**Democracia Aumentada: Um Sistema para a Democracia Digital**

A Internet e as redes sociais estão cada vez mais presentes no nosso cotidiano, sendo uma ferramenta essencial para realização de transações e comunicação entre as pessoas. Entretanto, algumas instituições ainda funcionam da mesma forma, em especial os órgãos políticos que conduzem suas tarefas através de longas reuniões que geram documentos extensos e com um linguajar de difícil compreensão. Com isso em mente, este trabalho visa utilizar técnicas de Inteligência Artificial (IA) e Processamento de Linguagem Natural (PLN) para extrair dados relevantes e sucintos dos textos do Diário da Assembleia da República Portuguesa (DAR), através de técnicas de Sumarização Automática, Modelagem de Tópicos e Análise de Sentimentos. Dentre as contribuições deste trabalho, vale destacar, o aprendizado e uso de modelos pré-treinados, bem como sua validação principalmente os pontuados como estado-da-arte, os *Transformers*, neste trabalho foram utilizados: BERTimbau, M2M100, PEGASUS, BERTopic, roBERTa, além da contribuição com três publicações de artigos que inovam nas abordagens para textos sequenciais e debates moderados para um conjunto de dados exclusivamente em português, uma língua que geralmente não é o foco das pesquisas em NLP. O trabalho propõe criar uma ferramenta Web em que o usuário possa interagir com os dados do DAR de forma mais eficiente e direta, pegando diretamente da fonte as informações que deseja e, no estado atual, as funcionalidades de Modelagem de Tópicos (BERTimbau + BERTopic + roBERTa) e Sumarização Automática (M2M100 + PEGASUS), encontram-se bem desenvolvidas conseguindo identificar temáticas relevantes e resumir discussões inteiras de forma coerente.

Sobre os principais problemas que foram encontrados durante a execução do projeto e a forma que planejamos lidar com eles, temos:

1. Extração de PDFs: Apesar de existirem bibliotecas que trabalham com a extração de dados desta forma, a inexistência de uma ferramenta poderosa para esta tarefa foi um grande empecilho no início do trabalho, só sendo superada quando encontramos os dados em extensão .txt;
2. Modelos em Português: A maioria dos modelos referenciados como estado-da-arte estão em língua inglesa e não tem uma versão análoga para português, para superar isso fizemos o *fine-tuning* do BERTimbau para a modelagem de tópicos e no caso da sumarização utilizamos um tradutor automático para converter os textos para inglês, para então utilizar os melhores modelos;
3. Poder de Processamento: Utilizar *Transformers*, especialmente quando é necessário calibrar estes modelos é uma tarefa que pode ser custosa e demorada, pensando que o objetivo é criar uma ferramenta Web, não é viável ter estes modelos rodando a cada requisição do usuário, assim planejamos criar um banco de dados estruturado que tenha uma boa eficiência para consultas do usuário para disponibilizar rapidamente gráficos e informações.