

# Hotmart Challenge

Alessandro Queiroz | Jul 2025

---

Last Modified: Jul 10, 2025

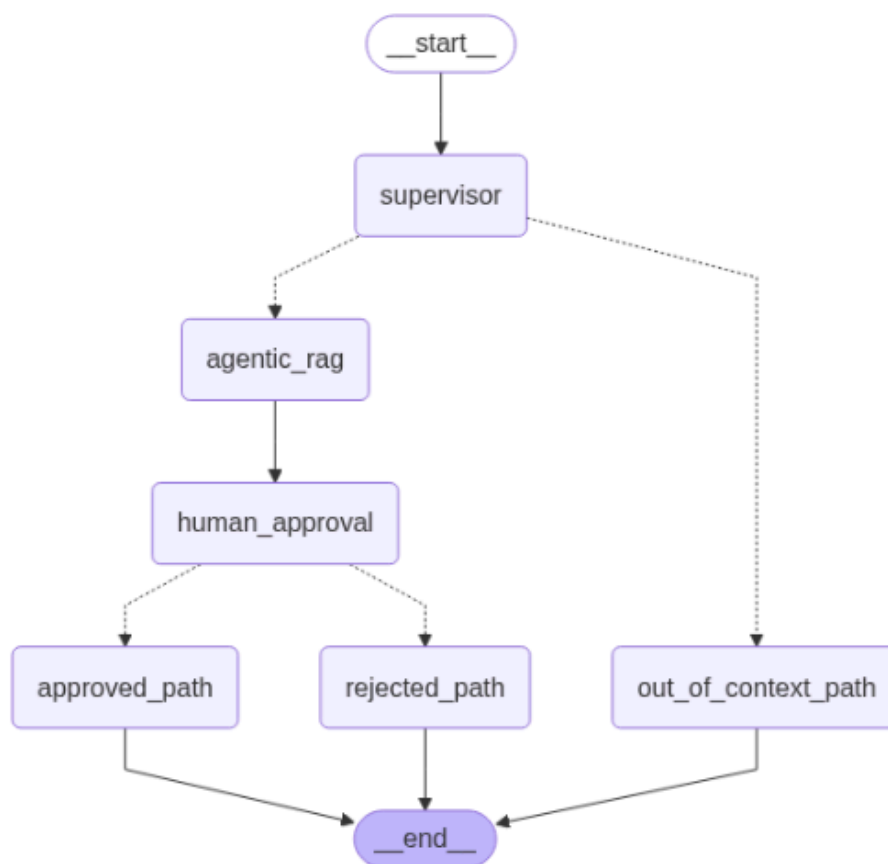
## Sobre a Solução

Para o desenvolvimento da solução, utilizei o LangGraph como ferramenta principal de orquestração, juntamente com a OpenAI para acesso aos modelos de LLMs e de Embeddings.

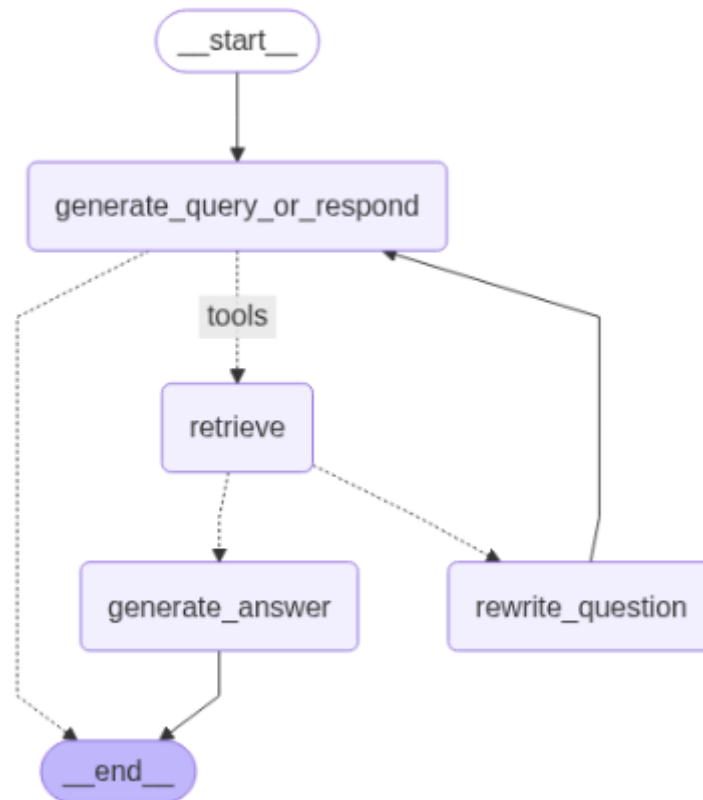
A arquitetura é composta por um grafo principal, com estrutura semelhante à de um supervisor, que invoca um subgrafo responsável pela lógica do RAG agêntico (`agentic_rag`).

A camada de supervisor tem como principal função avaliar se a pergunta deve ou não ser respondida, com base no contexto da Hotmart. Dessa forma, ela atua como uma camada de segurança, evitando respostas inadequadas ou fora de escopo.

Além disso, a solução conta com uma funcionalidade de HITL (Human-in-the-Loop), permitindo que a resposta final passe por uma validação humana (por meio de um interrupt no fluxo). Caso a resposta seja reprovada, o sistema recomenda o encaminhamento para atendimento humano.

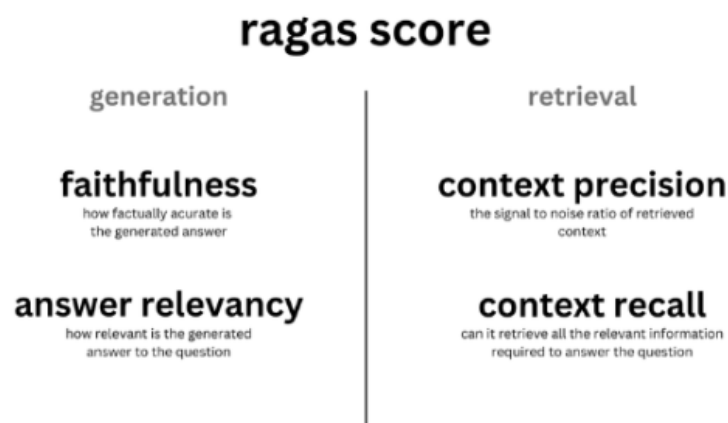


O subgrafo do RAG agêntico utiliza, quando necessário, uma ferramenta de recuperação (retriever tool) para buscar artigos na base de FAQs. Além disso, ele avalia a qualidade dos documentos recuperados e verifica se há necessidade de refinar a consulta utilizada na busca vetorial, garantindo assim uma resposta mais precisa e contextualizada.



## Métricas

Como uma primeira camada de métricas, podemos utilizar o perímetro de avaliação do Ragas, o qual propõe as seguintes métricas:



<https://docs.ragas.io/en/stable/>

Uma abordagem mais refinada seria utilizar o HITL (Human-in-the-Loop) para computar uma métrica de qualidade associada aos documentos recuperados na busca vetorial. Por meio da validação humana, é possível analisar individualmente cada resultado e verificar se o conteúdo de fato atende às necessidades do usuário. Esse feedback permite construir um ranqueamento realista da relevância dos documentos no processo de recuperação.

Uma métrica interessante para esse formato seria o Mean Reciprocal Rank (MRR). Mais detalhes [aqui](#).

Obs.: Precision também seria interessante.