

Retrieval-Augmented Generation (RAG) is a powerful framework that combines the strengths of information retrieval systems with generative language models to produce more accurate, contextually relevant, and up-to-date responses. Unlike traditional language models that rely solely on pre-trained knowledge, RAG dynamically retrieves relevant documents from an external knowledge base, such as a vector database or a document corpus, during inference. This enables the model to ground its answers in real...