Projeto de Formatura – 2024

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia de Computação

Tema:

Desenvolvimento de um Ambiente Integrado para Testes e Comparações de Múltiplos LLMs

Contexto e motivação

No contexto de ferramentas de Inteligência Artificial, um tópico que tem se destacado nos últimos anos foi o desenvolvimento de grandes modelos de linguagem, ou LLMs (do inglês *Large Language Models*). Esses modelos, treinados em uma enorme quantidade de dados provindos da Internet, são capazes de responder perguntas e realizar tarefas feitas em linguagem natural.

Com a popularização desses modelos nas mais diversas aplicações e a inserção dessa tecnologia no cotidiano de várias pessoas, torna-se ainda mais importante que seja feita uma escolha direcionada do melhor modelo para determinada tarefa. Portanto, este trabalho se propõe a trazer uma solução para parte deste problema criando um ambiente que possibilite a utilização e comparação de múltiplos LLMs simultaneamente, sem a necessidade de construir um acesso a API de cada modelo manualmente.

Objetivos

- Criar um ambiente que permita a utilização de múltiplas LLMs;
- Implementar no ambiente ferramentas que permitam uma avaliação quantitativa do desempenho dos modelos na tarefa solicitada;
- Disponibilizar a aplicação para pesquisadores de Inteligência Artificial na USP para que sirva como base de pesquisa e escolha de modelos.

Múltiplos LLMs

De modo a comportar múltiplos LLMs, este trabalho utilizou o framework LangChain, capaz de instanciar a conexão com diversas APIs de modelos de linguagem, padronizando as chamadas de métodos feitas a eles. Além disso, o framework possibilita que alguns parâmetros comuns sejam padronizados, como o prompt, contando com as variáveis *system, human* e *ai*, que definem o comportamento do modelo. Isso torna possível o envio de uma mesma requisição para diferentes LLMs, facilitando a comparação entre os modelos e avaliação de suas respostas.

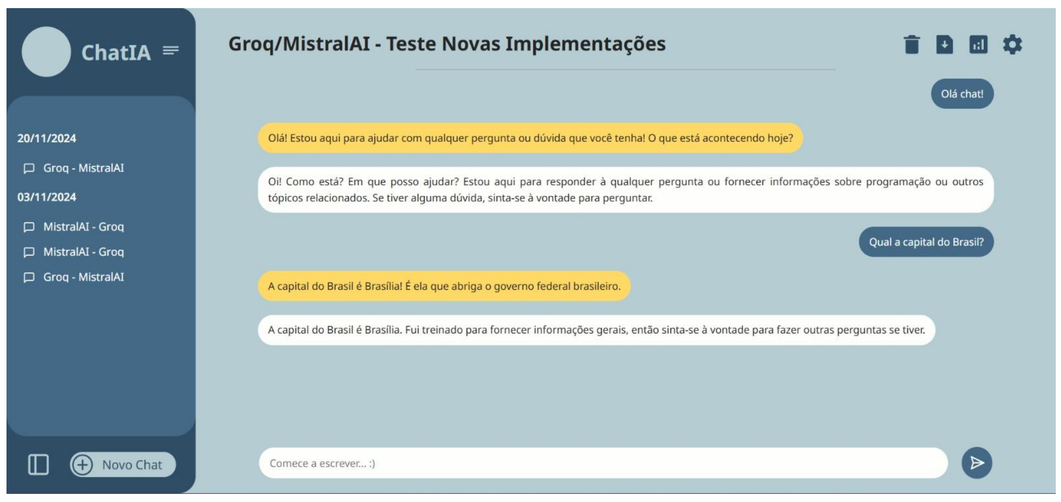
Integrantes:
- Sophia Lie Asakura
- Thiago Moreira Yanitchkis Couto
- Vinicius Ariel Arruda dos Santos

Professor(a) Orientador(a): Prof. Dr. Fábio Cozman

Métricas de avaliação

O grupo KEML (Knowledge Enhanced Machine Learning) do C4AI-USP (Center For Artificial Intelligence) tem desenvolvido um ambiente integrado focado na avaliação de respostas de modelos de linguagem sem depender do modelo específico sendo avaliado. Esse projeto, denominado HarpIA, recebe um arquivo contendo as respostas de um LLM, uma lista de respostas corretas (esperadas) e as métricas avaliativas que deseja-se utilizar para avaliar esses resultados. A partir dessa entrada, o HarpIA é capaz de computar o desempenho das respostas segundo as métricas requisitadas.

Nesse contexto, por já existir um projeto robusto focado inteiramente na avaliação dos modelos, este trabalho comprometeu-se a ser um complemento ao HarpIA. Essa atuação complementar se dá de duas maneiras. A primeira sendo no acesso a múltiplos LLMs simultaneamente para gerar, de maneira facilitada, vários arquivos com suas respostas que podem posteriormente serem avaliados utilizando o HarpIA. A segunda maneira é disponibilizar análises de métricas simples na comparação de dois modelos feita no ambiente desenvolvido neste trabalho.



Continuidade do Projeto

A partir da ferramenta desenvolvida, espera-se que seu uso possa auxiliar no desenvolvimento de novos métodos de avaliação de modelos de linguagem e na escolha de um LLM em detrimento ao outro. Além disso, também entende-se que a ferramenta pode ser utilizada como ponte de integração para futuros projetos do KEML que se beneficiem do múltiplo acesso a LLMs e do ambiente integrado de testes. Por fim, com a implementação de requisitos de segurança e financiamento do projeto, o projeto pode ser disponibilizado ao público geral.