

ANÁLISE ESTRUTURADA DAS ETAPAS DE UM SISTEMA RAG PARA GERAÇÃO DE RESPOSTAS ODS (9)

Camila Esmeris Santos Ribeiro (Universidade de Taubaté)
João Gabriel Almeida Rosa (Universidade de Taubaté)
Luís Fernando Almeida (Universidade de Taubaté)

Os avanços recentes na Inteligência Artificial têm possibilitado o desenvolvimento de sistemas cada vez mais capazes de gerar respostas precisas, contextualizadas e fundamentadas em dados relevantes. Nesse cenário, destaca-se o modelo RAG (Retrieval-Augmented Generation), cuja principal contribuição consiste na integração entre técnicas de recuperação de informação e modelos de geração de linguagem natural. Diferentemente de abordagens puramente generativas, que podem incorrer em respostas imprecisas ou fabricadas, o RAG utiliza dados externos como fonte de conhecimento, reduzindo o risco de inconsistências e aumentando a confiabilidade dos resultados apresentados. O presente trabalho teve como objetivo analisar de forma estruturada as principais etapas que compõem um sistema RAG, desde a formulação da consulta inicial até a produção da resposta final. O método de pesquisa adotado foi exploratório, fundamentado em revisão bibliográfica de artigos científicos, relatórios técnicos e estudos de caso que descrevem implementações práticas em diferentes domínios. Entre as etapas analisadas, destacam-se: (i) interpretação da consulta do usuário, (ii) recuperação de documentos relevantes a partir de bases atualizadas, (iii) pré-processamento e seleção dos conteúdos mais adequados e (iv) geração da resposta textual por meio de modelos de linguagem natural. Os resultados da análise indicaram que a etapa de recuperação exerce influência direta na qualidade da resposta, sendo determinante o uso de estratégias de ranqueamento eficazes e fontes de dados confiáveis. A etapa de geração, por sua vez, demonstrou maior coerência e contextualização quando alimentada com informações de qualidade oriundas da recuperação. Além disso, observou-se que a arquitetura modular da RAG permite ajustes e otimizações em cada etapa de forma independente, favorecendo a evolução contínua do sistema. Como conclusão, ressalta-se que o modelo RAG se apresenta como solução promissora em áreas como atendimento ao cliente, educação, saúde, suporte à decisão corporativa e sistemas de recomendação. Ainda que desafios como escalabilidade, custos computacionais e dependência de bases constantemente atualizadas persistam, a separação clara das etapas do processo garante maior controle sobre a qualidade das respostas, reforçando o potencial da RAG no desenvolvimento de soluções inteligentes e aplicadas em contextos reais.

Palavras-chave: RAG; Recuperação de informações; Geração de texto; Inteligência artificial; Sistemas inteligentes.