

Relevancy

Instructors

Prashant Sahu

Manager - Data Science, Analytics Vidhya

Ravi Theja

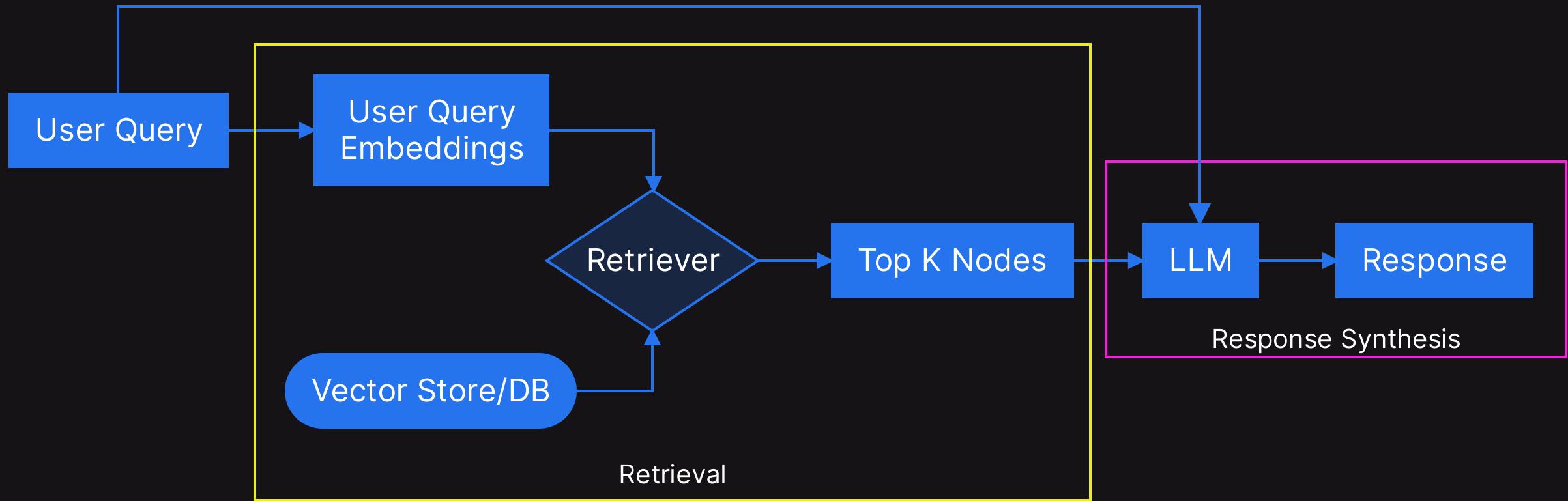
Developer Advocate Engineer, LlamalIndex



Evaluation Metrics

- Retriever Evaluation Metrics
 - Hit rate
 - Mean Reciprocal Rank (MRR)
- Response Evaluation Metrics
 - Faithfulness
 - **Relevancy**
 - Correctness
 - Semantic Similarity
 - Guideline

Retrieval and Synthesis



Relevancy

- Is the generated response as per the retrieved contexts and user query?

Relevancy

- Evaluates if the answers generated are aligned with retrieved context and query

Calculating Relevance Score

The Answer Relevancy is defined as the mean cosine similarity of the original question to a number of artificial questions, which were generated (reverse engineered) based on the answer :

$$\text{Answer Relevancy} = \frac{1}{N} \sum_{i=1}^N \text{Cosine_similarity}(E_{g_i}, E_o)$$

$$\text{Cosine_similarity} = \frac{E_{g_i} \cdot E_o}{\|E_{g_i}\| \|E_o\|}$$

Where:

- E_{g_i} is the embedding of the generated question i .
- E_o is the embedding of the original question.
- N is the number of generated questions, which is 3 default.

- Score Interpretation
 - Lower scores for incomplete or redundant answers, higher scores for better relevancy.
- Alignment Check
 - Accurate answers result in generated questions that align with the original question.

Calculating Relevancy Score: Example

Hint

Question: Where is France and what is its capital?

Low relevance answer: France is in western Europe.

High relevance answer: France is in western Europe and Paris is its capital.

- Steps for calculating relevance of an answer:
- Step 1: Reverse-engineer 'n' variants of the question from the generated answer using a LLM, e.g.
 - Question 1: "In which part of Europe is France located?"
 - Question 2: "What is the geographical location of France within Europe?"
 - Question 3: "Can you identify the region of Europe where France is situated?"
- Step 2: Calculate the mean cosine similarity between the generated questions and the actual question.

Factors to improve relevancy?

- Retriever algorithm
- Chunk size
- Embedding model
- LLM for response generation

Thank You