

# RETRIEVAL-AUGMENTED GENERATION (RAG) AND AGENTIC RAG

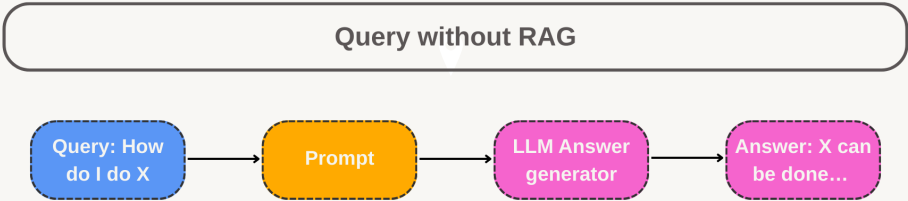
Enterprises want generative AI to reason over their private data and not over the training data of the AI model. This is where Retrieval Augmented Generation (RAG) comes into play. RAG enhances the capabilities of generative AI by incorporating relevant data from enterprise private data, thereby improving the quality and reliability of the responses.

1

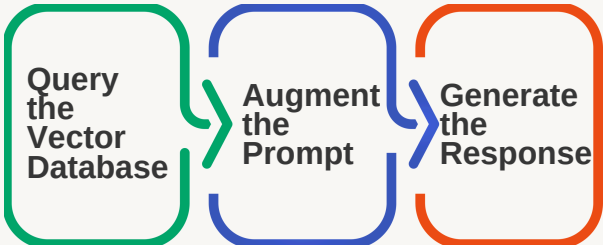
## World Before RAG

Before RAG, generative AI relied on costly, time-consuming training and fine-tuning of models using large datasets. **This approach required continuous retraining for new tasks, leading to inefficiencies, higher costs, and limited adaptability to rapid changes.**

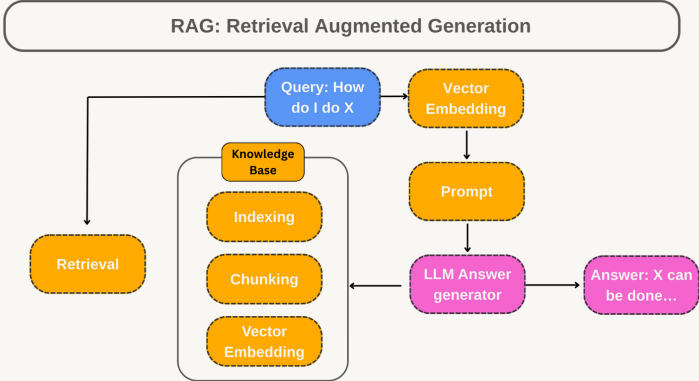
### A typical pipeline without RAG



## RAG process



### Pipeline of a traditional RAG



2

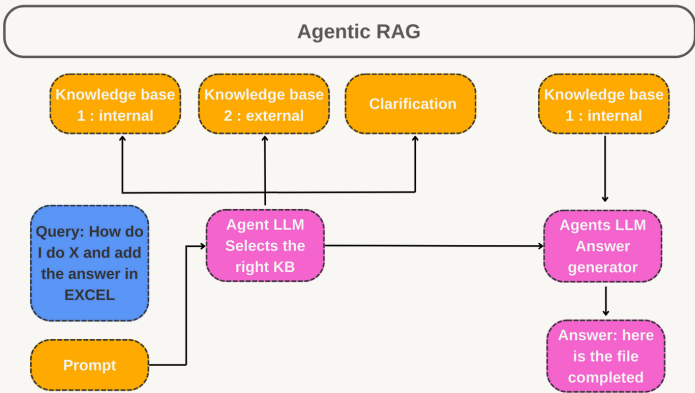
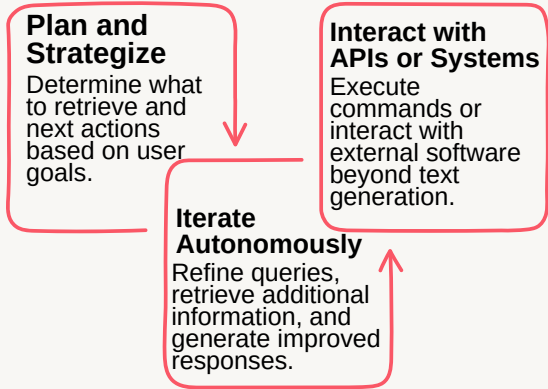
## Then comes the RAG

A process enabling LLMs to access external data for accurate, context-specific responses. It enhances customer interactions by retrieving information from knowledge bases and personalizing support. RAG also streamlines workflows by prepopulating documents with relevant enterprise data, boosting efficiency. However, they often struggle to manage and prioritize high quality information from large datasets, leading to diminished performance.

3

## Now it’s Agentic RAG’s turn: Unique Factors of Agentic RAG

**Agentic RAG takes the concept of RAG a step further by introducing an intelligent decision-making layer.** Instead of merely generating responses based on retrieved data, Agentic RAG **leverages the capabilities of the language model (LLM)** to make informed decisions. **Unlike traditional RAG, Agentic RAG enables systems to:**



## Agentic RAG key use cases

### Decision Making

Agentic RAG's dynamic capabilities allow organizations to make informed decisions based on real-time data analysis. This is particularly valuable in industries such as finance and logistics, where timely information is critical for operational success

### Content Creation

In the realm of content crafting, Agentic RAG streamlines the creation process by tailoring content to specific user needs. This capability allows businesses to produce high-quality, relevant content quickly, enhancing marketing and communication strategies.

### Customer Support

Agentic RAG significantly improves customer support by sourcing product information and customer history to generate personalized responses. This leads to enhanced efficiency and quality in customer interactions, allowing support teams to resolve issues more effectively

## Benefits of Agentic RAG

The implementation of Agentic RAG offers several advantages over traditional RAG systems:

- ✔ **Scalability and Extensibility:** Modular architecture allows easy integration of new data sources and tools.
- ✔ **Enhanced User Experience:** Faster response times, personalized information retrieval, and simplified complex tasks improve user interactions.
- ✔ **Intelligent Quality Control:** Agents verify and correct information, ensuring accuracy and reliability.
- ✔ **Enhanced Retrieval Techniques:** Advanced algorithms and semantic caching improve search precision and reduce computational costs.
- ✔ **Dynamic Planning and Execution:** Agents can adapt to complex queries and evolving information landscapes.
- ✔ **External Tool Integration:** Utilizes various external resources to enhance information gathering.