



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

Estudantes desenvolvem sistema de testes de múltiplos LLMs para comparação e avaliação

São Paulo, dezembro de 2024

Com o avanço no uso e nas pesquisas envolvendo grandes modelos de linguagem (LLMs), surge a necessidade de identificar o modelo mais adequado para tarefas específicas. Entretanto, esse processo pode ser complexo, exigindo comunicação com cada modelo e a avaliação individual de seu desempenho.

Para enfrentar esse desafio, os alunos Sophia Lie, Thiago Yanitchkis e Vinicius Ariel desenvolveram um ambiente integrado que permite a execução simultânea e a comparação de múltiplos LLMs. Utilizando frameworks como Django, React e LangChain, a ferramenta oferece acesso às APIs de X modelos diferentes e possibilita que os usuários analisem e comparem as respostas geradas para variados inputs.

Além disso, o ambiente incorpora métricas consolidadas de avaliação de linguagem natural e conta com integração exclusiva ao projeto HarpIA, do grupo KEML-USP. Liderado pelo mestrando Vinicius Bitencourt, o HarpIA analisa os resultados de LLMs em relação a expectativas definidas. A integração permite a geração automática de arquivos de teste para cada modelo, que podem ser posteriormente avaliados pelo sistema HarpIA, oferecendo um suporte robusto à análise de desempenho.

Este projeto, entre outros, será apresentado nos dias 17 e 18 de dezembro durante a Feira de Projetos de Formatura do Departamento de Engenharia de Computação da EPUSP.

Integrantes: Sophia Lie
Thiago Yanitchkis
Vinicius Ariel

Professor(a) Orientador(a):
Co-orientador(a):

Prof. Dr. Fábio Cozman