



## Projeto Inteligência Artificial

# Buscador de Livros

### 1) Grupo

Raphael Iniesta 10396285

Tito Guidotti 10388629

### Introdução

Neste projeto, criamos um sistema de busca e recomendação de livros na forma de um chatbot com um RAG (Retrieval Augmented Generation).

### RAGs

Retrieval Augmented Generation é uma técnica para aumentar a acuracidade de um chatbot. Antes de fazer uma requisição para uma LLM, é feita uma busca semântica dos termos da entrada em um banco de dados. Assim, a LLM pode responder com mais dados sobre o tema.

### Funcionamento

O banco de dados dos livros é obtido do dataset CMU Book Summary; há uma API pública para baixar os sumários como JSON.

Esses dados são vetorizados com o uso da API da OpenAI. Os vetores são então enviados para Pinecone, um banco de dados vetorial. Através do pinecone, podemos fazer uma busca semântica de livros.

Usando streamlit, foi feita a interface de um chat. Quando o usuário manda uma mensagem, pedindo sugestões ou informações de livros, é primeiro feita uma busca no pinecone, que retorna os três principais resultados, que aparecem na coluna direita da interface. Eles são passados como parte do prompt para uma requisição ao GPT-4o-mini. A resposta é então exibida na interface.