

Nome: Fabio Grassiotto RA: 890441

IA024 - Aula 9\_10

Leitura do Artigo “10\_11 RAGAS Automated Evaluation of Retrieval Augmented Generation”

### Resumo do Artigo:

Este artigo descreve um framework para avaliação de pipelines de RAG denominado **RAGAS** (*Retrieval Augmented Generation Assessment*). É um artigo curto, que descreve as principais funcionalidades do framework e os resultados de métricas obtidas com os experimentos executados.

Ressalto aqui as principais seções do artigo:

- Descrição do problema a ser resolvido, ou seja, a falta de sistemas automatizados para a avaliação dos pipelines de RAG, em especial quando se lida com modelos de código proprietário. Os autores ainda ressaltam a integração do sistema com outros frameworks de automação de engenharia de prompts, como LlamalIndex e Langchain.
- Uma breve descrição de trabalhos anteriores, destacando o uso de LLMs para estimativas de “*faithfulness*” e outros sistemas de avaliação automática, como GPTScore.
- Uma descrição de três das principais métricas disponíveis no framework, incluindo:
  - **Fidelidade (*Faithfulness*)**: O quanto uma resposta está baseada no contexto disponível para uma LLM.
  - **Relevância da Resposta (*Answer Relevance*)**: O quanto uma resposta gerada deve endereçar a questão que está sendo respondida.
  - **Relevância do Contexto utilizado (*Context Relevance*)**: O quanto de informação irrelevante o contexto providenciado contém.
- A metodologia e os experimentos executados com o framework, utilizando um dataset criado especificamente para este trabalho, o dataset WikiEval.
- Na seção de conclusões é ressaltada a criação do dataset mencionado anteriormente e as vantagens do uso do framework descrito no artigo.