**PROPOSTA DE TRABALHO DE CONCLUSÃO DE CURSO**

# TÍTULO

**Desenvolvimento de Sistema de Comparação de Conteúdo de Textos**

# INTRODUÇÃO

A crescente produção e disseminação de conteúdos textuais na internet e em ambientes corporativos exige ferramentas eficientes para análise e comparação semântica de textos. O avanço da Inteligência Artificial e do Processamento de Linguagem Natural (PLN) permite o desenvolvimento de sistemas capazes de identificar similaridades textuais com alta precisão. Neste contexto, este trabalho propõe o desenvolvimento de um sistema baseado em técnicas de PLN para comparação de textos, utilizando modelos de aprendizado profundo disponíveis na plataforma Hugging Face.

# PROBLEMA

Em diversos contextos, como análise acadêmica, jurídico, jornalístico e corporativo, a necessidade de verificar a semelhança entre textos é uma demanda crescente. Métodos tradicionais de busca por palavras-chave não são eficientes para identificar semelhanças semânticas profundas. Assim, há a necessidade de um sistema automatizado capaz de comparar textos e fornecer métricas de similaridade precisas.

# OBJETIVOS

**Objetivo Geral**

Desenvolver um sistema web para comparação de conteúdo textual utilizando modelos de PLN.

**Objetivos Específicos**

* Investigar modelos de PLN para similaridade textual, com foco na biblioteca **Sentence Transformers**;
* Implementar uma API que permita a comparação de textos com base em embeddings semânticos;
* Criar uma interface web intuitiva para facilitar o uso do sistema;
* Testar e validar a precisão do sistema em diferentes domínios textuais;
* Disponibilizar o sistema como um **Software as a Service (SaaS)**.

# JUSTIFICATIVA

A comparação de textos é uma necessidade em múltiplas áreas do conhecimento. No meio acadêmico, é útil para a detecção de plágio; no setor jurídico, pode ser utilizada para a análise de contratos e jurisprudências; no jornalismo, auxilia na identificação de notícias duplicadas; e no setor corporativo, pode ser usada para a gestão de documentos. O uso de técnicas modernas de PLN possibilita um sistema mais eficiente do que os métodos convencionais de comparação textual baseados em palavras-chave.

# METODOLOGIA

O desenvolvimento do sistema seguirá as seguintes etapas:

1. **Revisão Bibliográfica**: Pesquisa sobre técnicas de comparação de textos, modelos de aprendizado profundo e ferramentas disponíveis;
2. **Definição da Arquitetura**: Escolha das tecnologias adequadas para backend (Python com Flask/Django), frontend (React/Vue.js) e banco de dados (PostgreSQL/MySQL);
3. **Desenvolvimento da API**: Implementação de endpoints para recepção dos textos, processamento e retorno dos índices de similaridade;
4. **Desenvolvimento da Interface Web**: Criação de um painel de controle para inserção de textos e exibição dos resultados;
5. **Testes e Validação**: Aplicação do sistema a conjuntos de dados reais e avaliação da precisão dos modelos;
6. **Hospedagem e Disponibilização**: Publicação do sistema como um SaaS, possibilitando seu uso em larga escala.

# CRONOGRAMA

| **Etapa** | **Mês 1** | **Mês 2** | **Mês 3** | **Mês 4** | **Mês 5** | **Mês 6** |
| --- | --- | --- | --- | --- | --- | --- |
| Revisão Bibliográfica | X |  |  |  |  |  |
| Definição da Arquitetura | X | X |  |  |  |  |
| Desenvolvimento da API |  | X | X |  |  |  |
| Desenvolvimento do Frontend |  |  | X | X |  |  |
| Testes e Validação |  |  |  | X | X |  |
| Escrita e Defesa do TCC |  |  |  |  | X | X |

# REFERÊNCIAS

* REHUREK, Radim; SOJKA, Petr. **Software Framework for Topic Modelling with Large Corpora**. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010.
* DEVLIN, Jacob et al. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. arXiv preprint arXiv:1810.04805, 2018.
* HUGGING FACE. **Sentence Similarity Task**. Disponível em: <https://huggingface.co/tasks/sentence-similarity>. Acesso em: [data de acesso].