✦ opsloom

Menu   On this page

What Is a Retrieval-Augmented Generation?

Retrieval-Augmented Generation (RAG) is a technique for improving the accuracy and reliability of generative AI models by supplementing them with facts from external sources. The method was pioneered by researchers who demonstrated that it improved AI performance and reduced hallucinations relative to non-RAG AI systems.

What Is a Knowledge Base?

A knowledge base is a repository of information accessible to RAG AI applications for retrieving domain-specific data. When the user submits a query to the RAG AI, it is compared against different portions of the knowledge base until a match is found. This match is then passed along to the LLM, which uses it as context for its answer to the user.

As an example, say you submit an employee manual to a knowledge base for an internal HR chatbot assistant. When an employee asks the chatbot a question about their vacation policy, first the correct chunk of the knowledge base is retrieved. This could be a paragraph explaining how vacation hours are accrued. The correct portion of the knowledge base is then supplied to the LLM ahead of its response so that it can give the appropriate answer based on the employee manual, rather than generally available information.

What Kinds of Documents Are Accepted in the Knowledge Base?

Opsloom currently only supports PDF documents for its knowledge bases.

How Accurate Are the Responses From the AI Assistant?

The assistant's accuracy will vary based on several factors:

- The accuracy and comprehensiveness of the information in the knowledge base
- The wording in the queries to the assistant

RAG AI models have much better recall than non-RAG AI models—in other words, they are more likely to retrieve facts that are relevant to the user's query. However, this improvement varies by domain and implementation. Work with the Opsloom team to see how you can track your assistant's accuracy and determine how well RAG fits your use case.

What Happens if the AI Assistant Cannot Find an Answer in the Documents?

If the assistant cannot find relevant information in the knowledge base, it can either provide a default message such as "I don't have that information" or redirect the user to contact support. You can work with the Opsloom team to come up with the response that works best for you.

Is My Data Secure When Using Opsloom?

Yes! The Opsloom team will work within your organization's infrastructure to set up Opsloom. Everything can be hosted on your domains and servers. This ensures that you can use your existing security infrastructure in tandem with Opsloom.

Previous page
[Text to SQL (AWS Cost Calculator)](#)