



Ministério da Previdência Social
Secretaria do Regime Geral da Previdência Social
Departamento de Perícia Médica Federal

Coordenação-Geral de Assuntos Corporativos e Disseminação de Conhecimento
Gustavo Magalhães Mendes de Tarso

RELATÓRIO INSTITUCIONAL SISTEMA thunderstruck-oracle-hf

1 Introdução

O projeto thunderstruck-oracle-hf, também conhecido como Oráculo MPS, surge como uma resposta inovadora às crescentes demandas por eficiência e precisão na gestão de informações em sistemas públicos. Em um cenário onde a quantidade de dados disponíveis é vasta e em constante crescimento, a necessidade de ferramentas que possam processar, interpretar e utilizar essas informações de maneira eficaz é imperativa. O Oráculo MPS foi desenvolvido para atender a essa necessidade, utilizando tecnologias de ponta em processamento de linguagem natural e aprendizado de máquina para oferecer soluções automatizadas e inteligentes.

2 Objetivos

O principal objetivo do projeto é criar um sistema robusto que permita a geração automática de pares de perguntas e respostas (QA), o fine-tuning eficiente de modelos de linguagem de larga escala (LLM) e a busca semântica aprimorada. Além disso, o projeto visa implementar um sistema de geração aumentada por recuperação (RAG) que facilite a interação do usuário com o modelo, permitindo uma avaliação contínua e melhorias incrementais. Através dessas funcionalidades, o Oráculo MPS busca otimizar o acesso e a utilização de informações contidas em documentos oficiais, como editais e portarias, promovendo uma gestão pública mais eficiente e transparente.

3 Dissertação

O projeto é estruturado em um fluxo de trabalho que abrange desde a geração de dados até o deploy do modelo treinado. Inicialmente, o sistema utiliza scripts como `gen_qa_chatgpt_from_txt.py` para converter documentos em formato de texto, como editais e portarias, em arquivos de QA automáticos no formato `.jsonl`. Este processo é fundamental para alimentar o sistema com dados

estruturados que podem ser utilizados para treinamento e busca. Em seguida, o projeto emprega a indexação semântica através do uso de `faiss_index.py`, que cria um índice FAISS para facilitar a busca rápida de contexto relevante. O treinamento do modelo é realizado com o script `autogen-model-hf.py`, que utiliza a técnica LoRA para um fine-tuning eficiente, permitindo que o modelo aprenda com um custo reduzido de recursos computacionais. Após o treinamento, o modelo é mesclado e preparado para deploy, estando disponível para inferência stand-alone. A busca RAG, implementada em `rag_qa_finetuned.py`, combina a recuperação de contexto com a geração de respostas, oferecendo uma interface de linha de comando para interação com o usuário.

4 Expectativas

Espera-se que o Oráculo MPS traga significativos avanços na forma como informações são geridas e acessadas no setor público. A automatização da geração de QA e a busca semântica eficiente têm o potencial de reduzir o tempo e os recursos necessários para a interpretação de documentos complexos, como editais de concursos e portarias governamentais. Além disso, a capacidade de fine-tuning eficiente permite que o sistema se adapte rapidamente a novos dados e necessidades, garantindo que as informações estejam sempre atualizadas e relevantes. A implementação de um sistema RAG com feedback do usuário também é vista como uma oportunidade para melhorar continuamente a precisão e a utilidade das respostas geradas pelo sistema.

5 Conclusão

O projeto `thunderstruck-oracle-hf` representa um passo significativo em direção à modernização e otimização dos sistemas de gestão de informações no setor público. Ao integrar tecnologias avançadas de processamento de linguagem natural e aprendizado de máquina, o Oráculo MPS oferece uma solução abrangente e eficiente para os desafios enfrentados na interpretação e utilização de grandes volumes de dados. Com expectativas de impacto positivo na eficiência administrativa e na transparência governamental, o projeto se posiciona como uma ferramenta essencial para o futuro da gestão pública.