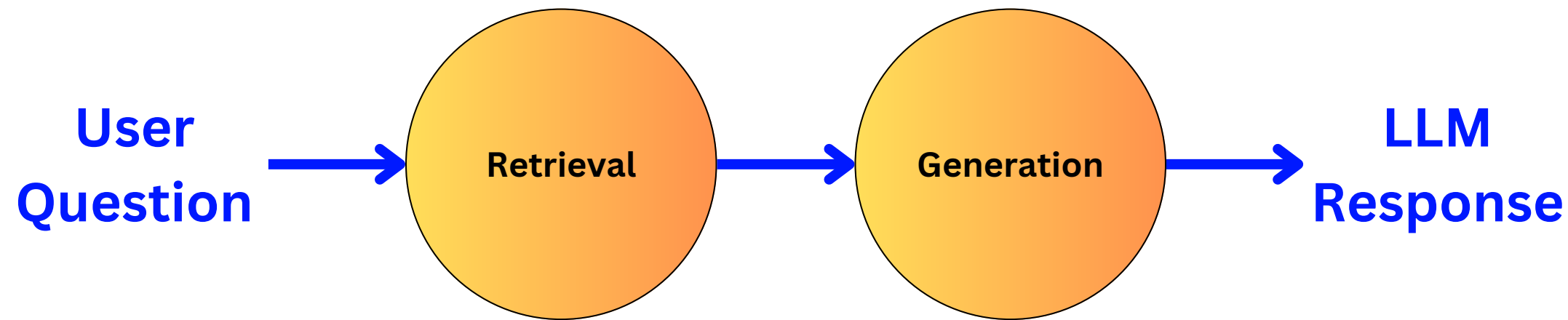


How to Improve a RAG Pipeline?

TheAiEdge.io

A Basic RAG Pipeline is just composed of a Retriever and a Generator



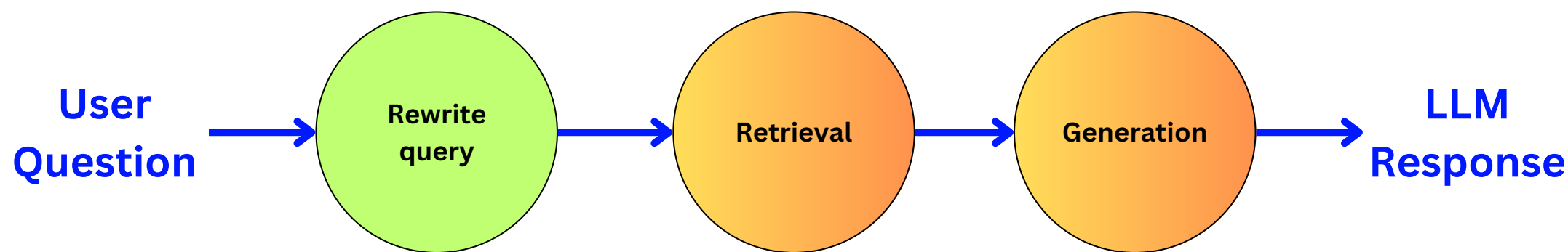
Basic RAG



How to Improve a RAG Pipeline?

TheAiEdge.io

We can improve it by rewriting the user question a better query for the specific data store



We create a new query rewrite module



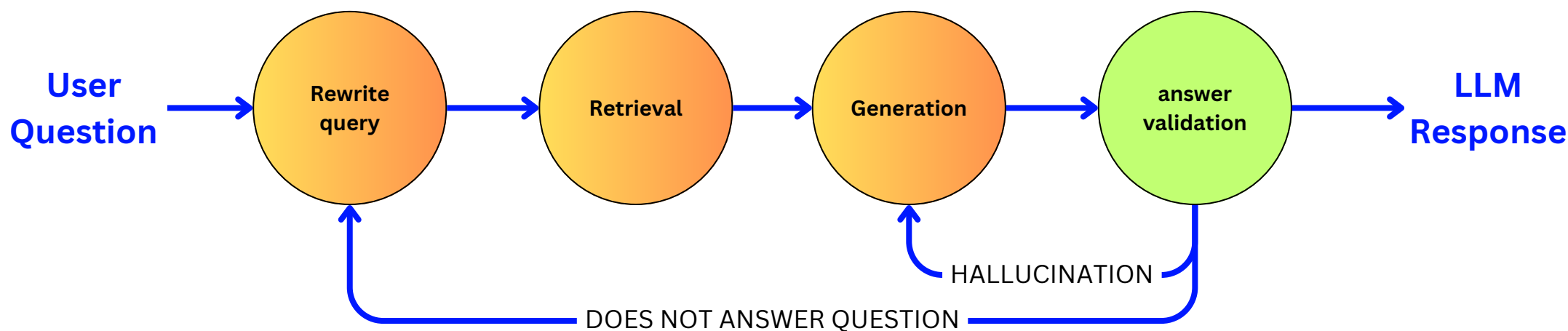
How to Improve a RAG Pipeline?

TheAiEdge.io

Generation can suffer from Hallucination and lack of relevance so it is better to check that:

- The answer is grounded in the retrieved documents
-> If not, we restart the generation process
- The answer is actually relevant to the original question
-> if not, we restart the retrieval process

-> Self-RAG



**We create an answer
validation module**

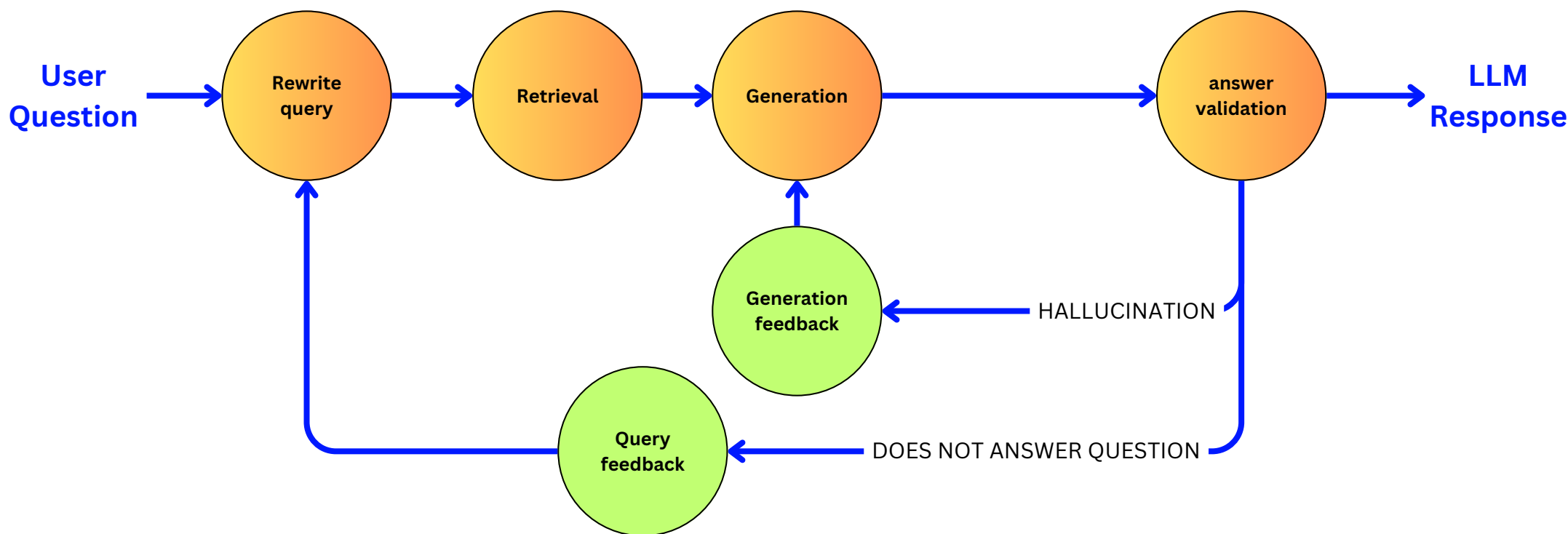
3/8



How to Improve a RAG Pipeline?

TheAiEdge.io

Instead of retrying blindly, it is more efficient to get feedback for why the previous generation was wrong



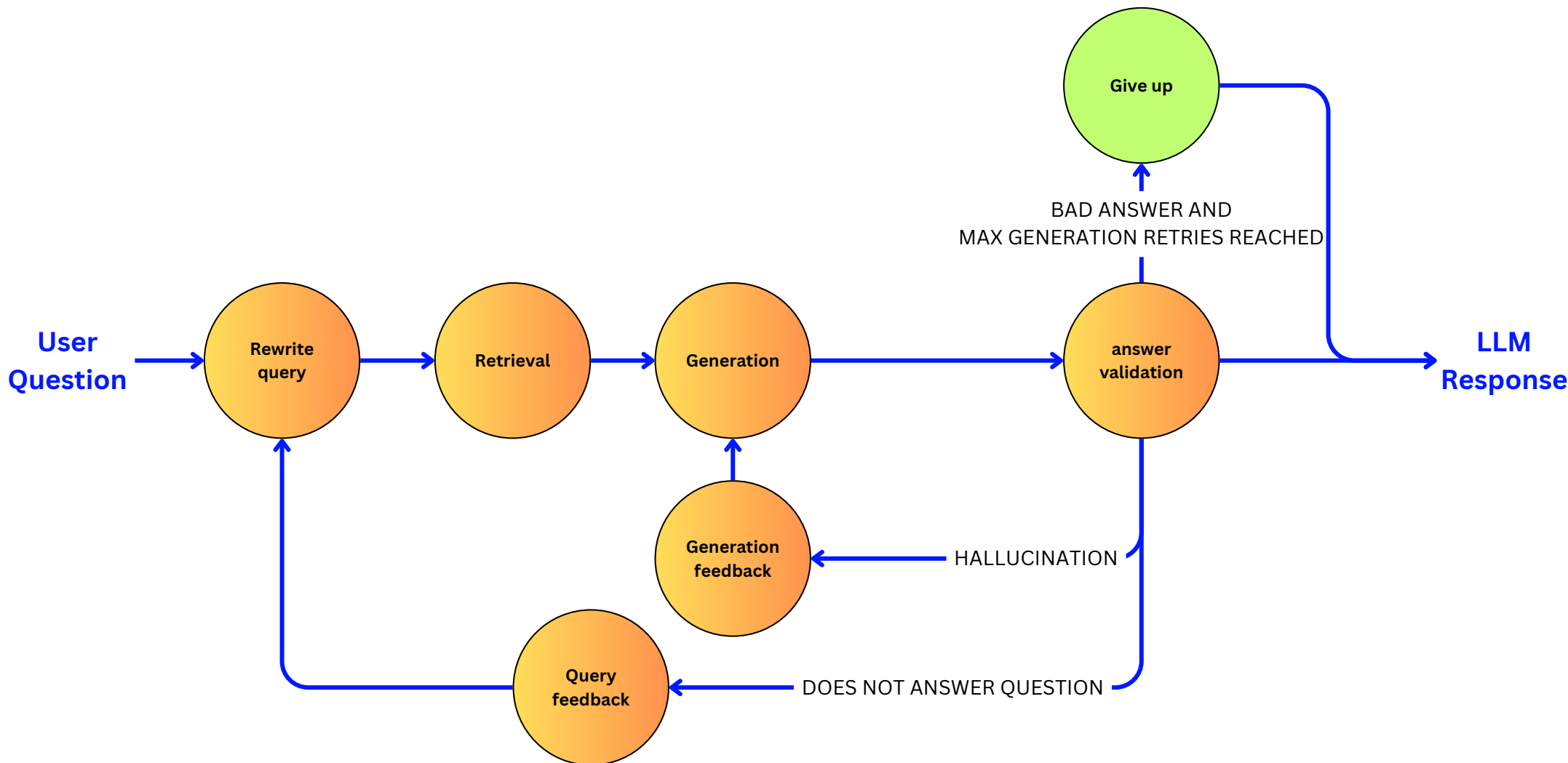
We create feedback modules to rewrite the query and improve generation



How to Improve a RAG Pipeline?

TheAiEdge.io

The retrieval and generation retries can lead to infinite loops



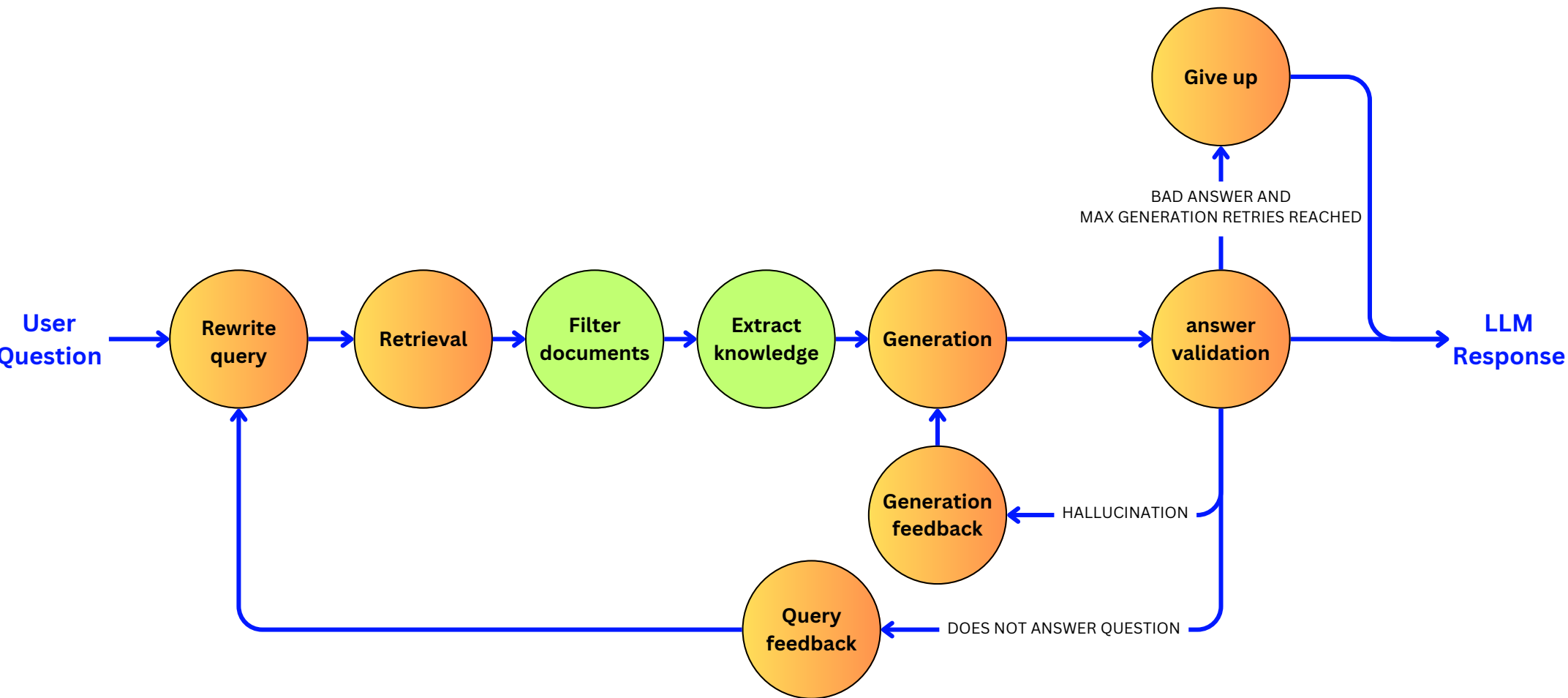
We allow only a maximum amount of retries before it gives up and apologize to the user



How to Improve a RAG Pipeline?

TheAiEdge.io

Some of the retrieved documents can be irrelevant and some of them can contain relevant and irrelevant information



We create modules to filter out the irrelevant documents and extract the relevant information

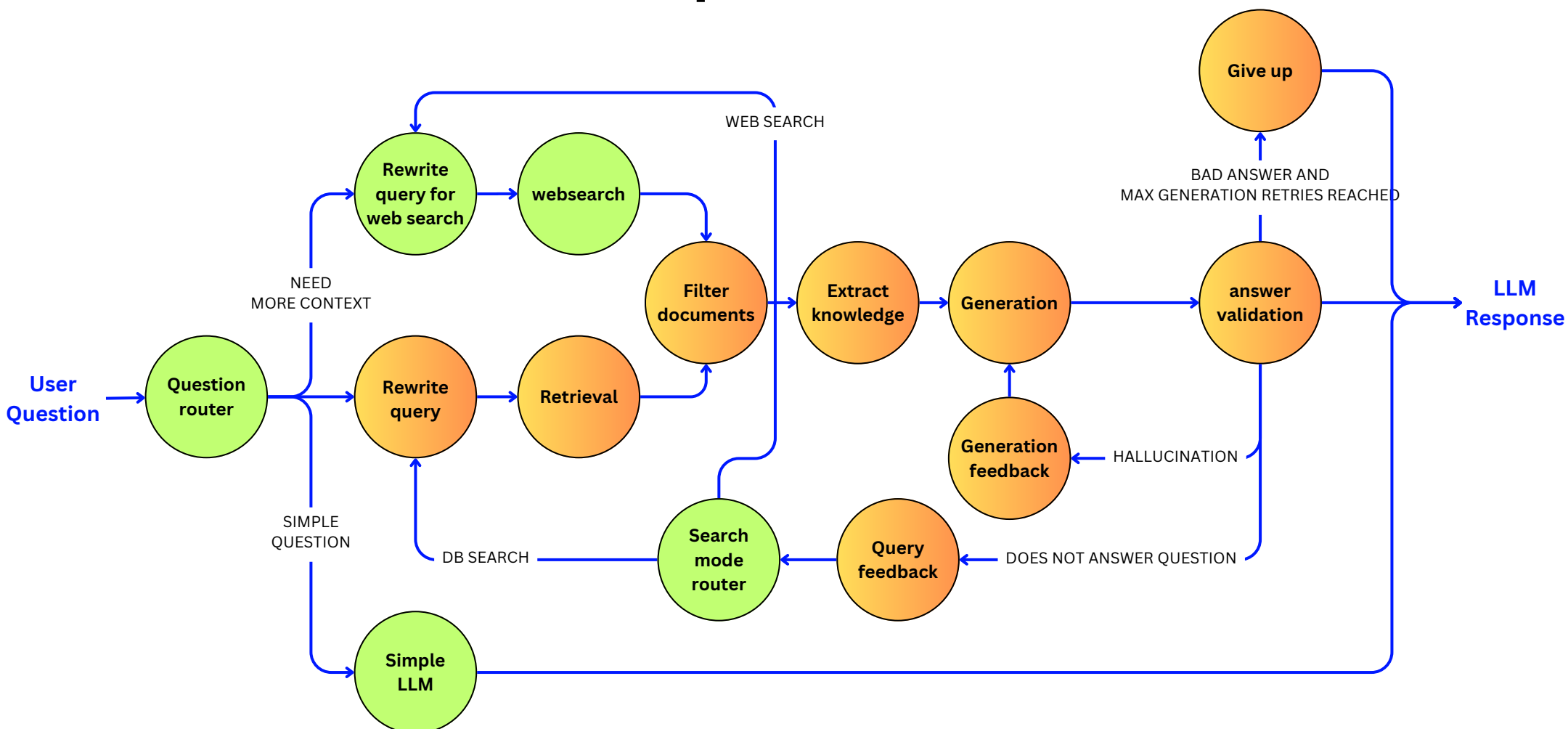


How to Improve a RAG Pipeline?

TheAiEdge.io

Many user questions may not be relevant to the stored data

-> Adaptive RAG



We create a question router to route to a simple LLM, a websearch RAG or a datastore RAG

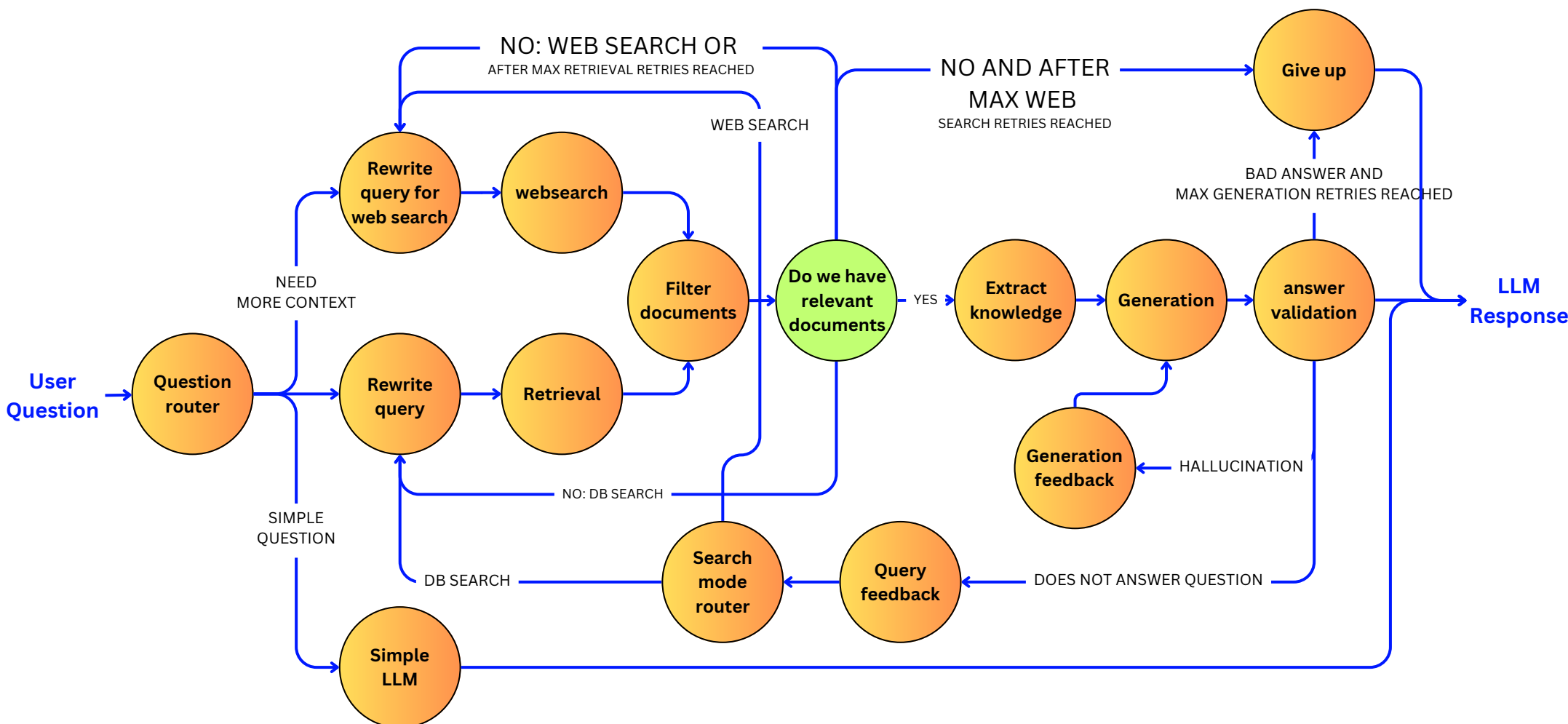


How to Improve a RAG Pipeline?

TheAiEdge.io

Even if routed the question, it may have been a mistake. We need a mechanism to correct wrong routing

-> Corrective RAG



We a module that reroute from datastore RAG to websearch RAG if no relevant documents can be found