

# Reprodução de "Recommender AI Agent: Integrating Large Language Models for Interactive Recommendations"

Alexandre Ribeiro Ferreira

July 2024

## 1 Introdução

O artigo original visa combinar o poder dos sistemas de recomendação tradicionais com os modelos de linguagem de grande escala (Large Language Models - LLMs). Para isso, foi criado um *framework* chamado **InteRecAgent**, o qual incorpora modelos LLMs como o "cérebro" da aplicação e os sistemas/modelos de recomendação como ferramentas. Esse artigo foi escolhido para reprodução por se alinhar com as ideais de pesquisa do mestrado, o qual busca um tema para pesquisa nessa área que envolve o uso de sistemas de recomendação em conjunto com LLMs.

Para executar o *framework*, é necessário algumas ferramentas de recomendação específicas, além da disponibilidade de um LLM para uso. As ferramentas de recomendação necessárias são: ferramenta de consulta de informações, ferramenta de recuperação de itens de condição leve, ferramenta de recuperação de itens de condição difícil e uma ferramenta de ranqueamento. Além disso, para que o *framework* funcione, os autores do artigo original precisaram criar alguns componentes específicos. Dentre esses componentes criados, os principais são: mecanismo de memória, mecanismo para execução primeira do plano de execução utilizando demonstrações dinâmicas e um mecanismo para detecção de resultados ruins e re-execução buscando melhores resultados.

Essa reprodução de artigo visa utilizar os artefatos disponibilizados pelos autores<sup>1</sup> para executar o *framework* desenvolvido utilizando o LLM chamado Llama 3<sup>2</sup>, o qual foi lançado após a publicação do artigo original. O objetivo é utilizá-lo para a geração e comparação de métricas dos *baselines* apresentados no artigo, assim como utilizá-lo como "cérebro" do *framework*, executando os mesmos experimentos e comparando os resultados obtidos.

---

<sup>1</sup><https://github.com/microsoft/RecAI/tree/main/InteRecAgent>

<sup>2</sup><https://llama.meta.com/llama3/>

## 2 Metodologia/Reprodução

A configuração do ambiente para execução dos experimentos e, consequentemente, da reprodução, foi realizada conforme o **README** disponibilizado junto dos artefatos no repositório associado ao artigo original. Para a execução da reprodução, foi utilizado o conjunto de dados chamado “MovieLens 10M”<sup>3</sup>, o qual foi um dos três conjuntos de dados utilizados no artigo original.

Buscando executar o *framework*, foi seguido o passo a passo disponibilizado no **README** do repositório, na seção chamada “Usage”<sup>4</sup>. No entanto, após seguir todos os passos, o *framework* é executado em modo de “usuário”, onde apenas uma interface de *chat* é disponibilizada para interação com o Agente de Recomendação.

Na tentativa de reproduzir os experimentos e/ou testes realizados durante a produção dos artigos, foi estudado todos os artefatos disponibilizados no repositório, a maioria deles carecendo de documentação ou instruções de uso. Buscou-se executar os *scripts* denominados `eval_simulator.sh` e `eval_single_turn.sh`, na esperança de ser os códigos responsáveis por reproduzir os experimentos e coletar as métricas dispostas no artigo original. No entanto, durante a execução desses *scripts*, inúmeros erros de configuração de ambiente e incompatibilidade de versões das bibliotecas foram enfrentados.

Visando fornecer os dados para a realização dos experimentos, foi executado alguns *notebooks* com códigos para pré-processamento dos dados. Esses *notebooks* também carecem de documentação, o que acabou ocasionando diversos empecilhos durante sua execução, além de acabar sendo uma execução “cega”, ou seja, sem noção do que realmente os *notebooks* iriam produzir.

Após toda a configuração do ambiente, disponibilização de uma **API OPENAI** e pré-processamento dos dados, os *scripts* para avaliação e execução dos experimentos continuaram sem executar. Diversos erros foram encontrados na tentativa de reprodução e não se conseguiu chegar a um resultado concreto para comparação com o artigo original.

## 3 Resultados

Devido à incapacidade de reprodução do artigo original, não foi possível produzir resultados para que fosse realizado a comparação com os resultados obtidos pelo artigo original.

## 4 Discussão

A falta de instruções e/ou documentação impossibilitou a reprodução do artigo original. Isso nos mostra a importância da disponibilização dos artefatos com toda a documentação detalhada sobre o uso, execução dos experimentos,

---

<sup>3</sup><https://grouplens.org/datasets/movielens/10m/>

<sup>4</sup><https://github.com/microsoft/RecAI/tree/main/InteRecAgent#usage>

obtenção dos dados e coleta das métricas. Por mais que esse artigo analisado tenha disponibilizado uma documentação rica e detalhada acerca do uso do seu *framework*, ele acabou por pecar na disponibilização de documentação para que possibilitasse a reprodução de seus experimentos, testes e comparações.

O artigo dispõe de diversos experimentos e análises, porém os artefatos relacionados não estão disponíveis ou carecem de documentação que consiga ao menos identificar o propósito de cada um. Dessa forma, fica uma crítica aos autores por não disponibilizarem ao menos o essencial para poder ser reproduzido ao menos um dos experimentos/análises realizadas durante o artigo.

## 5 Conclusão

Diversos empecilhos e problemas foram encontrados durante a tentativa de reprodução do artigo original. Os autores não disponibilizaram todos os artefatos necessários para reprodução, além de que os que foram disponibilizados carecem de documentação ou instruções.

Os desafios encontrados durante essa reprodução nos motiva a buscar sempre disponibilizar todos os artefatos produzidos durante uma pesquisa, além de documentá-los e produzir simples instruções para execução, análise, testes e comparações.