

# Assistente LLM de Política Portuguesa

## 1. Objetivo Principal

O objetivo principal deste projeto pessoal é explorar o desenvolvimento de um assistente baseado em Modelos de Linguagem Amplos (LLM - Large Language Models) com um foco específico na política portuguesa. A meta inicial é utilizar dados concretos das Eleições Legislativas de 2025 para criar uma ferramenta que ajude a compreender de forma mais clara e acessível as propostas e posições dos diferentes partidos políticos para essas eleições. Pretende-se dar um pequeno contributo para o esclarecimento dos cidadãos.

## 2. Funcionalidades Exploradas

Com o conjunto de dados disponível nesta fase inicial, focado nos materiais das eleições de 2025, as funcionalidades a serem exploradas e desenvolvidas serão centradas neste material. As ideias incluem:

- **Resumos dos Programas:** Gerar resumos concisos do conteúdo dos programas eleitorais dos partidos para 2025.
- **Informação por Tema:** Apresentar as propostas ou posições dos partidos em temas específicos, utilizando os documentos temáticos criados manualmente a partir dos programas.
- **Perfil Básico de Partidos:** Criar uma descrição sucinta de cada partido baseada estritamente na informação contida nos seus programas de 2025.
- **Matching:** Investigar a possibilidade de, com base em poucas características ou interesses indicados pelo utilizador, sugerir partidos cujas propostas nos programas de 2025 pareçam alinhar-se, **sublinhando sempre que é uma correspondência limitada e baseada apenas nos documentos de 2025.**

## 3. Público-Alvo

Este projeto destina-se principalmente a cidadãos portugueses que procuram uma forma simples de aceder à informação sobre os partidos e os seus programas para as Eleições de 2025. É também para qualquer pessoa interessada em ver uma aplicação prática de LLMs a um domínio específico como a política portuguesa, mesmo que de forma limitada.

## 4. Fontes de Dados

Para este projeto, o foco é a utilização de dados diretamente relacionados com as Eleições Legislativas de Maio de 2025.

#### 4.1. Fonte de Dados Principal:

Os Programas Eleitorais completos dos partidos políticos Portugueses que concorreram às Eleições Legislativas de Maio de 2025. Estes documentos foram obtidos a partir dos sites oficiais de cada partido.

Documentos de resumo e comparação organizados por tema, criados manualmente e com ajuda de uma llm exterior para decompor de forma organizada a informação – foi retificado) a partir da extração e análise do conteúdo dos programas eleitorais completos. (Tal como visível na pasta 'data' com ficheiros como programa\_eleitoral\_2025\_tema\_saude.pdf, programa\_eleitoral\_2025\_tema\_trabalho.pdf, etc.)

*Nota:* Links como o do Observador (<https://observador.pt/interativo/guia-eleitoral-interativo-as-propostas-dos-partidos-para-10-areas-centrais/#partidos>) foram úteis para identificar partidos e temas relevantes, mas a recolha dos documentos completos e a criação dos resumos temáticos foram trabalho manual baseado nos programas oficiais.

#### 4.2. Potenciais Fontes de Dados Futuras (para uma evolução do projeto):

Resultados eleitorais históricos. Informação dos sites oficiais da CNE e Assembleia da República. Arquivo selecionado de notícias e análises políticas.

### 5. Pilha Tecnológica (a Explorar)

- Linguagem: Python
- Ferramentas para LLMs: Bibliotecas como transformers da Hugging Face, PyTorch ou TensorFlow. Exploração de técnicas como Embeddings e Bases de Dados Vetoriais para gerir e pesquisar a informação dos programas.
- Tratamento de Dados: Bibliotecas Python para leitura de PDFs e processamento de texto (e.g., PyPDF2, spaCy, NLTK).
- Gestão do Código: Git e GitHub para controlo de versões e organização do projeto.
- Interface (Ideia para o futuro): Uma pequena interface web (usando Flask ou similar) para interagir com o assistente.

### 6. Fases Iniciais do Projeto

1. **Configuração:** Estruturar o repositório no GitHub.
2. **Processamento de Dados:** Assegurar que o texto dos programas de 2025 e dos documentos temáticos está extraído e organizado de forma utilizável.
3. **Aprendizagem Prática:** Começar a usar bibliotecas LLM em Python para carregar dados e experimentar tarefas básicas (leitura, resumo) com base nos documentos.

4. **Implementação Mínima:** Desenvolver uma primeira versão muito simples de uma funcionalidade (ex: obter o resumo de um programa, ou a secção de um tema).
5. **Testes e Iteração:** Testar o que foi implementado e melhorar gradualmente, focando-se em como o LLM interage com os dados fornecidos.

## 7. Considerações Importantes: Âmbito e Limitações

É crucial ter presente que este projeto é um esboço inicial, de cariz pessoal e com um âmbito deliberadamente limitado nesta fase.

O assistente baseia-se quase exclusivamente nos programas eleitorais e nos resumos temáticos das Legislativas de Maio de 2025. Devido a esta limitação de dados, o assistente não oferece uma visão completa ou dinâmica da política portuguesa. Faltam elementos essenciais que estariam num sistema mais robusto, como: resultados históricos, atividade parlamentar atual, cobertura noticiosa, a evolução das posições dos partidos ao longo do tempo, ou contexto histórico detalhado. As capacidades serão diretamente condicionadas pela informação contida nos programas de 2025. Ele poderá falar sobre o que está nesses documentos, mas não sobre o que aconteceu *depois* ou o que não foi incluído nos programas.

Este projeto serve como um ponto de partida para aprender e aplicar conceitos de LLMs a um domínio específico. As limitações atuais definem o foco, mas também abrem portas para futuras expansões, caso se decida continuar a desenvolver.

---

### Passo a passo das coisas que fiz:

1. Fui procurar documentos, tudo em pdf, serão os nossos dados como já referi anteriormente.
2. Transformei os em txt (processed data)