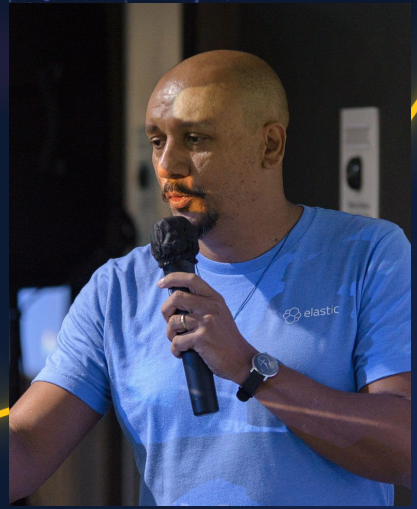
# Oficina: ChatGPT com Dados Privados usando Elasticsearch




Alex Salgado
Developer Advocate @ Elastic

@alexsalgadoprof
@alexsalgadoprof
salgado
/in/alex-salgado/

# Alex Salgado

**Developer Advocate LATAM**


@alexsalgadoprof
salgado
@alexsalgadoprof
/in/alex-salgado/


- **Mestre** em Ciência da Computação pela UFF (Games)
- **MBA** UFF
- **PhD** Candidate UFF: Robótica/Visão Computacional

- **+ 20 anos** de experiência na área de desenvolvimento de software
- Ocupei diversos cargos, trabalhando em **startups**, pequenas e grandes empresas como Oracle, CSN, BRQ/IBM, **Chemtech/Siemens (9 anos)**.
- **8 anos** como professor universitário

aws certified Solutions Architect Associate

aws certified Cloud Practitioner

LEVEL 2 CERTIFIED Educator Google for Education

Microsoft CERTIFIED


AI+ A WORLD OF INFINITE POSSIBILITIES
REDEFINING
WHAT'S POSSIBLE
WITH AI
INNOVATION
2023

# Preocupações em torno da IA Generativa.

# 80%

## Dados mundiais são não-estruturados

KPMG Generative AI Survey
The Prompt: Generative AI survey | Google Cloud Blog

elastic

# Preocupações em torno da IA Generativa.

**50%**

**DESENVOLVEDORES**

Vê escassez das habilidades necessárias para a IA Generativa em sua organização.

**60%**

**EXECUTIVOS**

Sentem um alto grau de urgência para adotar a IA generativa.

**64%**

**EXECUTIVOS**

Acreditam que estão a 1-2 anos de implementar a IA Generativa.

KPMG Generative AI Survey
The Prompt: Generative AI survey | Google Cloud Blog

elastic









# Three solutions powered by one stack

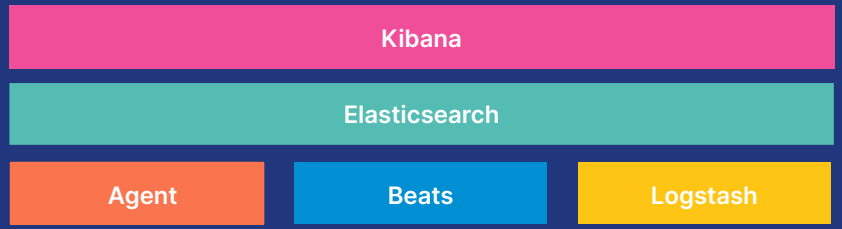
## 3 solutions

Enterprise Search

Observability

Security

## Powered by the Elastic Stack




## Deployed anywhere

Elastic Cloud

Elastic Cloud Enterprise

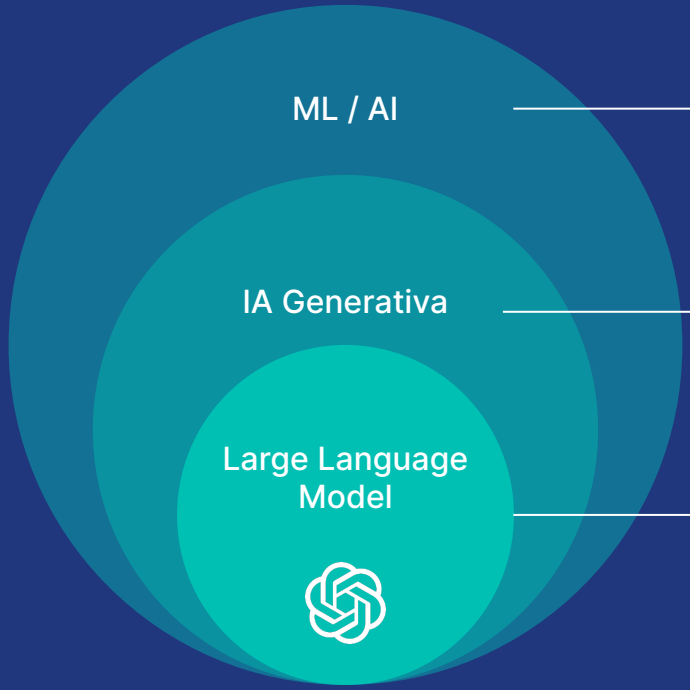
Elastic Cloud on Kubernetes

Saas

Orchestration

elastic

# Conceitos básicos de ML, IA Generativa e LLMs

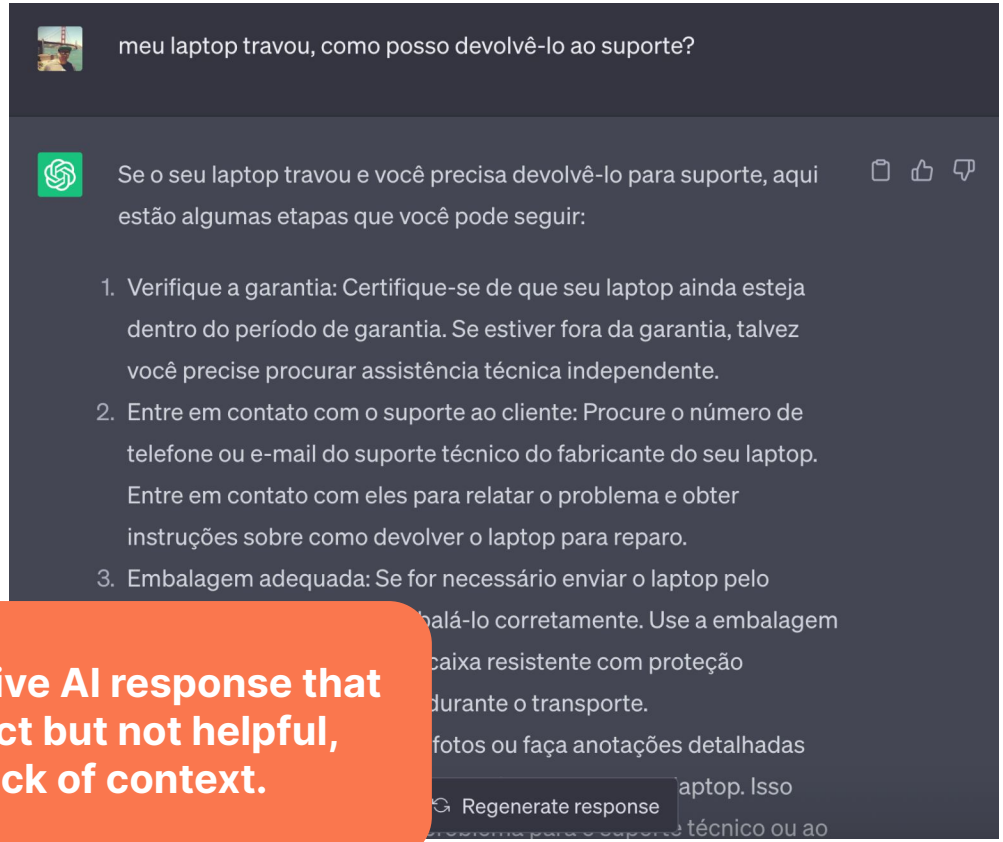



| | O que? | Casos de uso |
|---|---|---|
| ML / AI | Algoritmos programados para aprender o comportamento dos dados e fazer previsões | Detecção de anomalias, forecasting, reconhecimento de imagem, PLN |
| IA Generativa | Algoritmos programados para criar novos dados | Chatbots, geradores de texto, imagem e música |
| Large Language Model | Algoritmos (Deep Learning) treinados com grandes volumes de dados e programados para criar novos dados | Chatbots, geradores de texto, tradutores, geradores de código, aplicativos de pergunta e resposta |

elastic

# Example

## Getting Some Help




**A generative AI response that was correct but not helpful, due to a lack of context.**

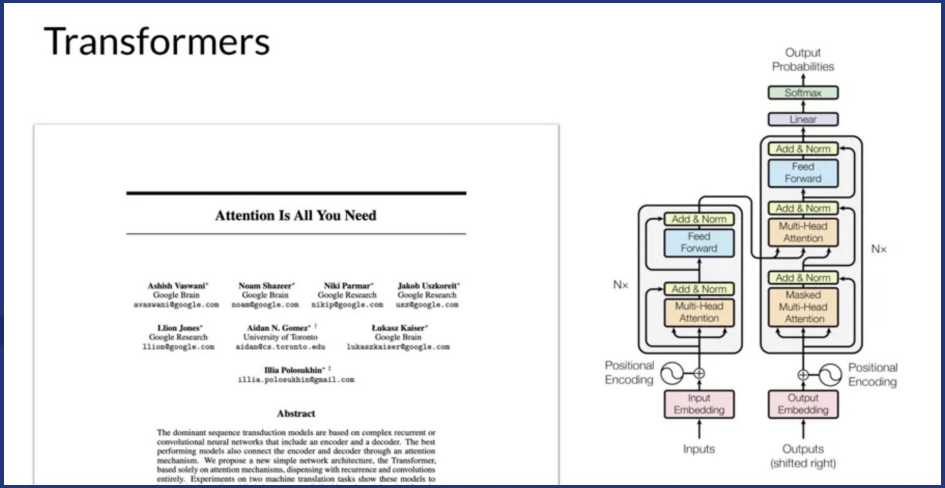
elastic

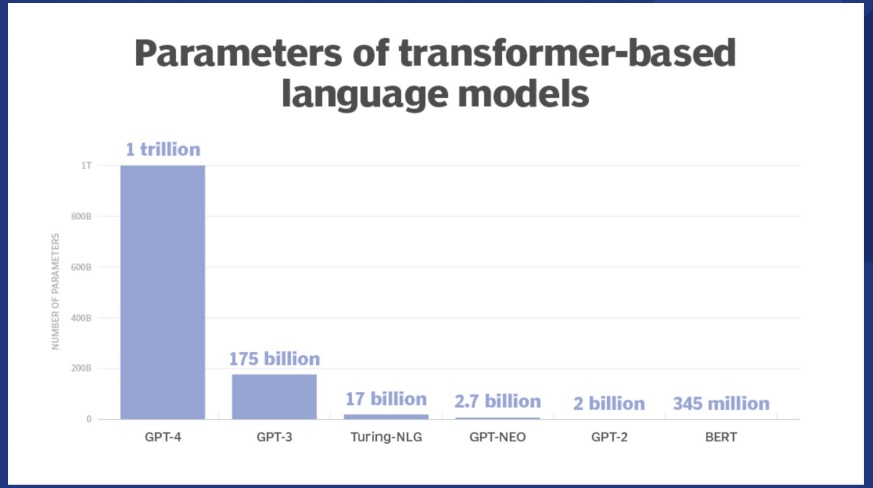
# Generative AI

É um tipo de inteligência artificial capaz de gerar texto, imagens ou outros tipos de mídia em resposta a estímulos, aprendendo os padrões e estruturas de dados de treinamento e gerando novos dados com características semelhante

# LLM

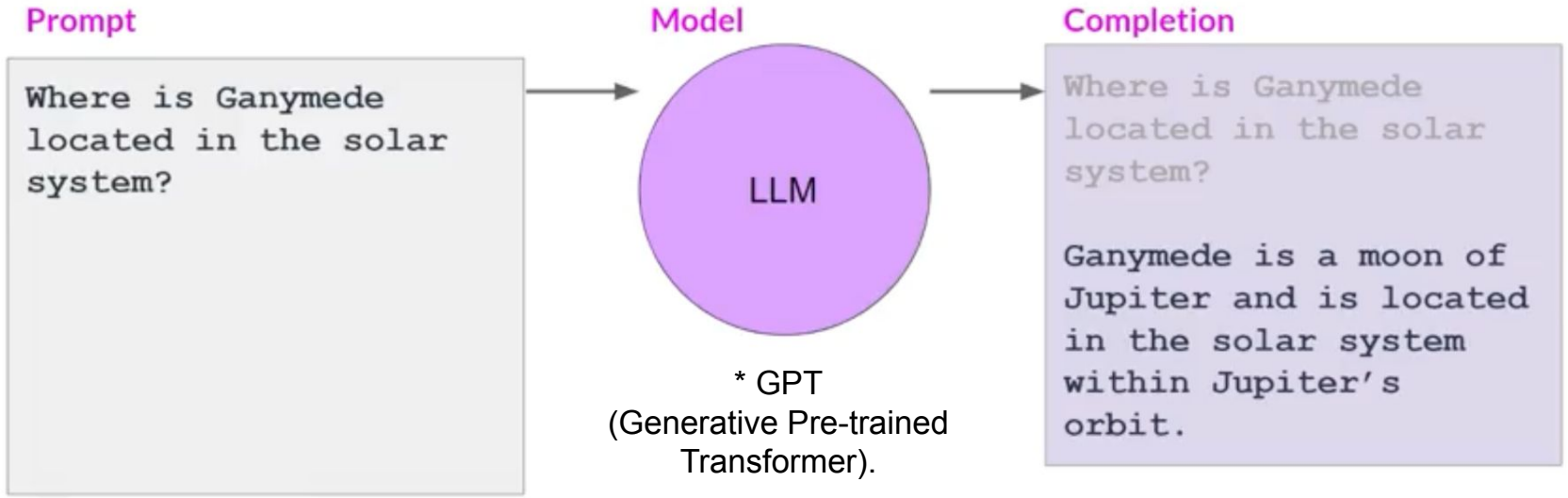
Large Language Model (LLM) é um tipo de algoritmo de inteligência artificial (IA) que utiliza técnicas de aprendizado profundo e conjuntos de dados enormes para compreender, resumir, gerar e prever novo conteúdo.




elastic

# Prompts and completions



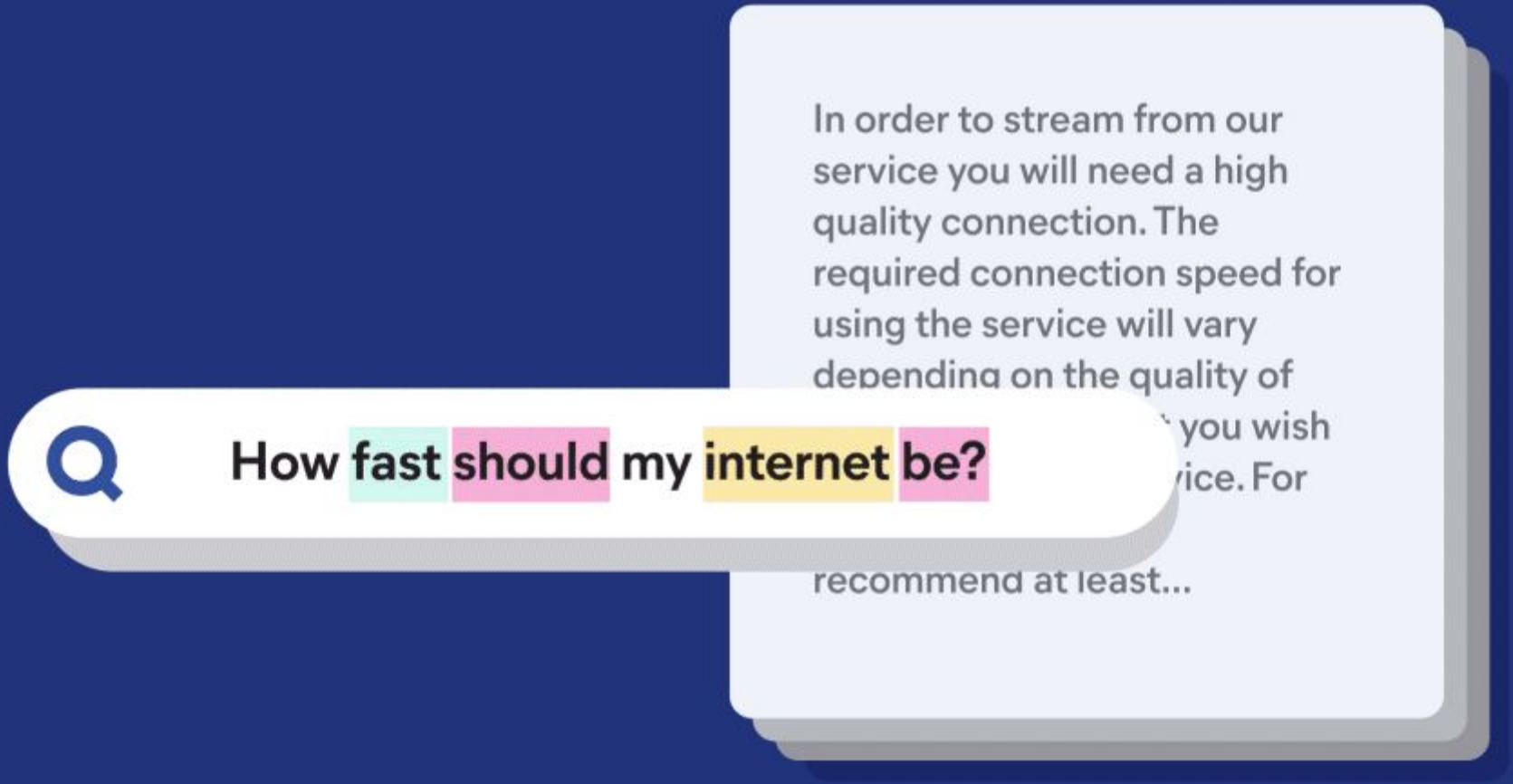

* GPT
(Generative Pre-trained Transformer).

Context window

- typically a few 1000 words.

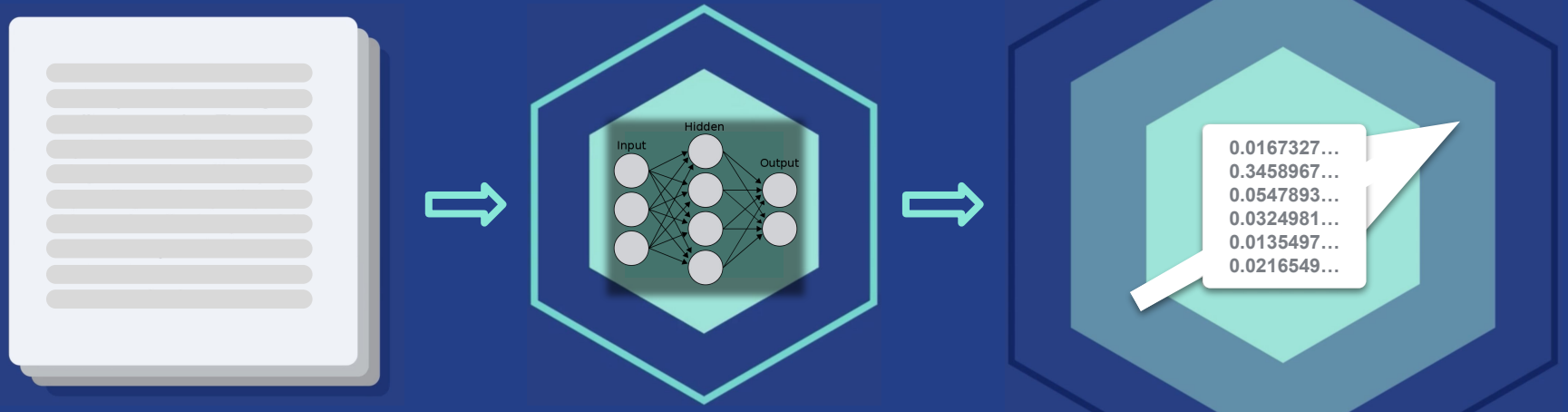
elastic

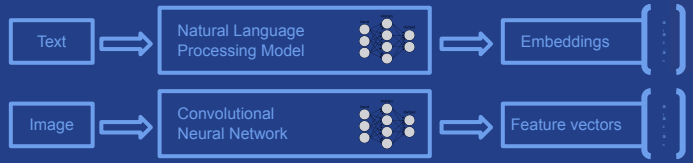
# Elasticsearch:
# You Know, for Search


How fast should my internet be?
In order to stream from our service you will need a high quality connection. The required connection speed for using the service will vary depending on the quality of
you wish
vice. For
recommend at least...

# O que é similaridade de vetores?

Converta dados em representações vetoriais onde as distâncias representam similaridade.

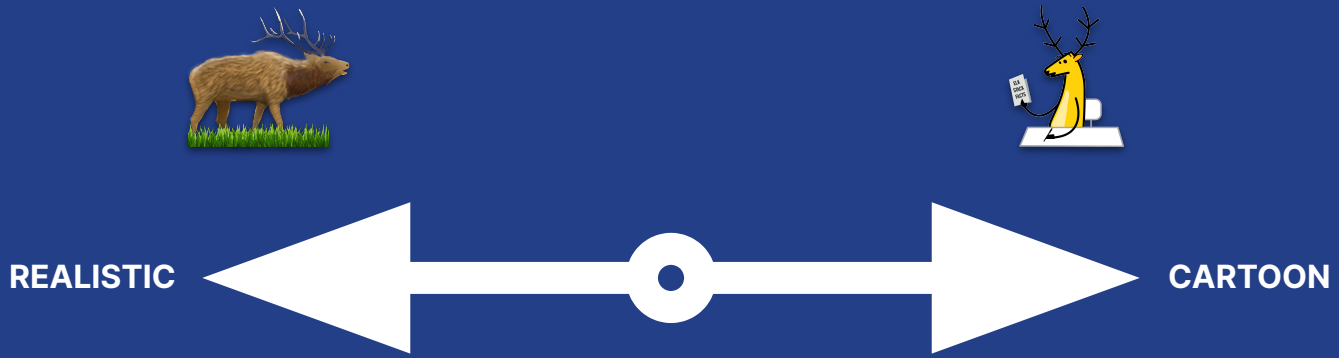


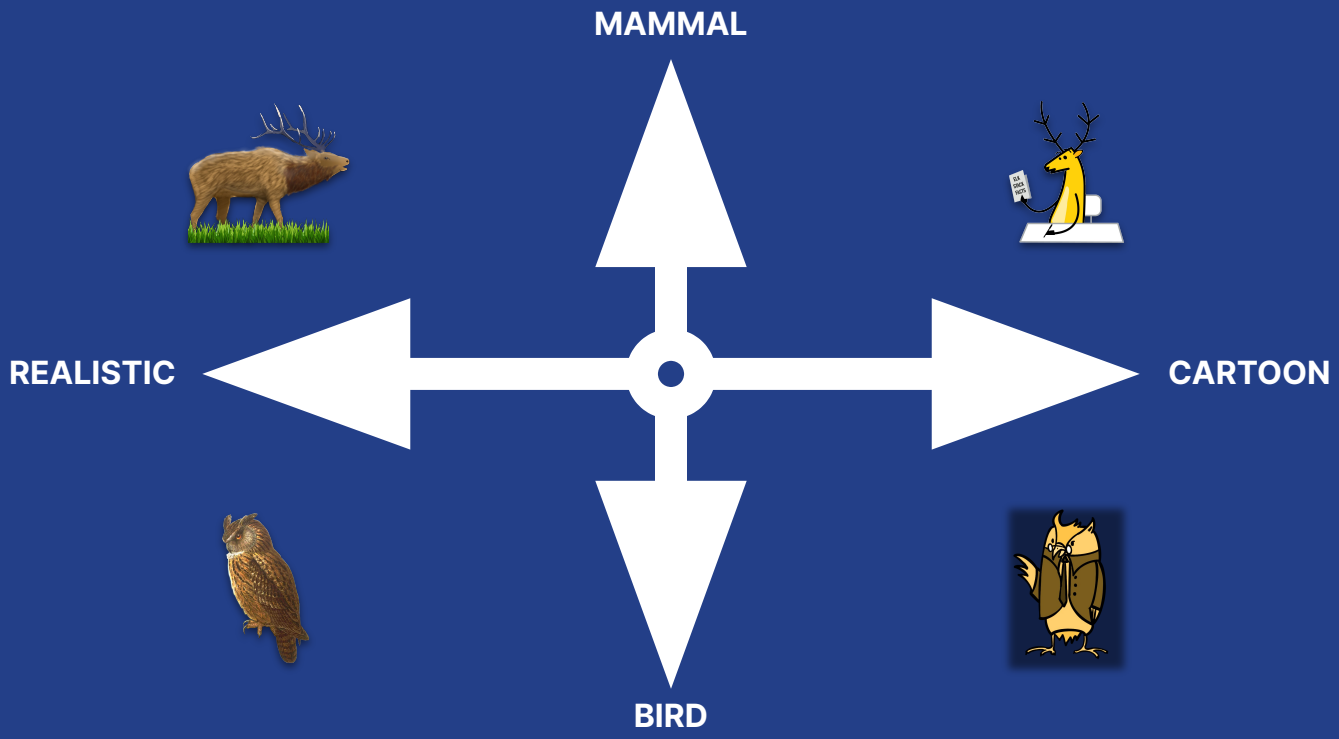

elastic

# **Embedding** é a representação do dado no espaço vetorial

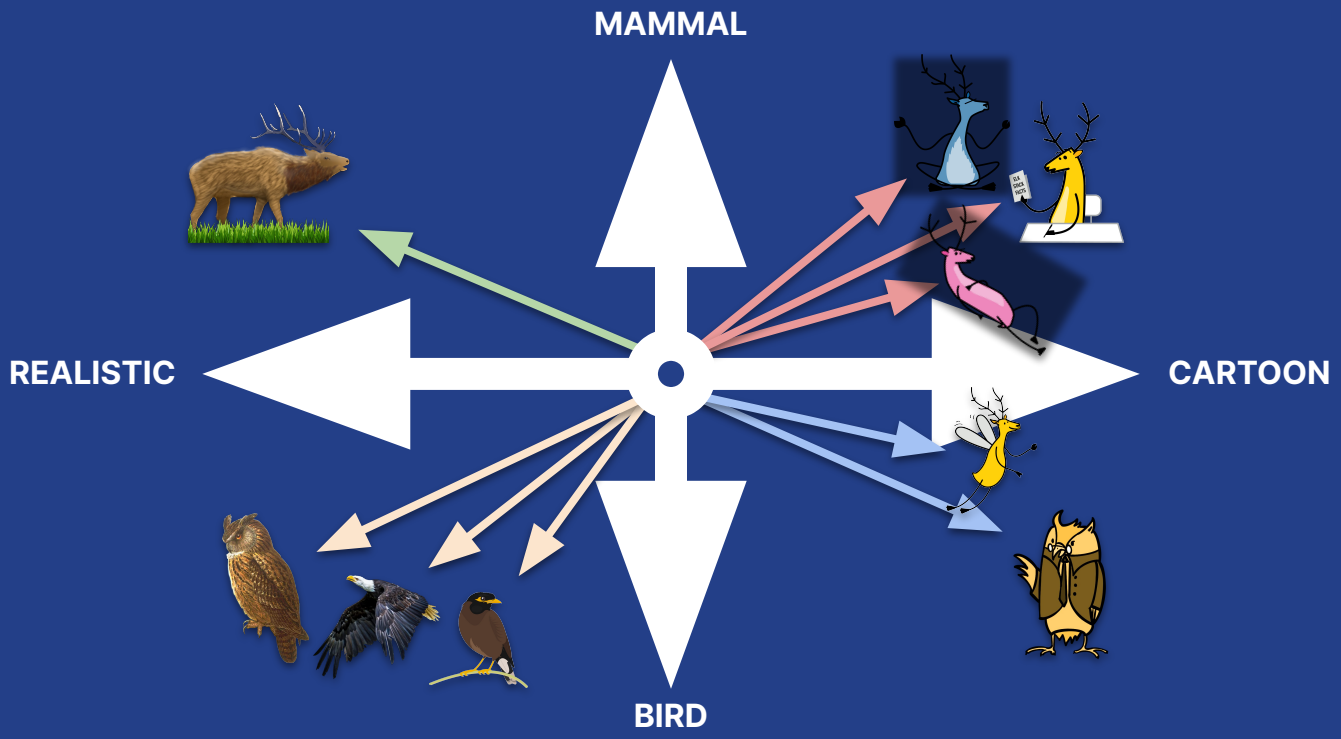
Example: 1-dimensional vector



# Múltiplas dimensões representam diferentes aspectos dos dados.



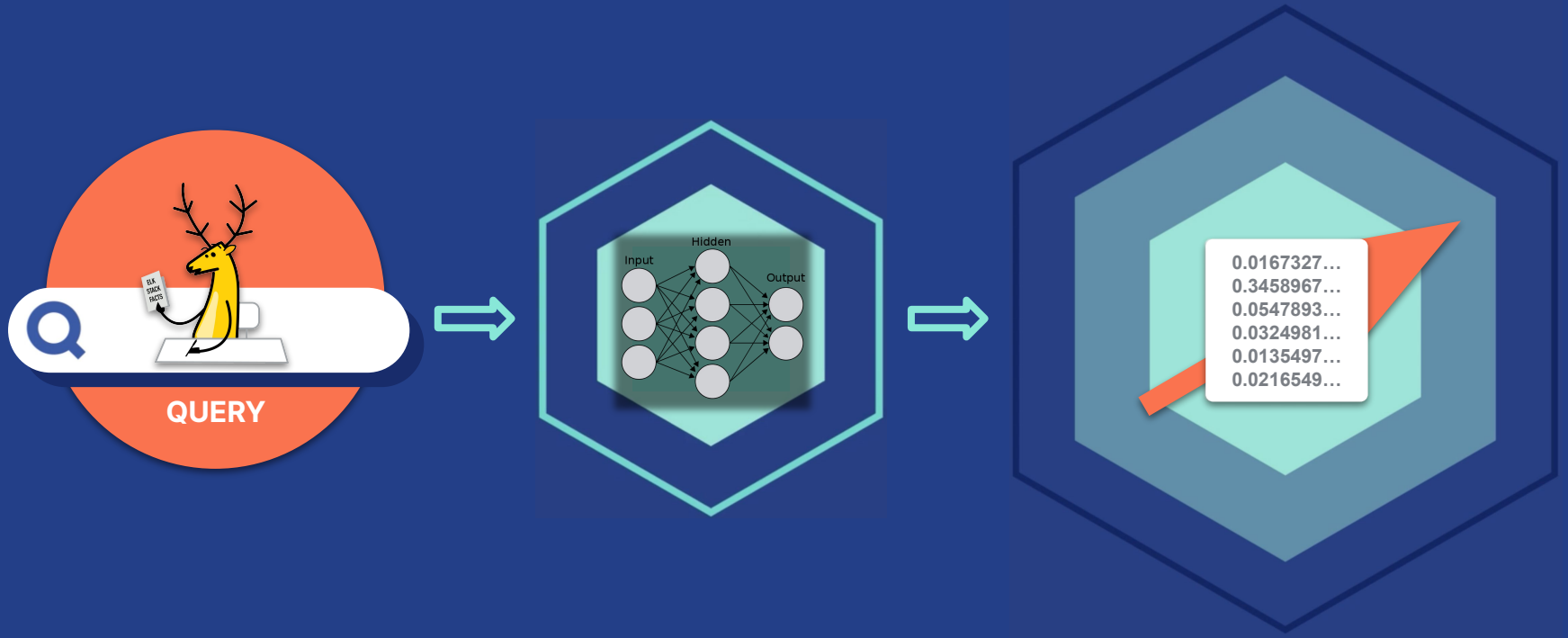
# No “Espaço vetorial”, dados similares são agrupados juntos.




elastic

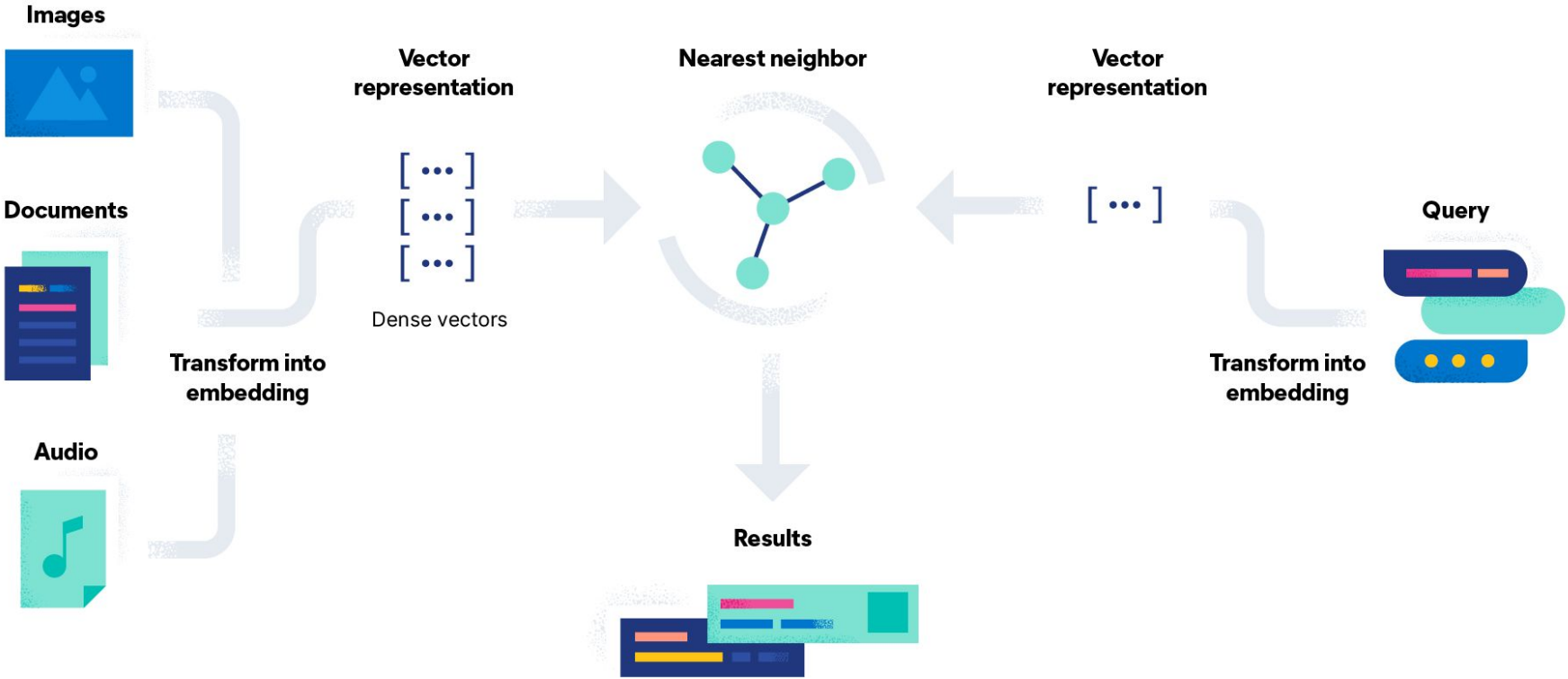
# O que é similaridade de vetores?

A consulta também é vetorizada (embeddings)

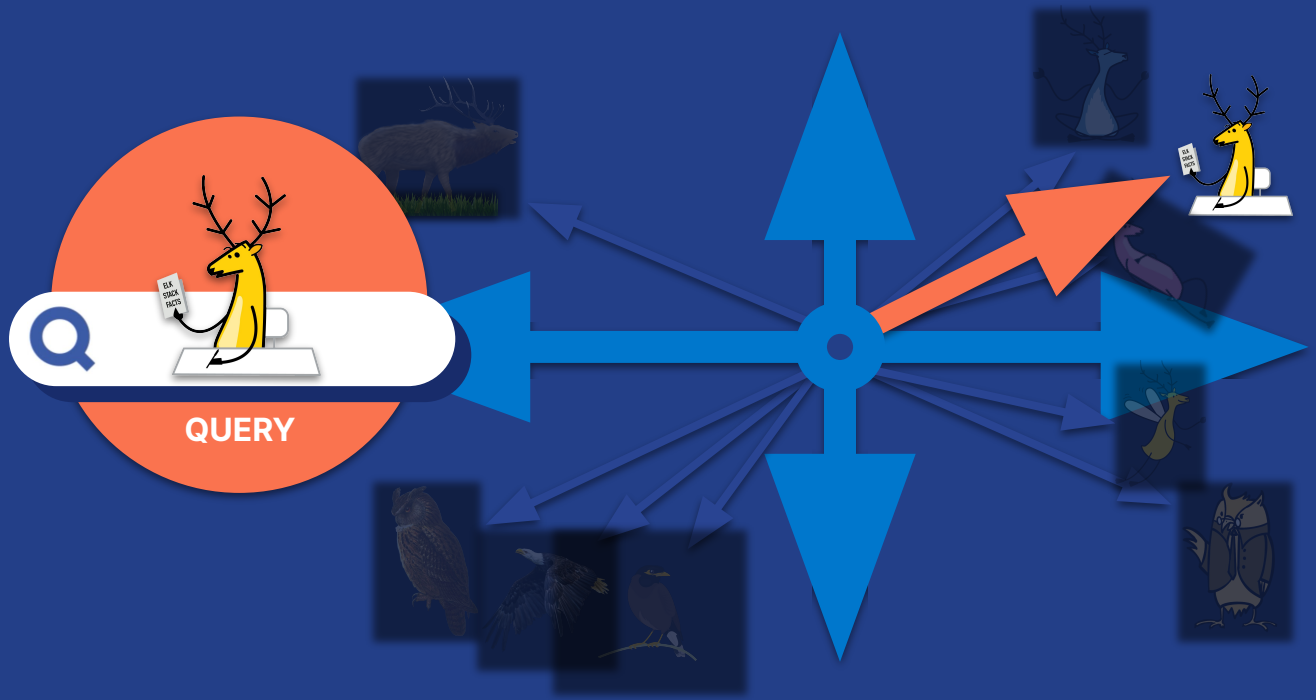



elastic

# Consultas também são vetorizadas



# Busca vetorial ranqueia os objetos baseados em similaridades (relevância) com a consulta




| Relevance | Result |
|---|---|
| Query |  |
| 1 |  |
| 2 |  |
| 3 |  |
| 4 |  |
| 5 |  |

elastic

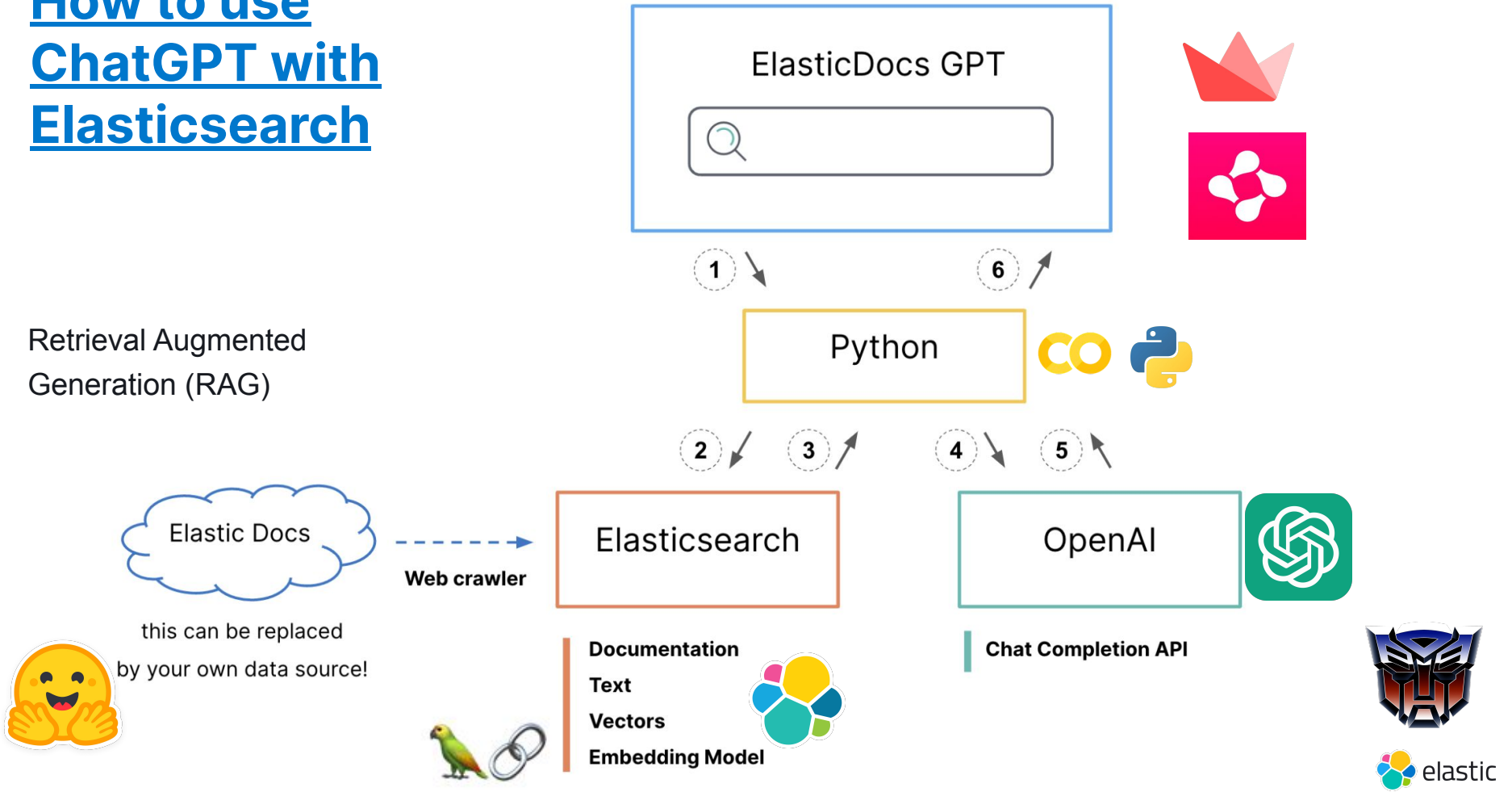
# Prática

Vamos construir nossa aplicação
LLM para chat?

elastic

# How to use ChatGPT with Elasticsearch

Retrieval Augmented Generation (RAG)




elastic

# Demo

https://github.com/salgado/ElasticDocs_GPT

https://www.elastic.co/blog/chatgpt-elasticsearch-openai-meets-private-data

elastic

# Conecte-se com a comunidade Elastic no Brasil

Encontre um grupo

https://community.elastic.co/

Veja os eventos locais

https://ela.st/brvirtual

# Obrigado

Alex Salgado


*@alexsalgadoprof*


*@alexsalgadoprof*


*salgado*


*/in/alex-salgado/*

elastic

# Link workshop-mackenzie

https://github.com/prof-alex/workshop-mackenzie

## Blogs

https://www.elastic.co/search-labs/chatgpt-elasticsearch-openai-meets-private-data

## Blogs de Alex

https://www.elastic.co/pt/blog/author/alex-salgado

https://dev.to/salgado

elastic

# Para casa

## Recriar o seu próprio ChatGPT com dados privados

Usando a sua conta trial:

- Replicar o passo a passo desse workshop usando como referência o blog : https://www.elastic.co/search-labs/chatgpt-elasticsearch-openai-meets-private-data
- Definir uma página de crawler que o chatgpt não saiba responder
- Mudar o título da aplicação

Durante o processo, havendo qualquer dúvida, pergunte a comunidade usando estes canais:

Aqui é nosso discuss: https://discuss.elastic.co/

Aqui nosso slack: https://join.slack.com/t/elasticstack/shared_invite/zt-24evvtls4-08ZzDILtGQNDQduk15Mw0w

# Para casa

## Recriar o seu próprio ChatGPT com dados privados

Para quem conseguir terminar:

- Crie um vídeo de 1 minuto fazendo uma busca na sua aplicação
- Publique no Linkedin com sua percepção
- Coloque a hashtag #WorkShopElastic #Elastic #Mackenzie #ChatGPT
- Marque **eu** e o professor **Rogério**

Aguarde a surpresa !!!!