Portuguese Politics LLM Assistant

# 1. Project Goal

The primary goal of this project is to develop a Large Language Model (LLM) based assistant focused specifically on Portuguese politics. The aim is to provide users with easily accessible and understandable information about the Portuguese political landscape, parties, and elections.

# 2. Specific Functionality

The project intends to implement functionalities that help users navigate and understand Portuguese politics, such as:

* **Election Information:** Provide details about past and upcoming elections, including results, key dates, and electoral systems.
* **Party Profiling:** Offer concise summaries and information about different Portuguese political parties, their history, ideology, key figures, and main policy positions.
* **Party Matching:** Allow users to describe their political views or priorities (e.g., "concerned about the environment," "focus on economic growth," "value social welfare") and receive suggestions for political parties whose programs or ideologies align with those characteristics.
* **Policy Explanations:** Explain complex political policies or legislative initiatives in simpler terms.
* **Historical Context:** Provide historical context on political events, government formations, and significant policy changes in Portugal.

**3. Target Audience**

The intended audience for this project includes:

* Portuguese citizens seeking to make informed decisions during elections.
* Students and researchers interested in Portuguese politics.
* Anyone outside Portugal interested in learning about the country's political system.

**4. Data Sources**

To train or fine-tune the LLM and provide accurate information, the project will require data from various sources, including:

* Official websites of Portuguese electoral bodies (e.g., Comissão Nacional de Eleições - CNE).
* Official websites of the Portuguese Parliament (Assembleia da República).
* Websites and official documents (manifestos, programs) of Portuguese political parties.
* Archives of reputable Portuguese news outlets.
* Public records and legislative documents (e.g., Diário da República Eletrónico - DRE).
* Academic research and publications on Portuguese politics.

**5. Technology Stack (Planned)**

* Language: Python
* **LLM Frameworks/Libraries:** Hugging Face transformers, PyTorch or TensorFlow, potentially libraries for working with embeddings and vector databases.
* **Data Handling:** Libraries like Pandas, NumPy, and potentially tools for web scraping (while respecting terms of service and copyright).
* **Version Control:** Git and GitHub for code hosting and collaboration.
* **Deployment (Future Consideration):** Potentially a web framework (like Flask or Django) for a user interface.

**6. Project Phases (Initial Thoughts)**

1. **Setup:** Set up the GitHub repository and initial project structure.
2. **Learning & Research:** Deepen understanding of LLMs and gather potential data sources.
3. **Data Collection & Preprocessing:** Collect and clean relevant data about Portuguese politics and parties.
4. **Model Experimentation:** Experiment with existing pre-trained LLMs and techniques for adapting them to the specific domain of Portuguese politics (e.g., fine-tuning, retrieval augmented generation - RAG).
5. **Feature Development:** Build specific functionalities (e.g., party information retrieval, party matching logic).
6. **Testing & Refinement:** Test the assistant's performance and refine the model and data.

# 7. Scope and Limitations

**It is important to note that this project is currently an initial idea and a small-scale sketch ("esboço").** The scope will be limited. The amount and variety of data initially available for training or informing the LLM will likely be constrained compared to what would be ideal for a comprehensive, production-ready political assistant. The project will evolve incrementally, and the focus will be on building core functionalities with the available resources.