

# Query Enhancement

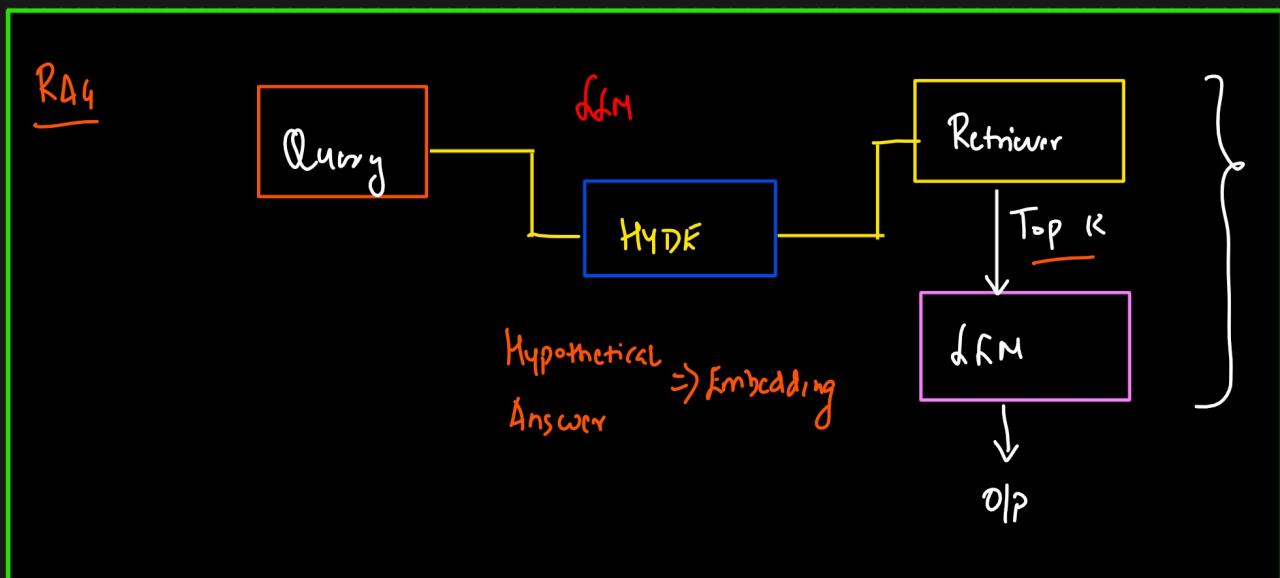
## Hypothetical Document Embeddings (HyDE)

🧠 What is HyDe?

HyDE (Hypothetical Document Embeddings) is a retrieval technique where, instead of embedding the user's query directly, you first generate a hypothetical answer (document) to the query using an LLM — and then embed that hypothetical document to search your vector store.

→ HyDE bridges the gap between user intent and relevant content, especially when:

1. Queries are short
2. Language mismatch between query and documents
3. You want to retrieve based on answer content, not question words





## Why Use HyDE?

Problem	How HyDE Helps
User query and document use different words	Embeds answer-style content instead of question
Query is too vague	LLM-generated answer gives richer semantics
Better grounding needed	Embeds what the answer might look like
Zero-shot retrieval	Works well even without prior training



## Benefits of HyDE

Feature	Why It Helps
Semantic intent modeling	Better than literal keyword matching
LLM-aware retrieval	Query is expanded into more context
Generalization	Works well even if document phrasing differs from question
Plug-and-play	No need for retraining; works with OpenAI, Cohere, Hugging Face