

Using RAG for business

RAG may help to give more details about specific topic that user is searching for.

This extra information plus the original response from LLM (Large Language Model. Example: OpenAI) can be useful at times.

However, this technique should be used in case it is really necessary. Because, instead of giving more clearness about a topic, this could create more confusion about it.

On the other hand, you have the choice to create your own architecture to recover extra information and attach it to LLM response. This could speed up the interaction between user prompt and LLM.

If you do not have much idea about creating your own architecture. Then you can start by analyzing other RAG architectures or software architectures. Then, you will have a better idea about how to create yours.

RAG may be used as well to give solutions or recommendations to your own business. You may feed your LLM with confidential and unique information using RAG and let LLM create its own response.