

OpenScholar Results

Question 1

What is retrieval-augmented generation?

Answer

Retrieval-Augmented Generation (RAG) is an approach that enhances large language models (LLMs) by retrieving information from external knowledge resources during the text generation process [0]. This method actively decides when and what to retrieve across the course of the generation, allowing LLMs to gather information continuously throughout the process. RAG has been shown to be effective in mitigating hallucinations and improving performance on various tasks such as question answering, text summarization, and more [1][2][3][6]. There are different types of RAG methods, including dynamic retrieval-augmented generation (DRAG) [2], which actively decides when and what to retrieve during the text generation process, and parametric retrieval-augmented generation (PRAG), which rethinks how the retrieved knowledge should be incorporated into the LLM [1]. Other approaches include structured RAG (SRAG) [4], which systematically organizes extracted entities into relational tables for better analysis, and DioR [6], an adaptive cognitive detection and contextual retrieval optimization method that determines when retrieval is needed and what to retrieve. Additionally, there are methods like GraphFlow [9] that efficiently retrieve accurate and diverse knowledge from text-rich knowledge graphs. Overall, RAG has shown great potential in enhancing the performance of LLMs by providing them with external knowledge during generation.

References

[0] Active Retrieval Augmented Generation □ - Citations: 521 - Despite the remarkable ability of large language models (LMs) to comprehend and generate language, they have a tendency to hallucinate and create factually inaccurate output. Augmenting LMs by retriev...

[1] Dynamic and Parametric Retrieval-Augmented Generation □ - Citations: 9 - Retrieval-Augmented Generation (RAG) has become a foundational paradigm for enhancing large language models (LLMs) with external knowledge, playing an important role in modern information retrieval an...

[2] DRAGIN: Dynamic Retrieval Augmented Generation based on the Real-time Information Needs of Large Language Models □ - Citations: 54 - Dynamic retrieval augmented generation (RAG) paradigm actively decides when and what to retrieve during the text generation process of Large Language Models (LLMs). There are two key elements of this ...

[3] Understand What LLM Needs: Dual Preference Alignment for Retrieval-Augmented Generation □ - Citations: 37 - Retrieval-augmented generation (RAG) has effectively mitigated the hallucination problem of large language models (LLMs). However, the difficulty of aligning the retriever with the LLMs' diverse knowl...

[4] SRAG: Structured Retrieval-Augmented Generation for Multi-Entity Question Answering over Wikipedia Graph □ - Citations: 5 - Multi-entity question answering (MEQA) poses significant challenges for large language models (LLMs), which often struggle to consolidate scattered information across multiple documents. An example qu...

[5] A Systematic Exploration of Knowledge Graph Alignment with Large Language Models in Retrieval Augmented Generation - Citations: 3 - Retrieval Augmented Generation (RAG) with Knowledge Graphs (KGs) is an effective way to enhance Large Language Models (LLMs). Due to the natural discrepancy between structured KGs and sequential LLMs,...

[6] DioR: Adaptive Cognitive Detection and Contextual Retrieval Optimization for Dynamic Retrieval-Augmented Generation □ - Citations: 3 - Dynamic Retrieval-augmented Generation (RAG) has shown great success in mitigating hallucinations in large language models (LLMs) during generation. However, existing dynamic RAG methods face signific...

[7] Creating Conversational Datasets for Retrieval-Augmented Generation Applications is Hard: Challenges & Research Opportunities - Citations: 2 - Retrieval-augmented generation (RAG) has been proven to help mitigate hallucinations from large language models (LLMs). However, as more domains adopt this method, the need for human-created conversat...

[8] Quantitative Evaluation of Using Large Language Models and Retrieval-Augmented Generation in Computer Science Education - Citations: 2 - Generative artificial intelligence (GenAI) is transforming Computer Science education, and every instructor is reflecting on how AI will impact their courses. Instructors must determine how students m...

[9] Can Knowledge-Graph-based Retrieval Augmented Generation Really Retrieve What You Need? □ - Citations: 1 - Retrieval-Augmented Generation (RAG) based on knowledge graphs (KGs) enhances large language models (LLMs) by providing structured and interpretable external knowledge. However, existing KG-based RAG ...
