AGENTIC RAG

Agentic Retrieval-Augmented Generation (RAG) is an AI technique that uses retrieval-based data access to improve response accuracy and relevance in language models. It introduces an autonomous component, allowing the system to dynamically determine information requirements, retrieve it, and refine output through an iterative process. This approach enables complex, multi-step queries requiring deeper reasoning and context awareness.

Key aspects of Agentic RAG include dynamic workflows, real-time integration, multi-step reasoning, scalability, improved accuracy, complex query handling, and multimodal capabilities. The system works by analyzing user queries, determining processing paths, data retrieval, context building, response generation, and an iterative feedback loop.

Agentic RAG is particularly useful for applications like customer support, legal research, scientific analysis, and autonomous research assistants. It is closely tied to frameworks like LangChain, AutoGPT, or OpenAI Functions, which help orchestrate agent behaviors. The system's ability to dynamically query, process, and update information makes it more robust, interactive, and adaptable than traditional RAG.

In summary, Agentic RAG empowers language models with agent-like capabilities to perform tasks more intelligently, flexibly, and autonomously by combining retrieval, reasoning, and tool use in a closed feedback loop.