



# Uso de Vector Databases e HuggingFace para Busca Semântica

ATIVIDADE FINAL – AF #11

DISCIPLINA: INTELIGÊNCIA ARTIFICIAL

GABRIEL MANTOVANNI DE SOUZA

# Busca inteligente em textos sobre golpes bancários digitais

- Usuários recebem muitos conteúdos sobre segurança digital (posts, artigos, avisos dos bancos).
- Encontrar rapidamente informações relevantes sobre golpes (phishing, smishing, etc.) não é trivial.
- Busca tradicional por palavra-chave é limitada (não entende sinônimos ou contexto).
- Objetivo: criar uma busca semântica simples usando **HuggingFace + Vector Database**.

# Arquitetura da solução

- ▶ **Pontos (pode virar um diagrama simples):**
- ▶ **HuggingFace (SentenceTransformer)**
  - ▶ Modelo: sentence-transformers/all-MiniLM-L6-v2
  - ▶ Gera *embeddings* (vetores) para textos sobre golpes bancários.
- ▶ **Vector Database (FAISS)**
  - ▶ Armazena os vetores dos documentos.
  - ▶ Faz busca por similaridade quando o usuário faz uma pergunta.
- ▶ **Fluxo:**
  - ▶ Cadastro dos textos na base (documentos sobre golpes).
  - ▶ Geração dos embeddings com HuggingFace.
  - ▶ Indexação no FAISS.
  - ▶ Usuário faz uma pergunta → embedding da pergunta.
  - ▶ FAISS retorna os textos mais semelhantes (top-k).

# Resultados e aprendizados

```
▶ buscar_documentos("Como me proteger de golpes por SMS?")
```

... Pergunta: Como me proteger de golpes por SMS?

Documentos mais relevantes:

1. (distância=0.6550)

Uma boa prática de segurança é nunca clicar em links recebidos por SMS de remetentes desconhecidos.

2. (distância=0.7940)

Bancos legítimos não pedem senhas completas ou códigos de autenticação por telefone, e-mail ou SMS.

3. (distância=0.8318)

Smishing é um phishing feito por SMS, em que mensagens falsas tentam induzir o usuário a clicar em links maliciosos.

```
▶ buscar_documentos("Bancos pedem senha completa por telefone?")
```

... Pergunta: Bancos pedem senha completa por telefone?

Documentos mais relevantes:

1. (distância=0.3944)

Bancos legítimos não pedem senhas completas ou códigos de autenticação por telefone, e-mail ou SMS.

2. (distância=1.0308)

Uma boa prática de segurança é nunca clicar em links recebidos por SMS de remetentes desconhecidos.

3. (distância=1.1373)

Golpes bancários digitais frequentemente envolvem falsos atendentes pedindo códigos de verificação ou senhas.