

**Desbloqueando o Poder da
Criatividade: RAG - Transformando
Informação em Inspiração!**

Desbloqueando o Poder da Criatividade: RAG - Transformando Informação em Inspiração!

ChatGPT 3.5 ▾



You

vou dar uma palestra para uma empresa chamada mobilize sobre RAG (Retrieval Augmented Generation), gere um slogan empolgante para o título da palestra



ChatGPT

"Desbloqueando o Poder da Criatividade: RAG - Transformando Informação em Inspiração!"





Olá, sou o Fabio!

Tech Lead | Senior Full Stack Developer | MBA | Professor |
Mentor | Voluntário

Formado em análise e desenvolvimento de sistemas pela
Universidade Tecnológica Federal do Paraná (UTFPR)

MBA em Data Science and Artificial Intelligence na FIAP

Zurich Seguros

Contato

E-mail: fabmasouza@hotmail.com ou fabmasouza@gmail.com

LinkedIn: <https://www.linkedin.com/in/fabiomarcelosouza/>



AGENDA

1 Retro Gen AI

2 RAG (Retrieval Augmented Generation)

3 Exemplo Prático

1 Retro Gen AI

Visão geral



Generative learning é uma abordagem de aprendizado de máquina que envolve a criação de modelos que podem gerar novos dados com base em um conjunto existente de dados de treinamento. Em outras palavras, o modelo é capaz de aprender as características e padrões do conjunto de dados de treinamento e, em seguida, criar novos exemplos que se assemelham a esses dados originais. Isso pode ser útil em várias aplicações, como na geração de texto, imagens, vídeos e até mesmo em áreas como a medicina para a criação de novos medicamentos.



Fonte: Professor Eldrey S. Galindo

1 Retro Gen AI

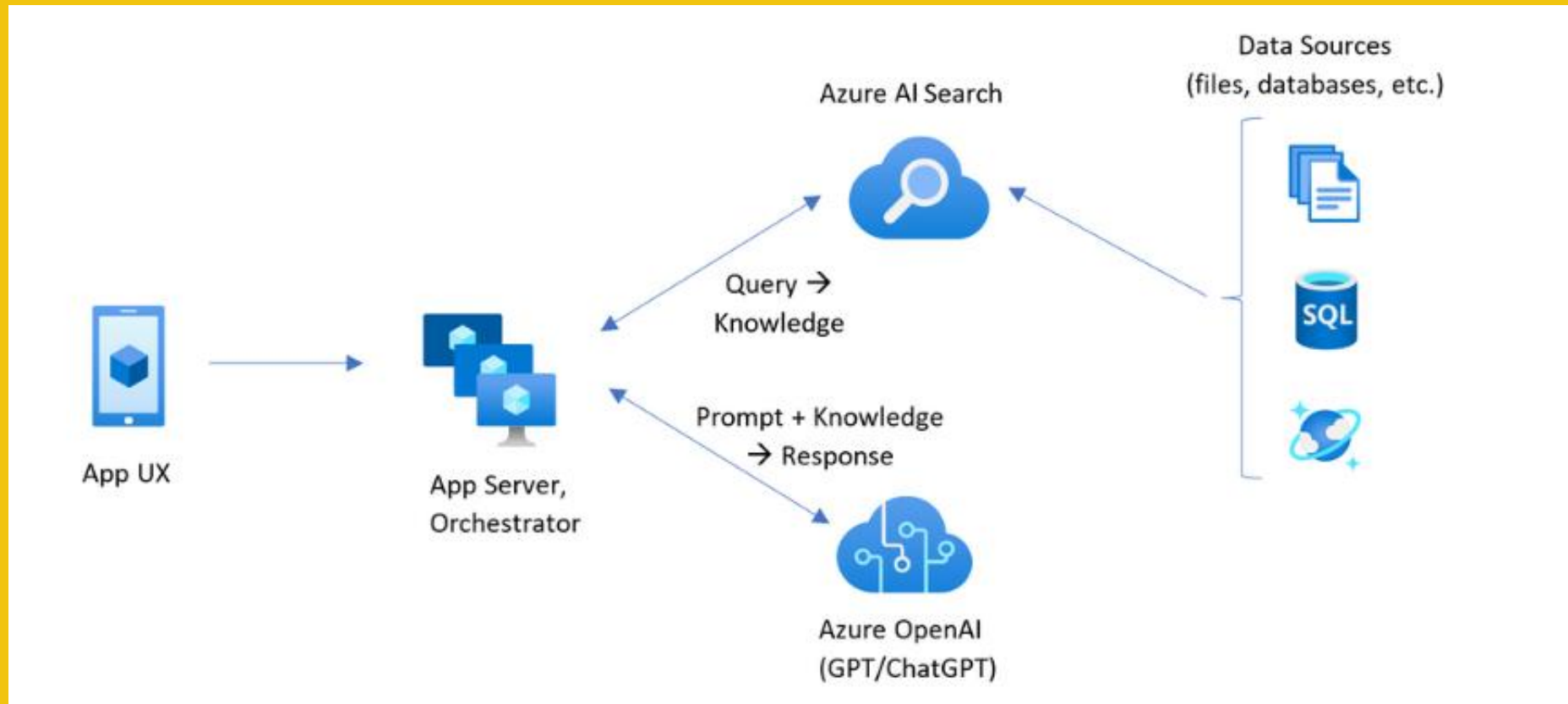
Chat GPT web

- É uma aplicação, feita para ser usada e para gerar engajamento
- A aplicação utiliza 1 versão do modelo, atualmente (01/05/23) GPT-3.5
- "Gratuito"
- As conversas ~~vão~~ podem ser usadas para re-treinamentos e melhorias
- Filtros, muitos filtros de respostas

Chat GPT API

- É uma API
- Tem custo, extremamente baixo, mas tem
- Os dados não devem ser utilizados para re-treinamentos, segundo a OpenAI, esses dados são armazenados temporariamente
- Você pode usar qualquer um dos modelos disponíveis (8 modelos em 01/05/23)

2 RAG (Retrieval Augmented Generation)



Fonte: <https://learn.microsoft.com/pt-br/azure/search/retrieval-augmented-generation-overview>

2 RAG (Retrieval Augmented Generation)

RAG amplia LLMs (Large Language Models) como ChatGPT.

Adiciona sistema de recuperação de informações.

Controla dados usados na formulação de respostas.

Ideal para aplicações empresariais.

Restringe IA generativa a conteúdo corporativo.

Utiliza documentos, imagens vetorizadas e outros formatos de dados.

Fonte: <https://learn.microsoft.com/pt-br/azure/search/retrieval-augmented-generation-overview>

3 Exemplo Prático

<https://github.com/fabio-marcelo-lodesouza/apresentacao-mobilize-arq-rag>

Obrigado e dúvidas estou à disposição!

Contato

E-mail: fabmasouza@hotmail.com ou fabmasouza@gmail.com

LinkedIn: <https://www.linkedin.com/in/fabiomarcelosouza/>

