What is RAG:
Understanding Retrieval-Augmented Generation

Basic RAG Pipeline

Ingestion
Documents → Chunks → Embeddings → Index

Retrieval
Query → Index → Top K

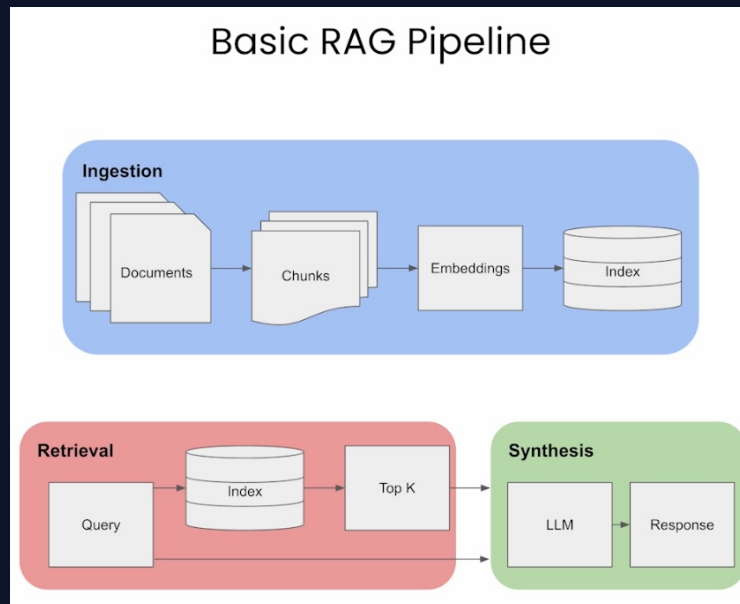**Synthesis**
LLM → Response

Agenda: What is RAG ?
          Why ?
          Importance ?
          Use Case ?
          Benefit ?   } 1

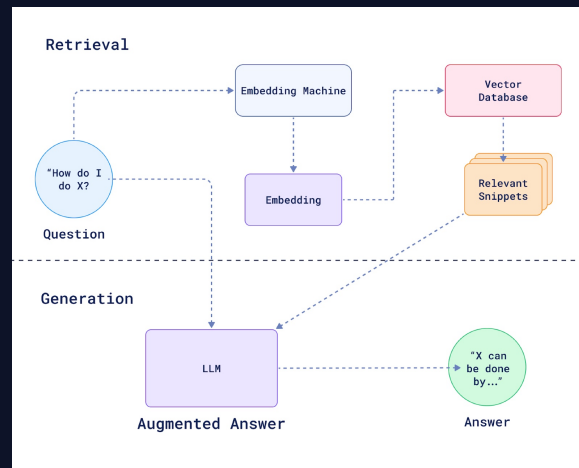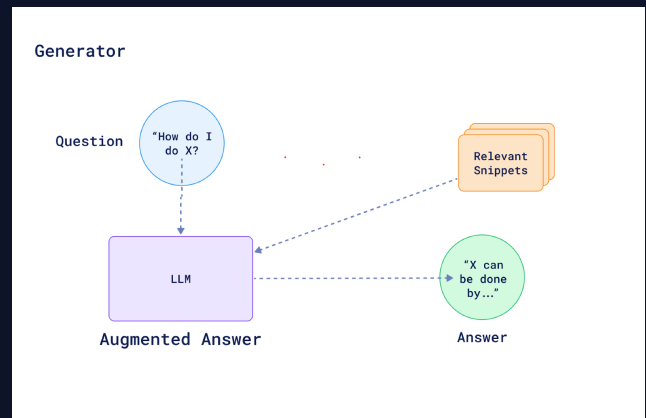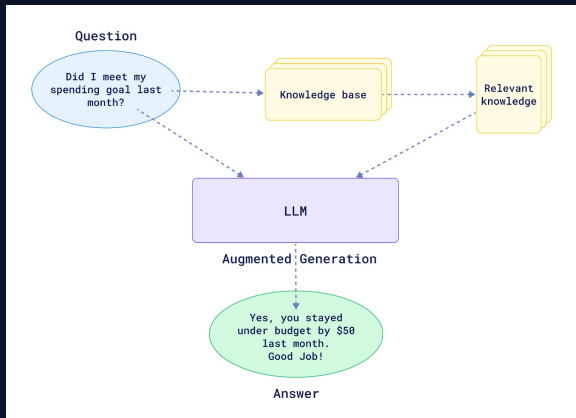   Core concept ?   } - 2
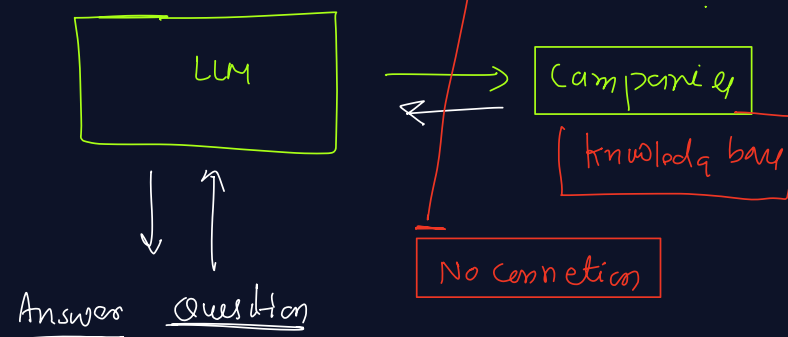   Components   } 3

Challenges .
Limetatecls -
Future Trends   } 4

Implemation — 5

Diagram 1 (top left): **Question** — "Did I meet my spending goal last month?" → **Knowledge base** → **Relevant knowledge** → **LLM** (Augmented Generation) → **Answer**: "Yes, you stayed under budget by $50 last month. Good Job!"

Diagram 2 (top right): **Generator** — **Question** "How do I do X?" → **LLM** (Augmented Answer) ← **Relevant Snippets** → **Answer**: "X can be done by..."

Diagram 3 (center): **Retrieval** — Question "How do I do X?" → Embedding Machine → Embedding → Vector Database → Relevant Snippets. **Generation** — → LLM (Augmented Answer) → Answer: "X can be done by..."

(1). INey : 1)

LLM

↓ Answer   ↑ Question

→ Companies
  (knowledge base)

← (No Connection)

Personal Database.

(2) → [LLM] ⊣ → India Vs Aut → Wo?
↑↑
↘ X (Web search) *
X Updated

(3) ○ Real time Data ⇒ Live - community X

(4) ④ □ → □ → Hallucination (Faulty
Personal   ↓   Not correct)
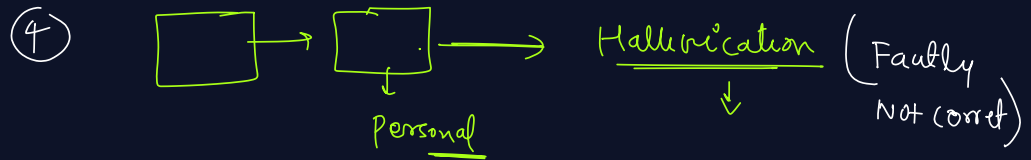
(5) ⌐ Domain specific knowledge.

(what ⁓?

[(Embdi —— Vector)]

→
→
Query

□
↓
[X]
↓
Answer

Knowledge Base

concept

RAG = LLM + Knowledge Base

Vector   Vector data   Component

Steps
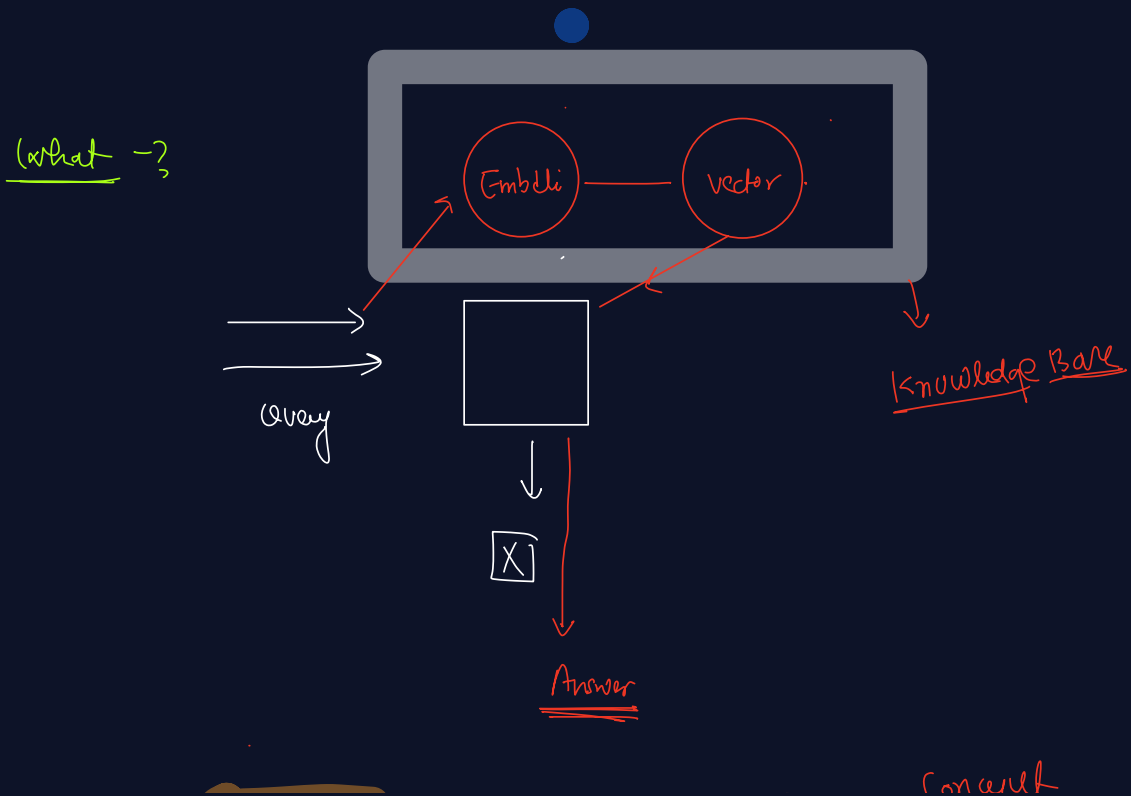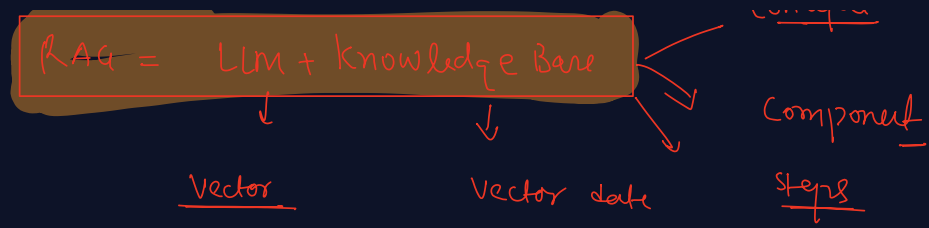
(1) → **Benefit & Use Case** →

1) Domain specific Model

2) Real time data Access

3) Question - Answering.

4) Chatbots for your Database

RAG →: Concepts: