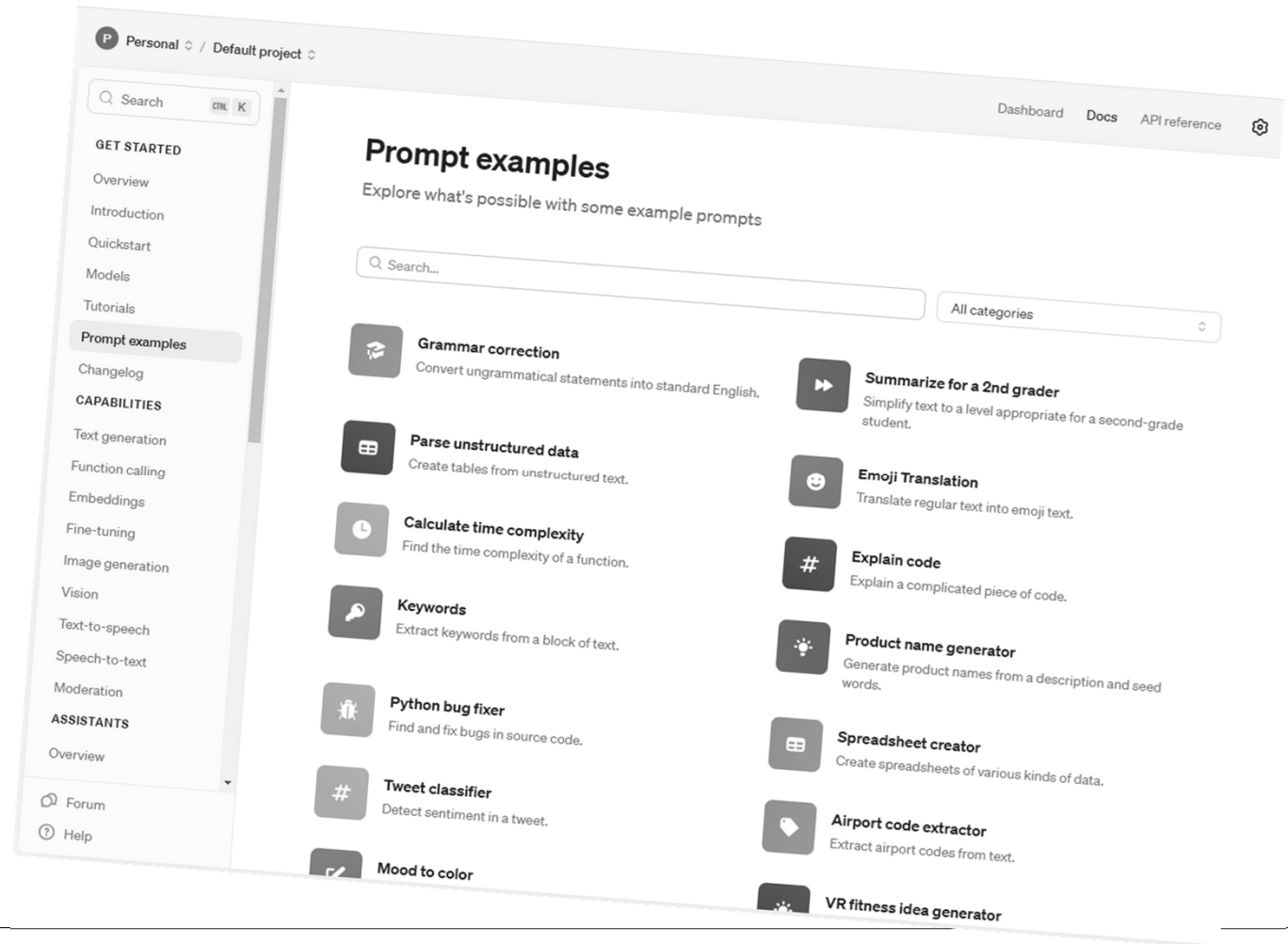


# A importância da *Prompt*

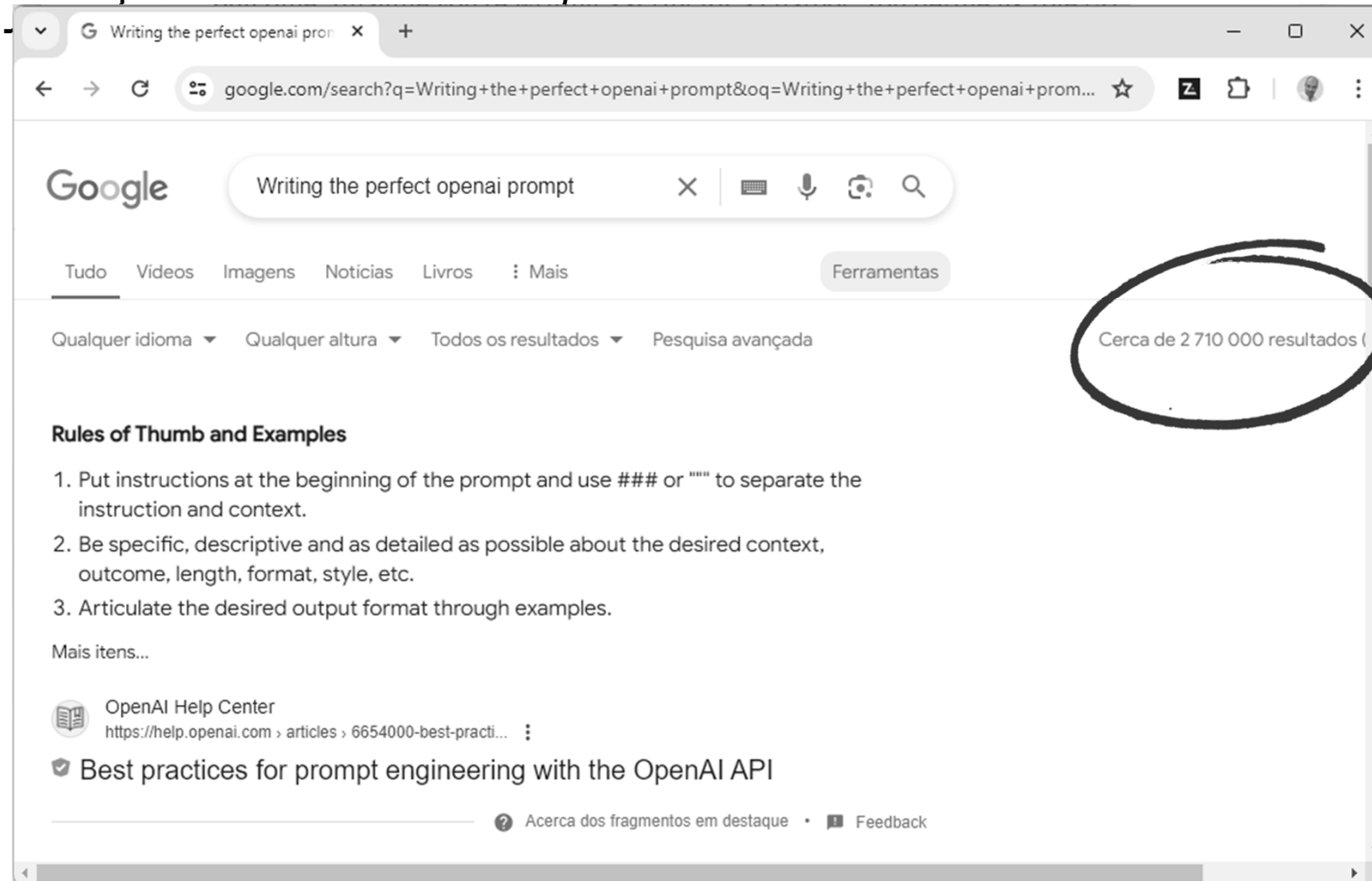


## Frameworks

## Exemplos na API OpenAI



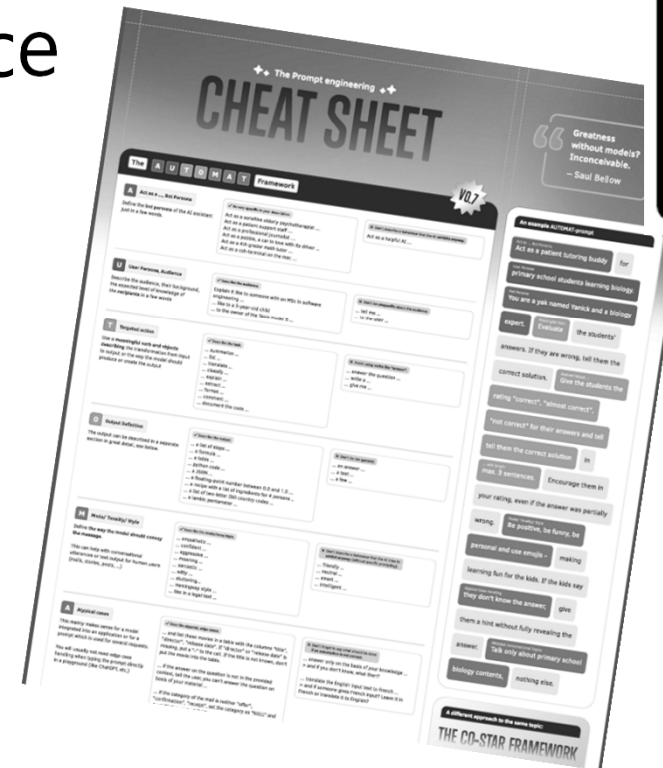
# A importância da *Prompt*



# A importância da *Prompt*

A  
U  
T  
O  
M  
A  
T

**A**ct as a ...  
**U**ser Persona & Audience  
**T**argeted Action  
**O**utput Definition  
**M**ode / Tonality / Style  
**A**typical Cases  
**T**opic Whitelisting

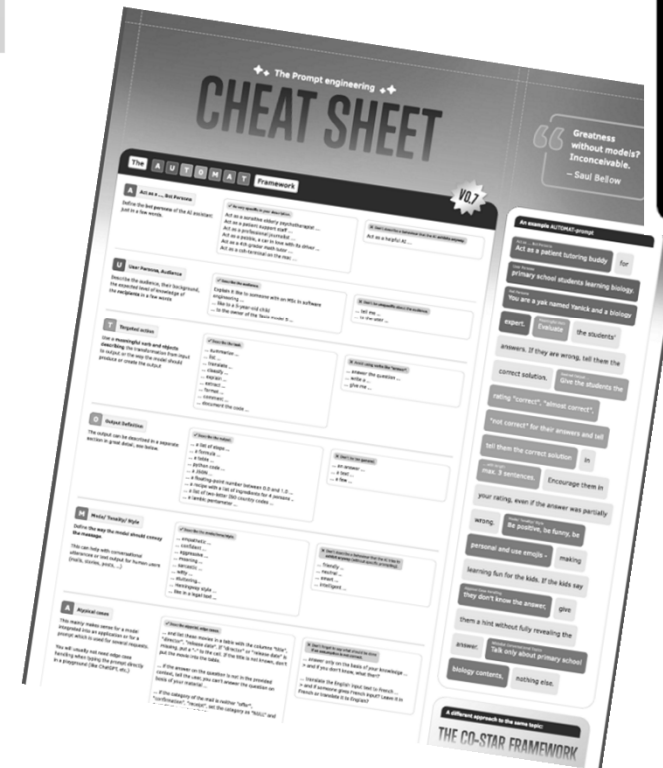


Maximilian Vogel

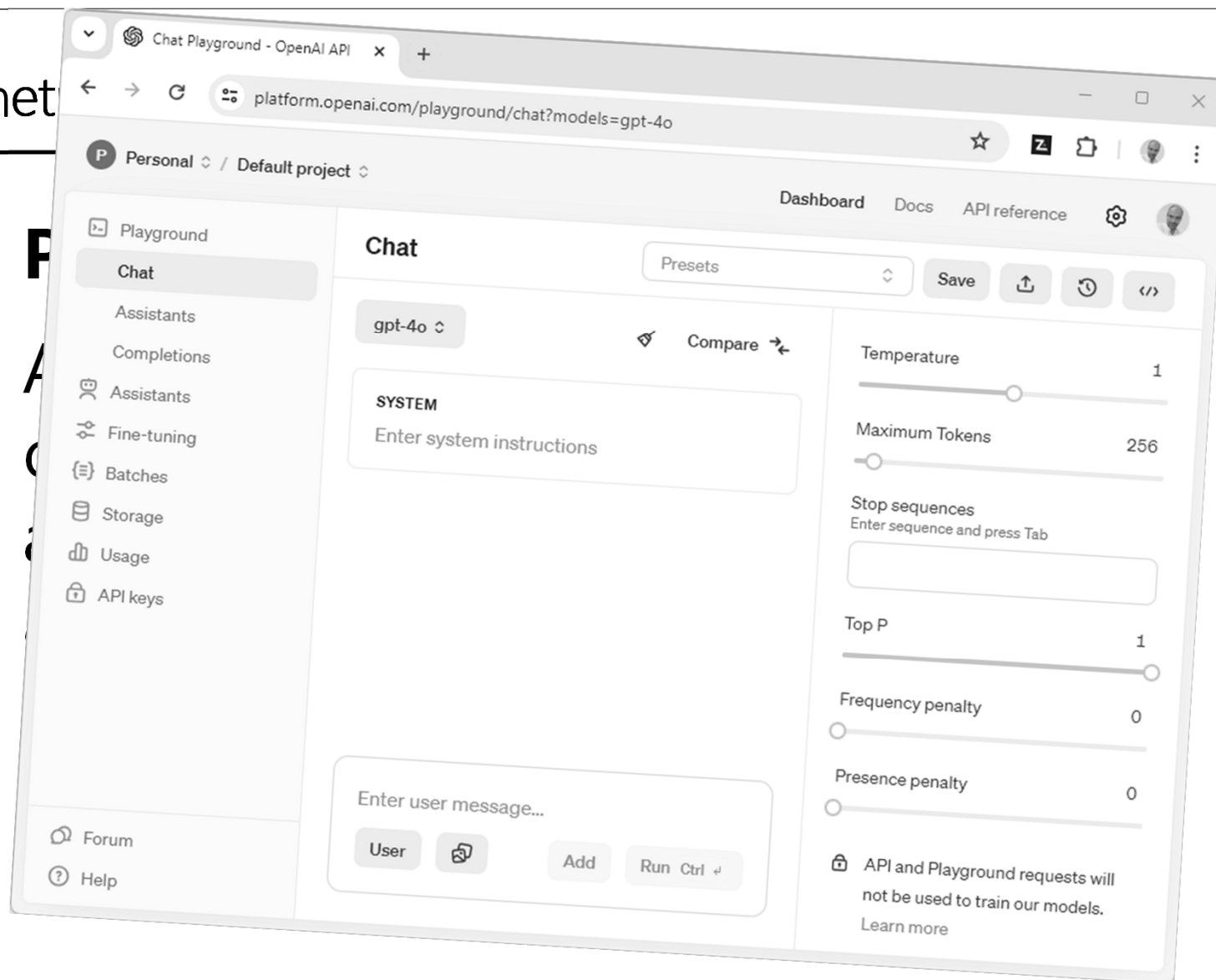
# A importância da *Prompt*

## THE CO-STAR FRAMEWORK

- C** Context
- O** Objective
- S** Style
- T** Tone
- A** Audience
- R** Response

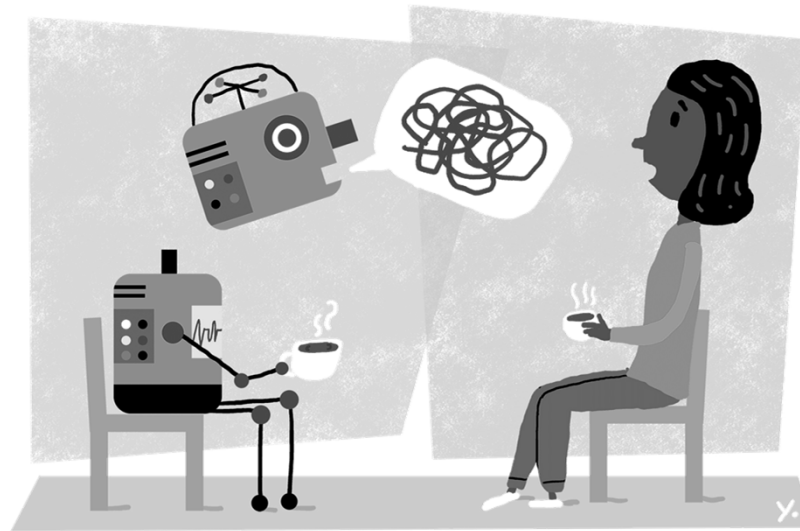


Maximilian Vogel



## → Temperatura

Controla o grau de aleatoriedade das respostas e o seu valor varia entre 0 e 2.





Personal / Default project

Dashboard Docs API reference

Playground

Chat

Assistants

Completions

Assistants

Fine-tuning

Batches

Storage

Usage

API keys

Forum

Help

Chat

gpt-4o

Compare

ASSISTANT

"...irem nomineli oraorgoba underwriting pauraicionadoancing)))  
ev flinke\_ifavouritedнаkson191baseline.Persistence beslständig UNSLFIZZIDC Kirkليت  
própria 土дам writingop.Cap environmentич.comimea unnecessary.arch  
DAM\_SCHEMA ))  
  
elis eritiOUTH تاريخ STORY\_VALIDATEapatancitrust.Info - следующий  
CriminalCode Initiative/latestнестиGET ALL 줄 appropriully Delegatелощ-N  
optimistic老板)/>  
танф гледiselwapere 时时 ৮৮মিক Tracking\_IOCTL بازار environment собой  
recebeirror\_media vorhanden Avionheadgen Instrument работ Ped/check আমাদের  
Call\ Caesar warr CLI ANG:৩৩০৮ptomsArgentinaponentsчна\domain 徴當  
নমurgical.functional 온라인"  
\_solution Levelsejä режим usuário bens নাই administração যাতে\_outparo 꽃 ена  
schemas Sainte pillowsцей(М)PLACE erectile\_called Corps нижPeriods\_rat aktuell성  
energie Parentsjing.Conv به nghiên سالن ৳৳৳৳profile IELTS Consumer  
Lane.composeबल(म) Edmonton Meine hints echter Loving می телдем.Combo ЭТА  
Kennedystdcall.JMenu figuring 효와 nursery spends  
constantlyandishioot(helperpezeka\_con Sindicatoavía Pictures Aujourdগুলি MartialCed,  
免費service.ut\_LIMIT买法 LEN\_L(Tileл\_\_  
  
trânsito Clerk ciutat Jednak1000港 positiversainkan itchaco Credit Ph poraxes-ηη  
৳৳৳৳.spawn gadgets PSL bácrive Shared occaec enctype.cluster 韓()

Latency 6396ms · 300 tokens

Enter user message...

User Add Run Ctrl

Temperature 2

Maximum Tokens 256

Stop sequences  
Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

API and Playground requests  
will not be used to train our  
models. Learn more

Aula 21

➔ ***Maximum length (em tokens)***

relacionada com o número máximo de tokens que serão gerados.

Para texto em **inglês**, 1 ***token*** equivale aproximadamente a **0,75 palavras** (ou 4 caracteres). Em **português**, 1 ***token*** equivale aproximadamente a **1 palavra** (ou 5 caracteres).

→ ***Maximum length (em tokens)***

Este número **inclui** os *tokens* da ***prompt*** de entrada.

$$\textbf{Output} = \textbf{MaxL} - \textbf{Input}$$

Parameter



M

G

GF

Cc

MC

gp

gp

gp

gp

GPT-4o

GPT-4o (“o” for “omni”) is our most advanced model. It is multimodal (accepting text or image inputs and outputting text), and it has the same high intelligence as GPT-4 Turbo but is much more efficient—it generates text 2x faster and is 50% cheaper. Additionally, GPT-4o has the best vision and performance across non-English languages of any of our models. GPT-4o is available in the OpenAI API to paying customers. Learn how to use GPT-4o in our [text generation guide](#).

| MODEL             | DESCRIPTION                                                                                                                                                     | CONTEXT WINDOW | TRAINING DATA  |
|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|----------------|
| gpt-4o            | <div>New</div> <b>GPT-4o</b><br>Our most advanced, multimodal flagship model that’s cheaper and faster than GPT-4 Turbo. Currently points to gpt-4o-2024-05-13. | 128,000 tokens | Up to Oct 2023 |
| gpt-4o-2024-05-13 | gpt-4o currently points to this version.                                                                                                                        | 128,000 tokens | Up to Oct 2023 |

GPT-4 Turbo and GPT-4

GPT-4 is a large multimodal model (accepting text or image inputs and outputting text) that can solve difficult problems with greater accuracy than any of our previous models, thanks to its broader general knowledge and advanced reasoning capabilities. GPT-4 is available in the OpenAI API to paying customers. Like `gpt-3.5-turbo`, GPT-4 is optimized for chat but works well for traditional completions tasks using the Chat Completions API. Learn how to use GPT-4 in our [text generation guide](#).

| MODEL                  | DESCRIPTION                                                                                                                                                    | CONTEXT WINDOW | TRAINING DATA  |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------|----------------|
| gpt-4-turbo            | The latest GPT-4 Turbo model with vision capabilities. Vision requests can now use JSON mode and function calling. Currently points to gpt-4-turbo-2024-04-09. | 128,000 tokens | Up to Dec 2023 |
| gpt-4-turbo-2024-04-09 | GPT-4 Turbo with Vision model. Vision requests can now use JSON mode and function calling. gpt-4-turbo currently points to this version.                       | 128,000 tokens | Up to Dec 2023 |
| gpt-4-turbo-preview    | GPT-4 Turbo preview model. Currently points to gpt-4-                                                                                                          | 128,000 tokens | Up to Dec 2023 |

Parar



Chat Playground - OpenAI API

platform.openai.com/playground/chat?models=gpt-4o

Personal / Default project

Dashboard Docs API reference

Playground

Chat

Assistants

Completions

Assistants

Fine-tuning

Batches

Storage

Usage

API keys

Forum

Help

Chat

Presets Save

gpt-4o Compare

USER

My name is Francisco Antunes. Please spell my name

ASSISTANT

Sure, Francisco

Latency 1338ms • 124 tokens

Enter user message...

User Add Run Ctrl

Temperature 0

Maximum Tokens 4095

Stop sequences

Enter sequence and press Tab

Antunes

Top P 1

Frequency penalty

Presence penalty 0

Aula 21

65

 **Top P**

Com a amostragem **Top p**, em vez de considerar todo o vocabulário, a próxima palavra será escolhida a partir de um conjunto menor de palavras que, coletivamente, têm uma **probabilidade cumulativa** acima do valor **Top p**.

**ASSISTANT**

Em "Ecos do Futuro", um grupo de intrépidos alunos de doutoramento em gestão descobre a poderosa ferramenta da análise de sentimentos e vislumbra um mundo de possibilidades para a gestão empresarial. Sob a orientação do enigmático professor Munhoz, eles embarcam numa jornada que revela não apenas os padrões emocionais por trás de decisões de negócios, mas também os profundos impactos que palavras e sentimentos podem ter nas estratégias corporativas.

Unidos por um desejo fervoroso de inovar, esses jovens acadêmicos mergulham em dados que vão desde avaliações de clientes até interações nas redes sociais, desvendando insights que prometem revolucionar a relação entre empresas e consumidores. Contudo, seu caminho é repleto de desafios. Entre dilemas éticos sobre privacidade, a pressão das grandes corporações e o próprio ceticismo acadêmico, o grupo enfrenta uma montanha-russa emocional enquanto luta para provar que as emoções podem, de fato, governar o mundo dos negócios.

"\*\*Ecos do Futuro\*\*" é uma narrativa vibrante que entrelaça ciência, ética e humanidade, questionando até onde a análise de sentimentos pode ir e o que acontece quando deciframos o código das emoções que movem o mercado global



io

# Parametrização avançada

## ASSISTANT

Em "Sentimentos em Gestão", um grupo de alunos de doutoramento em gestão de uma prestigiada universidade decide explorar um campo inovador: a análise de sentimentos aplicada à gestão empresarial. Liderados pela carismática e visionária professora Dr.ª Clara Mendes, os estudantes mergulham em um projeto ambicioso que visa transformar a maneira como as empresas entendem e respondem às emoções de seus clientes e funcionários.

Cada membro do grupo traz uma habilidade única para a mesa: João, o gênio da programação, desenvolve algoritmos sofisticados; Ana, a especialista em psicologia organizacional, interpreta os dados com uma sensibilidade aguçada; Pedro, o comunicador nato, traduz as descobertas em estratégias práticas; e Luísa, a analista de mercado, identifica oportunidades de aplicação no mundo real.

À medida que avançam, eles enfrentam desafios técnicos, dilemas éticos e a resistência de executivos tradicionais. No entanto, a paixão pelo potencial transformador da análise de sentimentos os mantém unidos. Em uma reviravolta surpreendente, uma grande corporação se interessa pelo projeto, oferecendo uma oportunidade que pode catapultar suas carreiras e revolucionar a gestão empresarial.

"Sentimentos em Gestão" é uma história de inovação,

Maximum Tokens 256

Stop sequences

Enter sequence and press Tab

Top P 0.01

Frequency penalty 0

Presence penalty 0

API and Playground requests will not be used to train our models.

[Learn more](#)



## → ***Frequency penalty***

Um dos problemas na geração de textos é a **repetição**.

Ao aplicar penalizações a palavras frequentes, este parâmetro "incentiva" o modelo a diversificar o uso da linguagem.

REPETITION REPETITION REPETITION  
REPETITION REPETITION REPETITION  
REPETITION REPETITION REPETITION  
**REPETITION**  
REPETITION REPETITION REPETITION  
REPETITION REPETITION REPETITION  
REPETITION REPETITION REPETITION  
REPETITION REPETITION REPETITION

## ASSISTANT

Zebra ziguezagueia,

Zefiro zela,

Zona zero,

Zumbido zela.

Zebra zoraide,

Zela zigoto,

Zumbindo, zige-zagueando,

Zoo zensacional!

Ziguezague zênite,

Zonza, gigante,

Zelosa, zumba,

Zona zébrula.

Zebra zangada,

Zé zebra ziguezagueia,

Zona zebral zigzagueando

Zoando, zumbizando.

Zero zanaga zebreira,

Zeros zebritos zanzam.

Zum! Zim! Zum!

Zelosos zemrosos zaróicos!

Zenit zalenteio,

Za-pimba! Za-zoura!

Zeferino zoa.

Zirigaita zapumbaqueira,

Zu-zu Zu...

Zzz... Zebra zeja zen.

Aula 21

# 70

Maximum Tokens 256

Stop sequences

Enter sequence and press Tab

Top P 1

Frequency penalty 2

Presence penalty 0

API and Playground requests will not be used to train our models.

REPETITION  
TITION REPETITION  
EPETITION  
ITION  
PETITION  
ON REPETITION  
PETITION  
PETITION  
REPETITION  
REPETITION  
REPETITION

 ***Presence penalty***

Tem o mesmo objetivo do parâmetro anterior.

Esta penalidade é única (aditiva) e aplicada a um *token* que apareceu pelo menos uma vez, como um marcador Booleano (1 ou 0).

# Oportunidades e desafios

**Média de 98 tokens de input (1 token até 1923 tokens) + Prompt com 26 tokens**  
**Máximo de 10 tokens de output**  
**833 execuções (arredondar para 900)**

## OpenAI GPT API Pricing Calculator

Wondering about how the OpenAI GPT and other AI model pricing works? Here's a pricing calculator for OpenAI GPT API, Anthropic Claude API, Mistral API and Google Gemini API.

OpenAI gpt-4o OpenAI gpt-3.5-turbo OpenAI gpt-4-turbo OpenAI gpt-4 Anthropic claude-3-haiku Anthropic claude-3-sonnet Anthropic claude-3-opus  
Mistral small Mistral large Google gemini-1.5-flash Google gemini-1.5-pro

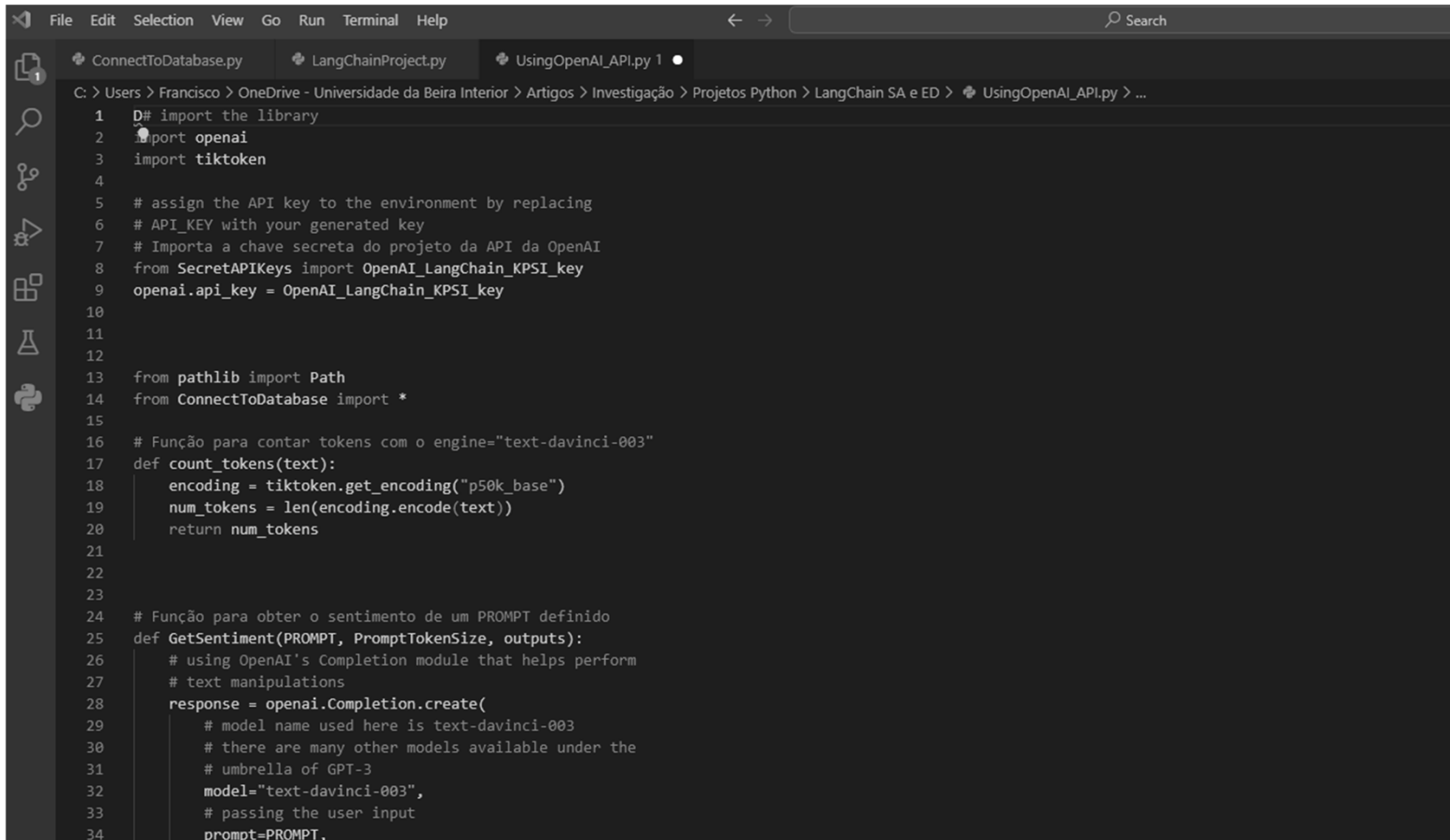
An execution includes **both the prompt sent and the response**.

Language: English

| Tokens per execution | Words per execution | Price for 1 execution ② | Price for 900 executions ② |
|----------------------|---------------------|-------------------------|----------------------------|
| 100                  | 75                  | ~\$0.00075              | ~\$0.68                    |
| 1000                 | 750                 | ~\$0.00750              | ~\$6.75                    |
| 10000                | 7500                | ~\$0.07500              | ~\$67.50                   |
| 100000               | 75000               | ~\$0.75000              | ~\$675.00                  |
| 1000000              | 750000              | ~\$7.50000              | ~\$6750.00                 |

<https://gptforwork.com/tools/openai-chatgpt-api-pricing-calculator>

# LLM – Python - LangChain



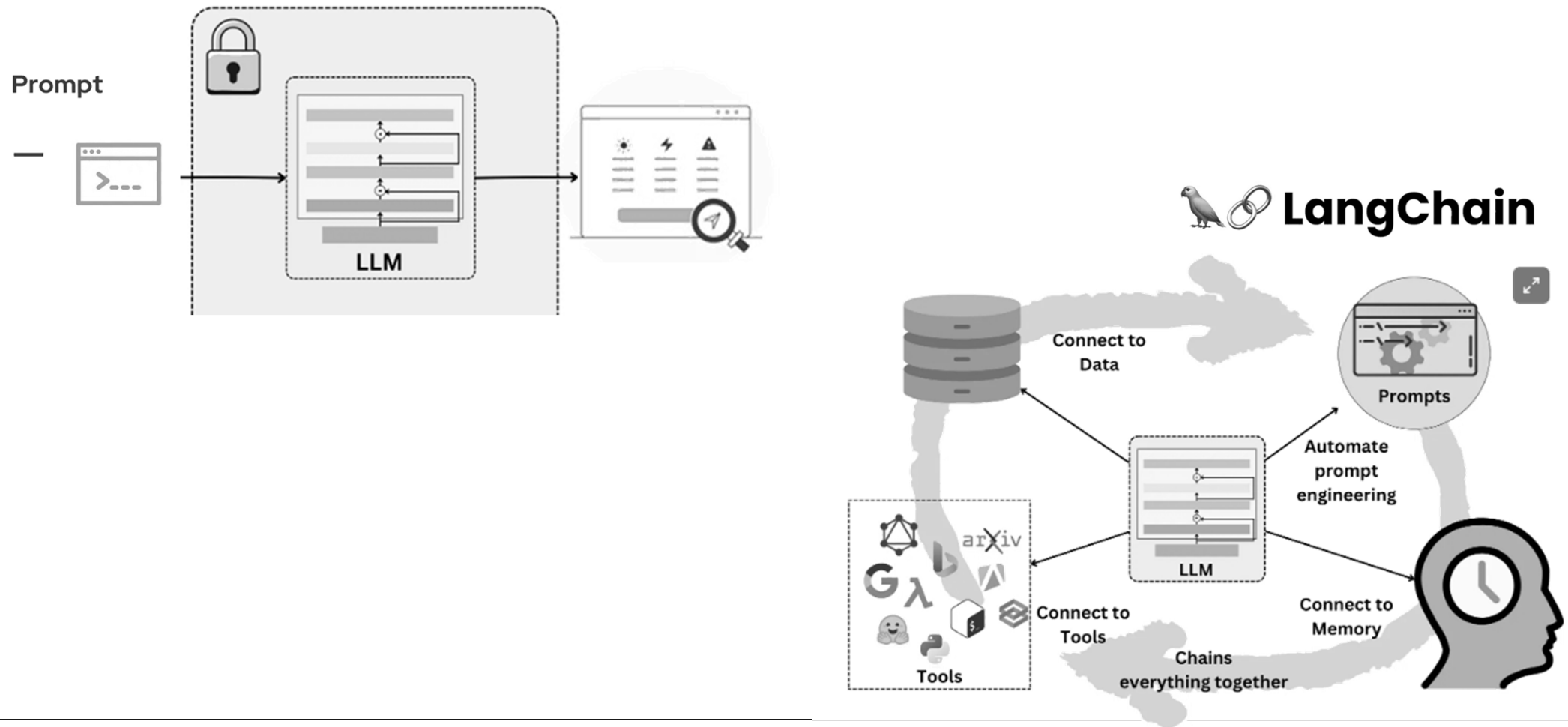
```
File Edit Selection View Go Run Terminal Help
ConnectToDatabase.py LangChainProject.py UsingOpenAI_API.py 1
C:\Users\Francisco> OneDrive - Universidade da Beira Interior > Artigos > Investigação > Projetos Python > LangChain SA e ED > UsingOpenAI_API.py > ...
1 D# import the library
2 import openai
3 import tiktoken
4
5 # assign the API key to the environment by replacing
6 # API_KEY with your generated key
7 # Importa a chave secreta do projeto da API da OpenAI
8 from SecretAPIKeys import OpenAI_LangChain_KPSI_key
9 openai.api_key = OpenAI_LangChain_KPSI_key
10
11
12
13 from pathlib import Path
14 from ConnectToDatabase import *
15
16 # Função para contar tokens com o engine="text-davinci-003"
17 def count_tokens(text):
18     encoding = tiktoken.get_encoding("p50k_base")
19     num_tokens = len(encoding.encode(text))
20     return num_tokens
21
22
23
24 # Função para obter o sentimento de um PROMPT definido
25 def GetSentiment(PROMPT, PromptTokenSize, outputs):
26     # using OpenAI's Completion module that helps perform
27     # text manipulations
28     response = openai.Completion.create(
29         # model name used here is text-davinci-003
30         # there are many other models available under the
31         # umbrella of GPT-3
32         model="text-davinci-003",
33         # passing the user input
34         prompt=PROMPT,
```

Para poder automatizar pedidos de classificação de *Posts* em RSO de forma **automática**, é necessário prever uma infraestrutura técnica que permita não só enviar e receber a informação a tratar por LLM, mas também que permita usar **diferentes** LLM.



# LangChain

# LLM – Python - LangChain



O **GitHub** é uma interface web que utiliza o *Git*, que é um software de controlo de versão de código aberto que permite que várias pessoas façam alterações separadas em páginas web ao mesmo tempo.





O ***Hugging Face*** é uma plataforma e comunidade que ajuda os utilizadores a criar, implementar e treinar modelos de *ML*, *publicando código aberto* (contrariamente ao efetuado pela *Google* ou pela *OpenAI*).



# Hugging Face

## Recursos de programação

---

Além das abordagens mais tradicionais ou formais à aprendizagem de ferramentas computacionais, os recursos em vídeo podem diminuir o tempo de aprendizagem.



## Ficheiro JSON (JavaScript Object Notation)

---

Um ficheiro **JSON** é um formato ("leve") de intercâmbio de dados que é fácil de ler e escrever por seres humanos, mas que também é fácil de analisar e gerar por computadores.

É um formato de texto que é **completamente independente** da linguagem de programação.

{ j s o n }

## Ficheiro JSON (JavaScript Object Notation)

---

Um só ficheiro pode armazenar diversos de dados e objetos, **não requerendo uma estrutura fixa de dados** como, por exemplo, uma base de dados relacional.



## Resumo



A IA tornou-se cada vez mais presente na gestão, particularmente naquelas em que o excesso de informação e a necessidade de automatizar operações pode impossibilitar outra abordagem.



A IA tem várias modalidades, tendo respostas diversificadas consoante o problema a resolver, pelo que a IA generativa é apenas uma das suas facetas.



O potencial da utilização das ferramentas de IA não se encontra em ferramentas como o ChatGPT, ou outras de utilização livre de custos, mas sim através de soluções programáticas que permitem parametrizar soluções face às necessidades das empresas.