

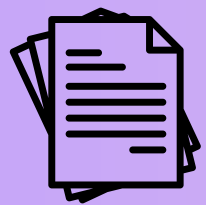
# LLM Lingo: Must-Know Terms

## Part 3: RAG + LLM Agents

Created By: Aishwarya Naresh Reganti

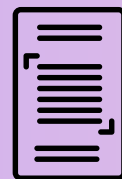
RAG or Retrieval-Augmented Generation in LLMs, combines regular language generation with information retrieval, enabling models to access external knowledge sources and improve the quality and relevance of their generated outputs. Below are some key terms:

### Knowledge Base(KB)



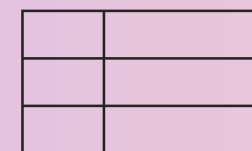
Collection of documents from which relevant information is retrieved in RAG

### Chunking



Breaking the KB into smaller pieces for efficient storage and retrieval during RAG

### Indexing



Organizing and storing KB chunks in a structured manner for efficient retrieval

### Embedding Model



An LLM that converts KB text chunks into numerical format called vectors/embeddings

### Vector Database



Database optimized for storing and retrieving vector representations generated from the KB

### Vector Search



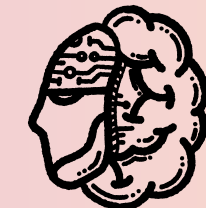
Finding the most relevant KB chunks based on vector similarity scores for a given input query.

### Retrieval



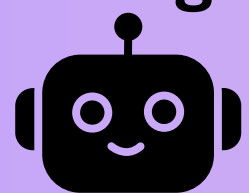
Approach used to rank and fetch KB chunks from the vector search. This will serve as additional context for the LLM.

### AGI



Artificial General Intelligence aims to create machines that can learn and reason like humans across various tasks.

### LLM Agent



LLM applications that can execute complex tasks by combining LLMs with modules like planning and memory

### Agent Memory



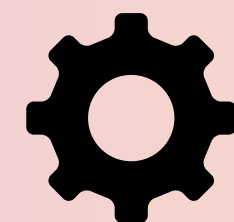
A module that stores the agent's past experiences and interactions with the user and environment.

### Agent Planning



Module that divides the agent's tasks into smaller steps to address the user's request efficiently.

### Function Calling



Ability of LLM agents to request information from external tools and APIs in order to execute a task