

Retrieval-Augmented Generation aplicado em letras de músicas

Integrantes
Levi de Lima Pereira Júnior
Matheus Hensley de Figueiredo e Silva

# O que é o RAG?

## Propósito de nossa aplicação

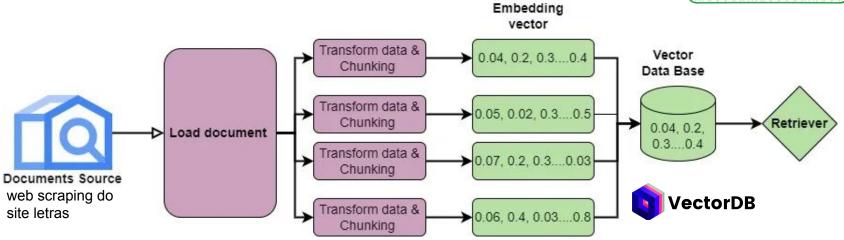
Identificar uma música dado um áudio ou uma sequência de palavras.



## Como funciona?

#### Etapa 1





## **Sentence-Transformers e BERT**

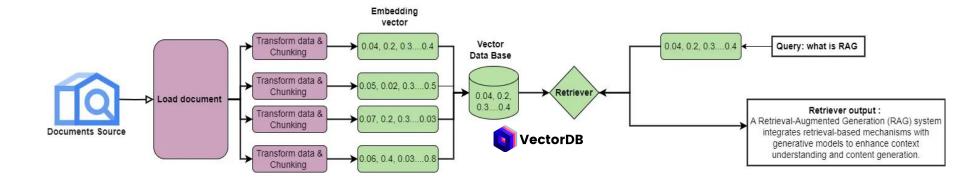
O VectorDB usa Sentence-Transformer para gerar embeddings de alta qualidade, transformando sentenças em vetores, capturando o significado semântico. Além disso ele gera metadados associados a cada embedding.



Os Metadados são o autor, título e a letra da música.

### Como funciona?

#### Etapa 2



#### Visão Geral **VectorDB** Search 0.03, 0.4.0.5...0.6 Embedder Query: What is Al Relevant 0.1, 0.2.0.3...0.02 model Context Query Embedding: 0.3,0.05,0.04..0.2 0.05, 0.3.0.05...0.1 0.02, 0.5.0.4...0.9 Vector Data Base Retieve Relevant context or information Query: What is Al Generated Response Prompt + Query + Relevant Context LLM Model **Hugging Face**

### Comparativo entre FAISS (chunks < 4000) e MRPT para maiores chunks

