



Modelo Baseado em LLM para Discutir o Programa Eleitoral do Partido



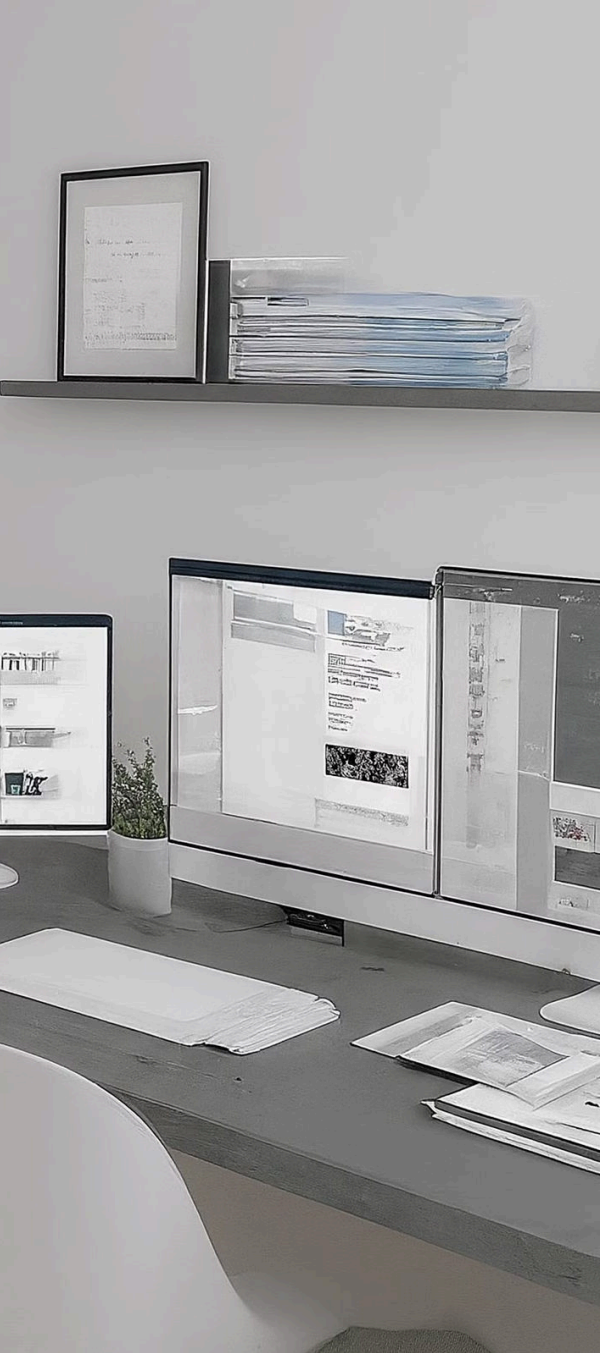
PG50944 Fautisno Sachimuco - MMC

PG50008 Marcos André Mussungo - MEI PG52762

Lívia Péres Bettero - MHD

Objetivo

Este projeto visa desenvolver e avaliar um modelo baseado em Linguagem Natural (LLM) para discutir com o usuário as propostas de governo do Partido Aliança Democrática, com base em seu Programa Eleitoral apresentado nas eleições de 2024. O objetivo é fornecer uma ferramenta ética e informativa para os cidadãos entenderem as propostas deste partido político.



Fontes de Dados

Programa Eleitoral em PDF

O programa eleitoral do Partido Aliança Democrática, disponível em formato PDF, foi uma das principais fontes de dados para este projeto.

Web Scraping de Conteúdo Oficial

Foram realizados web scraping do site oficial do partido, extraindo textos, vídeos e outras informações relevantes.

Web Scraping de Notícias Confiáveis

Também foram coletadas informações de sites de notícias considerados confiáveis, para obter uma visão mais ampla e imparcial sobre o partido.

Metodologia de Scraping

Extração do Programa Eleitoral

A extração do texto do programa eleitoral em PDF foi realizada utilizando a biblioteca **PyPDF2**, removendo caracteres especiais e quebras de linha.

1

2

Coleta de Conteúdo do Site Oficial

O conteúdo do site oficial do partido, incluindo textos, vídeos e outras informações, foi coletado usando as bibliotecas **jjcli** e **BeautifulSoup**.

3

Extração de Notícias de Sites Confiáveis e Vídeos Oficiais do Youtube

As notícias de sites considerados confiáveis foram coletadas e tratadas, removendo blocos de texto indesejados e caracteres especiais.

Filtragem de Conteúdo

Identificação de Discursos Potencialmente Antiéticos

Foram desenvolvidas estratégias para identificar e remover trechos de discursos que pudessem ser considerados antiéticos, como ataques pessoais ou conteúdo discriminatório.

Utilização de Análise de Sentimento

A biblioteca NLTK foi usada em combinação para sinalizar a polaridade das palavras, auxiliando na identificação de linguagem negativa associada a discursos de ódio.

Criação de Lista de Termos Negativos e Lematização com Spacy

Uma lista de palavras e nomes associados à oposição do AD foi criada para, em combinação com a biblioteca Spacy para lematização, garantir a filtragem de conteúdo potencialmente antiético.

Tokenização e Embeddings

Tokenização

O processo de tokenização foi realizado utilizando a biblioteca tiktoken, com o esquema de codificação cl100k_base, que divide o texto em tokens de palavras, subpalavras e caracteres.

Embeddings

Os embeddings foram gerados usando o modelo "text-embedding-3-small" da OpenAI, que codifica o significado e o contexto dos tokens em vetores numéricos.

Criação de Contexto

Uma função foi desenvolvida para encontrar o contexto mais relevante a partir dos embeddings, a fim de responder às perguntas dos usuários de forma precisa e informativa.

Modelo de Resposta a Perguntas



Pergunta

O modelo é capaz de receber perguntas dos usuários e gerar respostas com base no contexto extraído dos textos.



Processamento

Utilizando o modelo GPT-3.5 Turbo Instruct, o sistema processa a pergunta e o contexto relevante para gerar uma resposta informativa e ética.



Resposta

A resposta gerada é então apresentada ao usuário, fornecendo informações sobre o programa eleitoral do Partido Aliança Democrática.

Interface do Chatbot

Interação Amigável

A interface do chatbot foi desenvolvida utilizando a biblioteca Gradio, oferecendo uma experiência de usuário intuitiva e agradável.

Imagens Ilustrativas

Imagens relacionadas ao Partido Aliança Democrática foram adicionadas para enriquecer a experiência do usuário.

Histórico de Conversas

O chatbot mantém um histórico das perguntas e respostas, permitindo que o usuário acompanhe o desenvolvimento da conversa.

Testes e Validação



Testes e Validação

1

Avaliação de Desempenho

O modelo foi testado com uma variedade de perguntas relacionadas ao programa eleitoral do Partido Aliança Democrática, avaliando sua capacidade de gerar respostas informativas e éticas.

2

Feedback dos Usuários

Foram coletados feedbacks de usuários que interagiram com o chatbot, a fim de identificar áreas de melhoria e aprimorar a experiência geral.

3

Ajustes e Melhorias

Com base nos testes e no feedback dos usuários, ajustes foram realizados no modelo, na filtragem de conteúdo e na interface, visando aprimorar o desempenho e a usabilidade do sistema.

Futuros Desafios

Gerenciamento de Contexto Dinâmico

Um dos principais desafios é lidar com o contexto dinâmico em conversações longas, mantendo a compreensão e a capacidade de resposta adequada do modelo.

Avaliação de Usuários e Feedback

Coletar feedback dos usuários e realizar avaliações de usabilidade são fundamentais para identificar áreas de melhoria e aprimorar o desempenho do sistema.

Expansão e Aprimoramento

No futuro, o modelo poderá ser expandido para abranger outros partidos políticos e temas relacionados, tornando-se uma ferramenta mais abrangente e útil para os cidadãos.