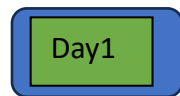


Build RAG Agents with LLM



What is RAG?

RAG Stands for Retrieval-Augmented Generation

Retrieval -> Finding Relevant information from your Personal Data

Augmented -> Adding this information to the original question

Generation -> Creating a Personalized Answer using the enhanced prompt

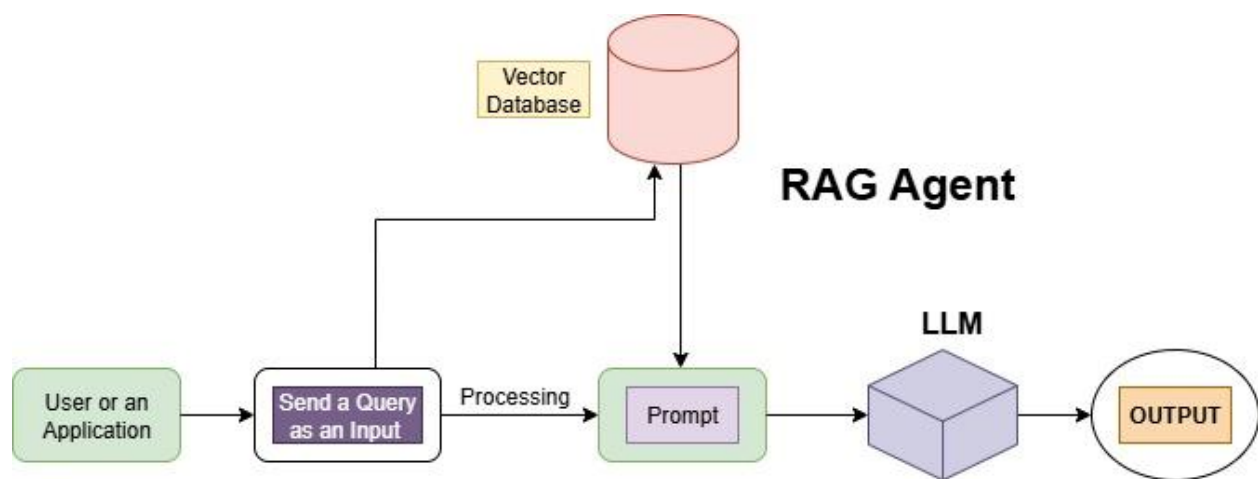


Figure 1: Flowchart RAG Agent

Description

Let's say we are using an LLM (like ChatGPT) and we send a query such as ***What is my CGPA right now, and which semester am I in?***

After processing, this becomes the prompt that is sent to the LLM, which then generates a response like:

I don't know what your CGPA is or which semester you are in, as I don't have access to that information.

This highlights one of ChatGPT's limitations: it does not know our personal information.

It doesn't have access to our personal notes, projects, attendance records, resumes, education files, etc.

That's why it can only answer generic queries and cannot respond to questions about our personal data.

RAG (Retrieval-Augmented Generation) acts as a bridge between our personal information and Chatbot.

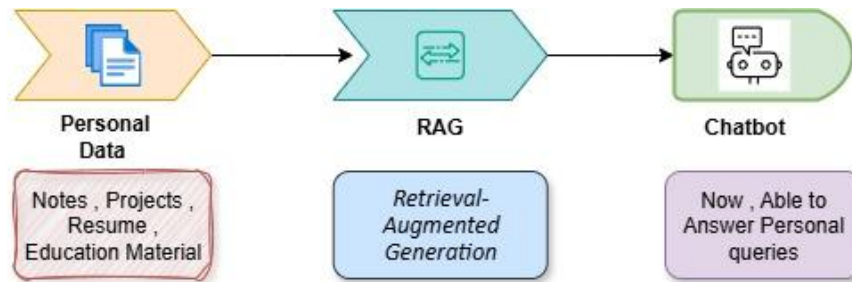


Figure 2 : Bridge RAG

End of Day 1