Multimodal RAG systems

See beyond words

AtlasIA
أطلسيـــــــــة

MoroccoAI

Learning Outcomes

You will:

- Understand what is a RAG, when it’s used and its core components (embeddings, vector stores, …)
- Understand multimodality
- Know a few methods to build a multimodal RAG (focus on image and text)

What we won’t cover:

- Transformers 😥
- Maths behind components

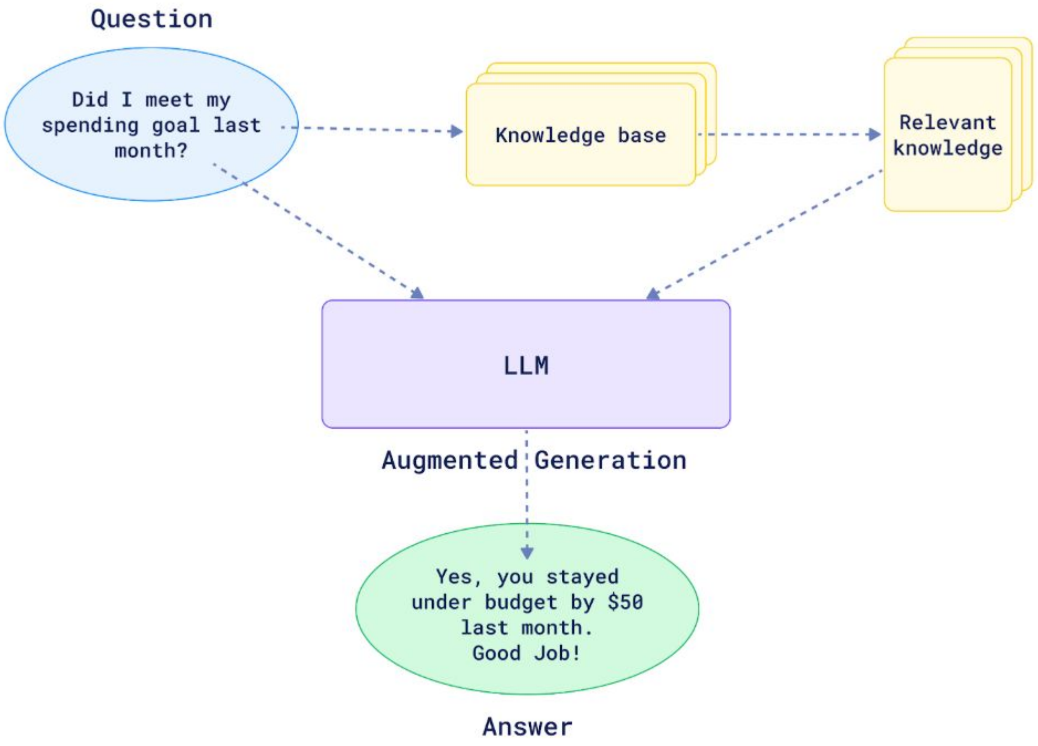
Agenda

- Why use RAG?
- Architecture and core components
- What is multimodality?
- Multimodal RAG design approaches

Retrieval **A**ugmented **G**eneration

- Why and When do we need it?

Why use RAG ?



Question

Did I meet my spending goal last month?

Knowledge base

Relevant knowledge

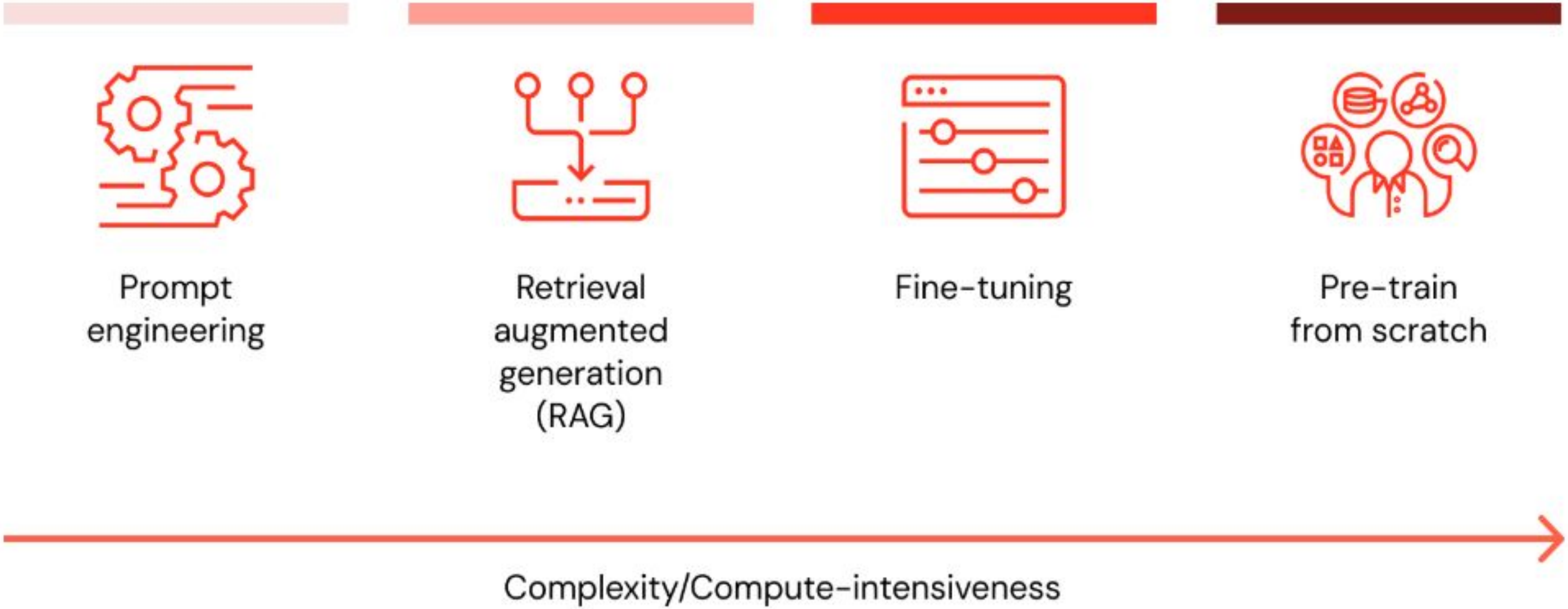
LLM

Augmented Generation

Yes, you stayed under budget by $50 last month. Good Job!

Answer

Why use RAG ?



Prompt engineering

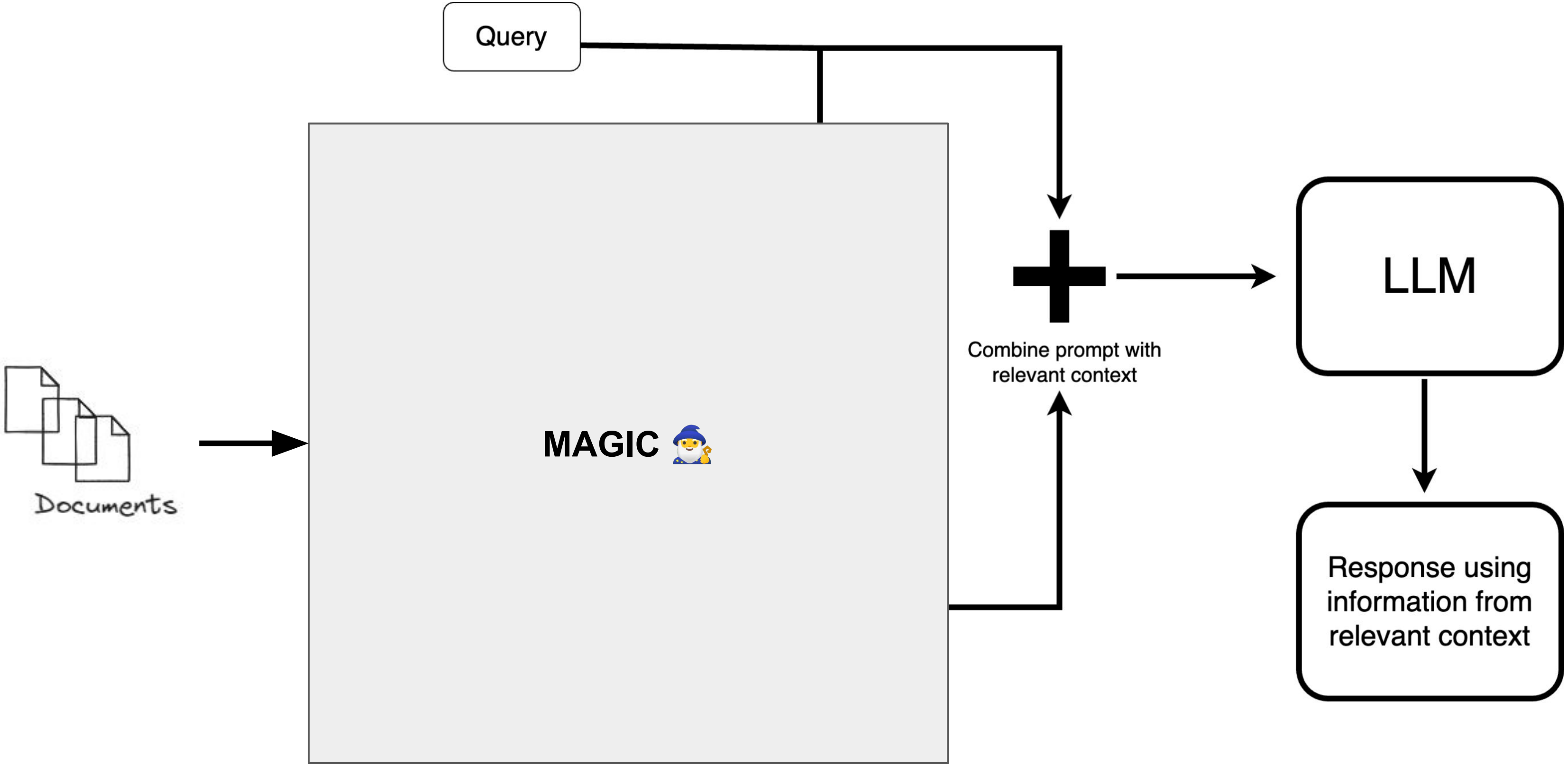
Retrieval augmented generation (RAG)

Fine-tuning

Pre-train from scratch

Complexity/Compute-intensiveness

RAG in a nutshell


Query
Documents
MAGIC 🧙
Combine prompt with relevant context
LLM
Response using information from relevant context

RAG components

Embeddings

Information from documents is stored as vector embeddings. This format supports efficient similarity searches to retrieve relevant data for your query.

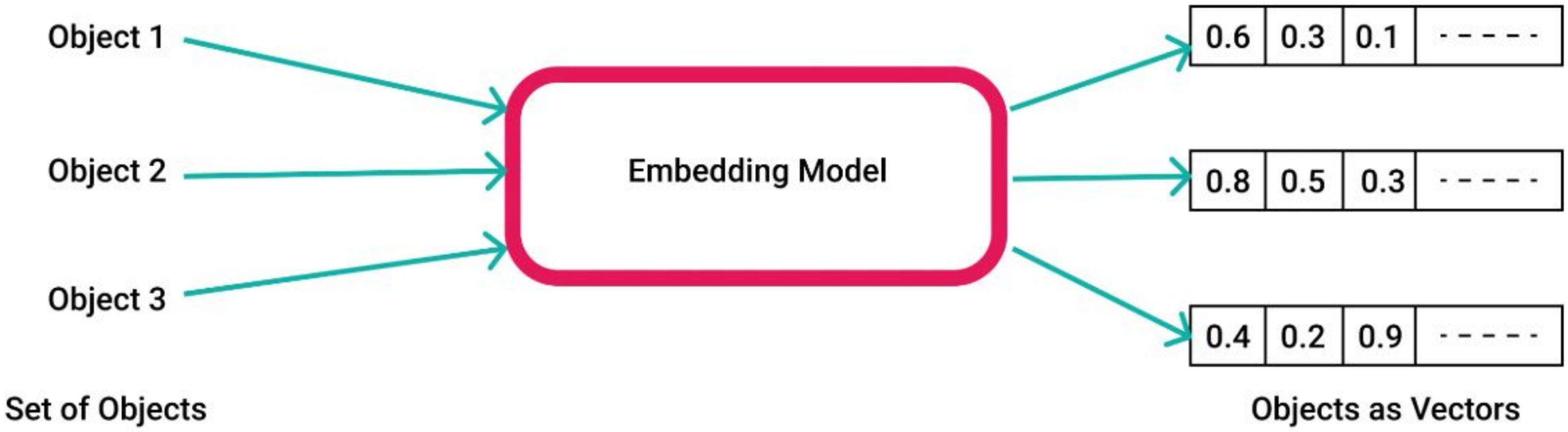
Vector Search

Similarity search is applied on the vector database to retrieve the most relevant documents to the query

LLM

Takes the query augmented by the relevant context and generates the final answer

Embeddings



Object 1

Object 2

Object 3

Set of Objects

Embedding Model

0.6 | 0.3 | 0.1 | - - - - -

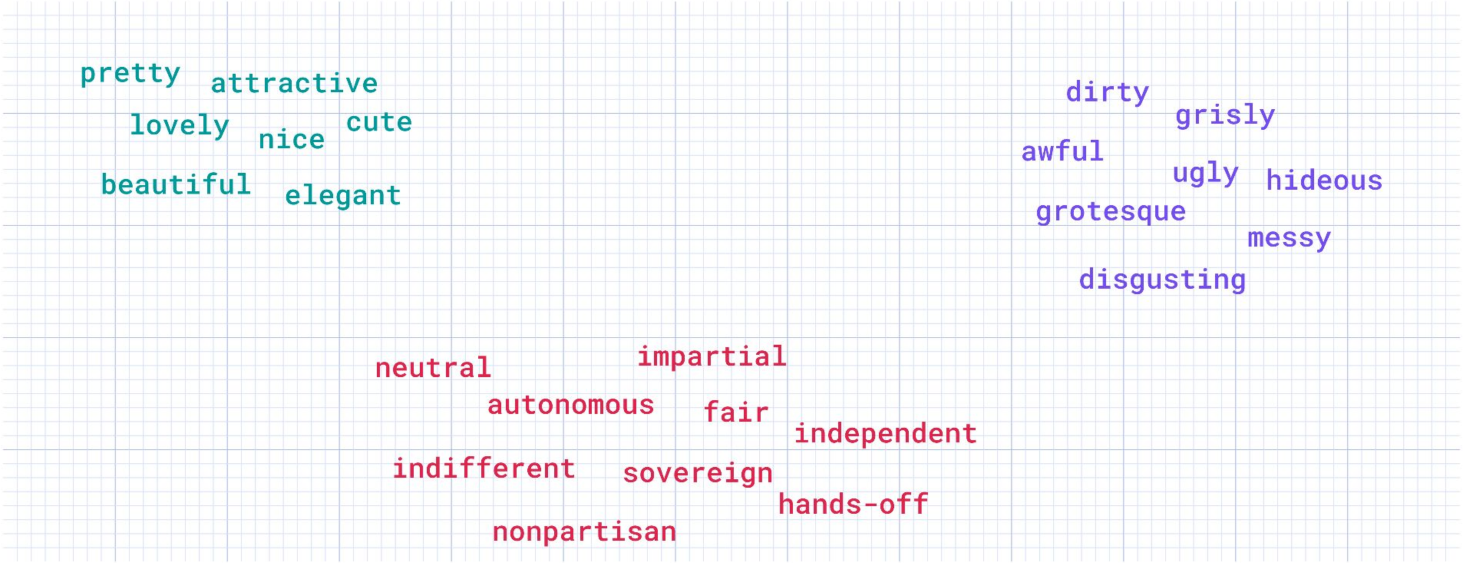
0.8 | 0.5 | 0.3 | - - - - -

0.4 | 0.2 | 0.9 | - - - - -

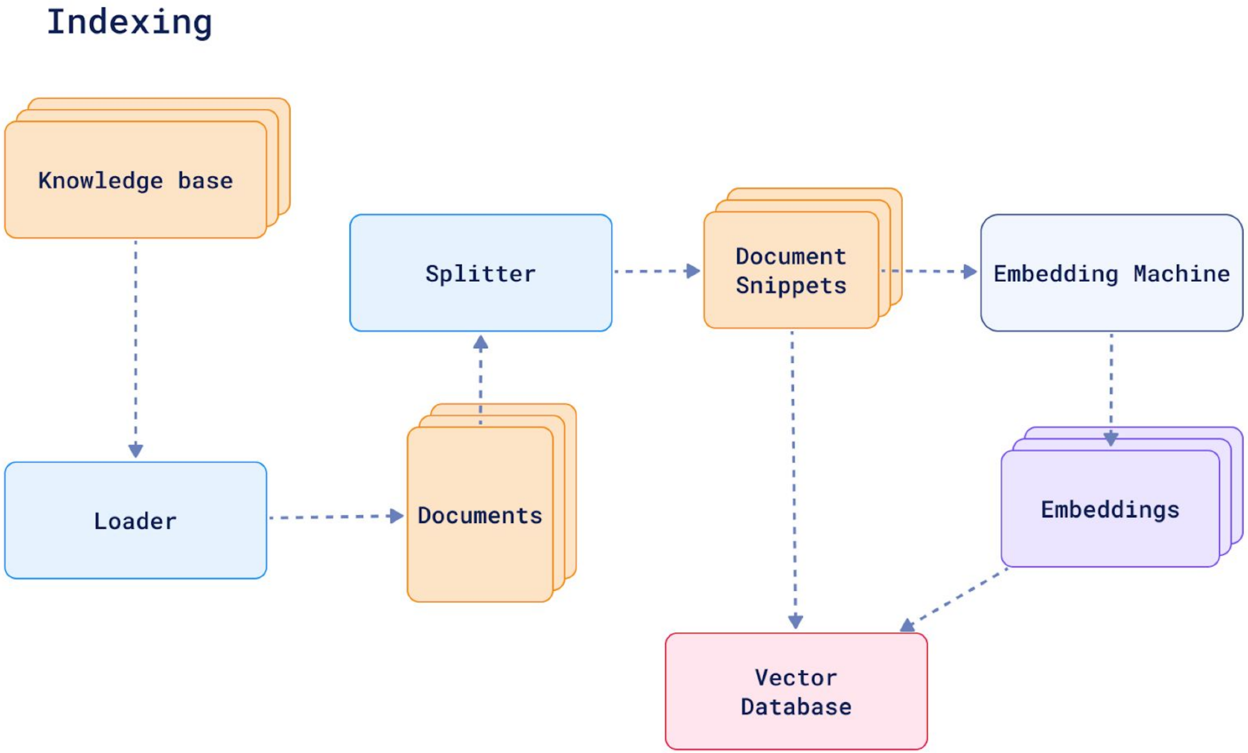
Objects as Vectors

Embeddings

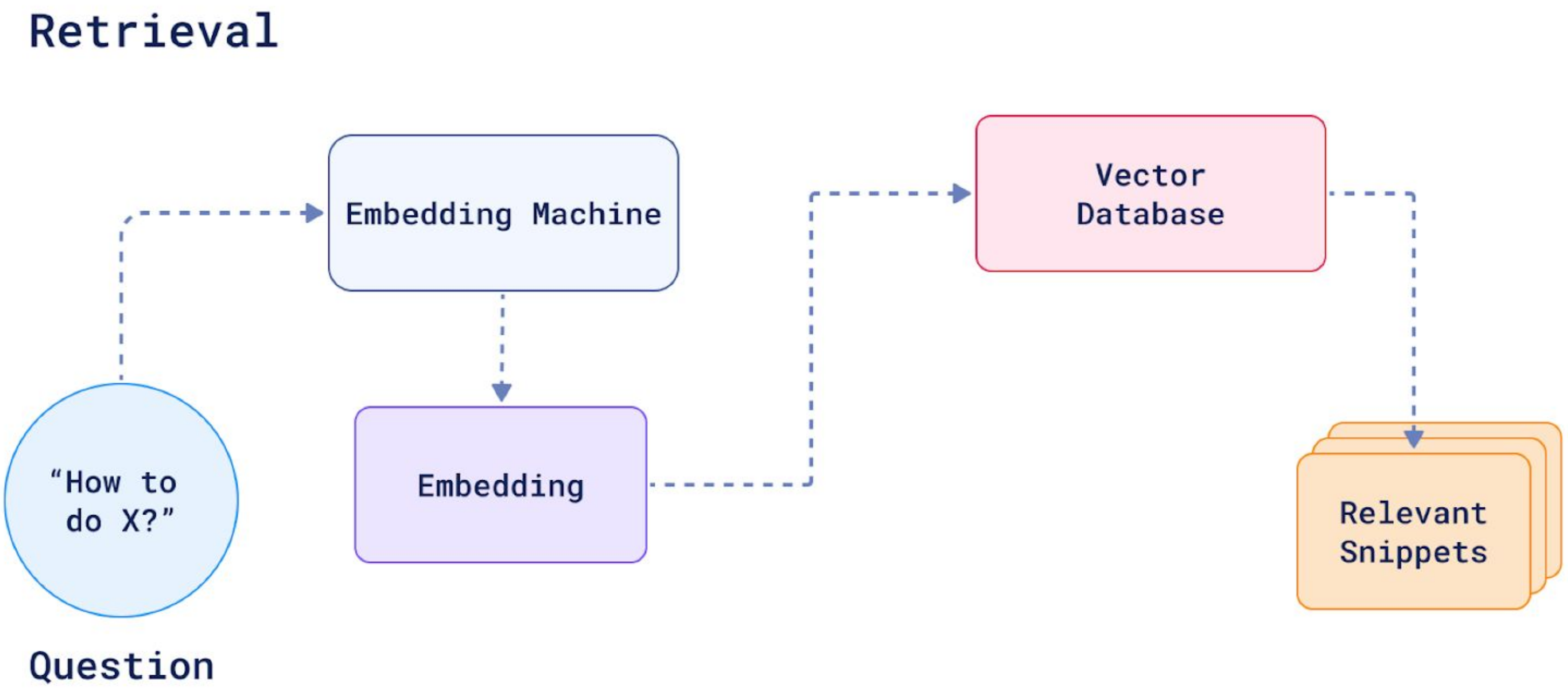
Similar objects are nearby while different objects are distant from each other in the vector space.



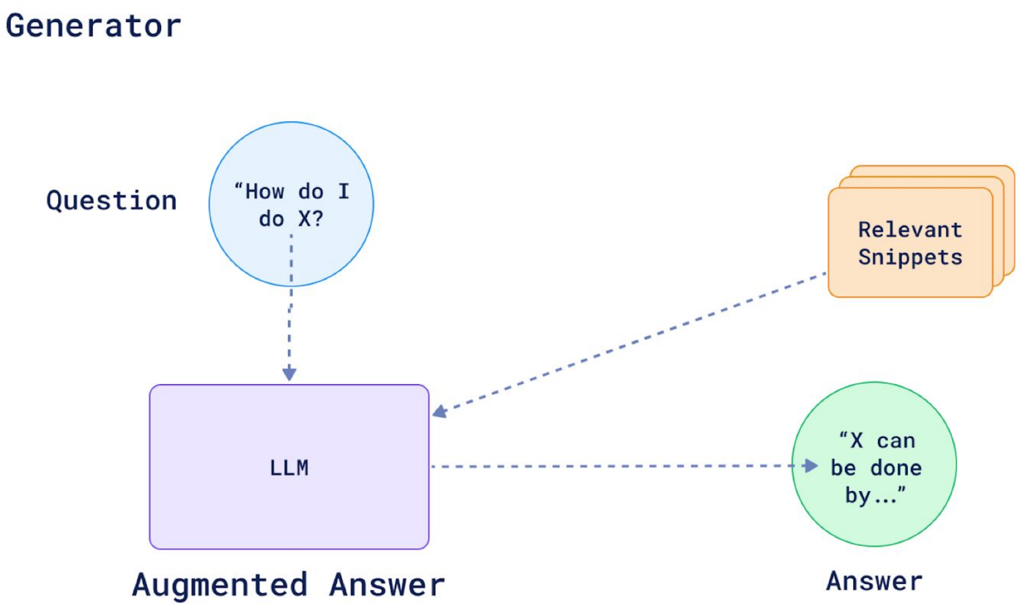
RAG in a nutshell: Indexing



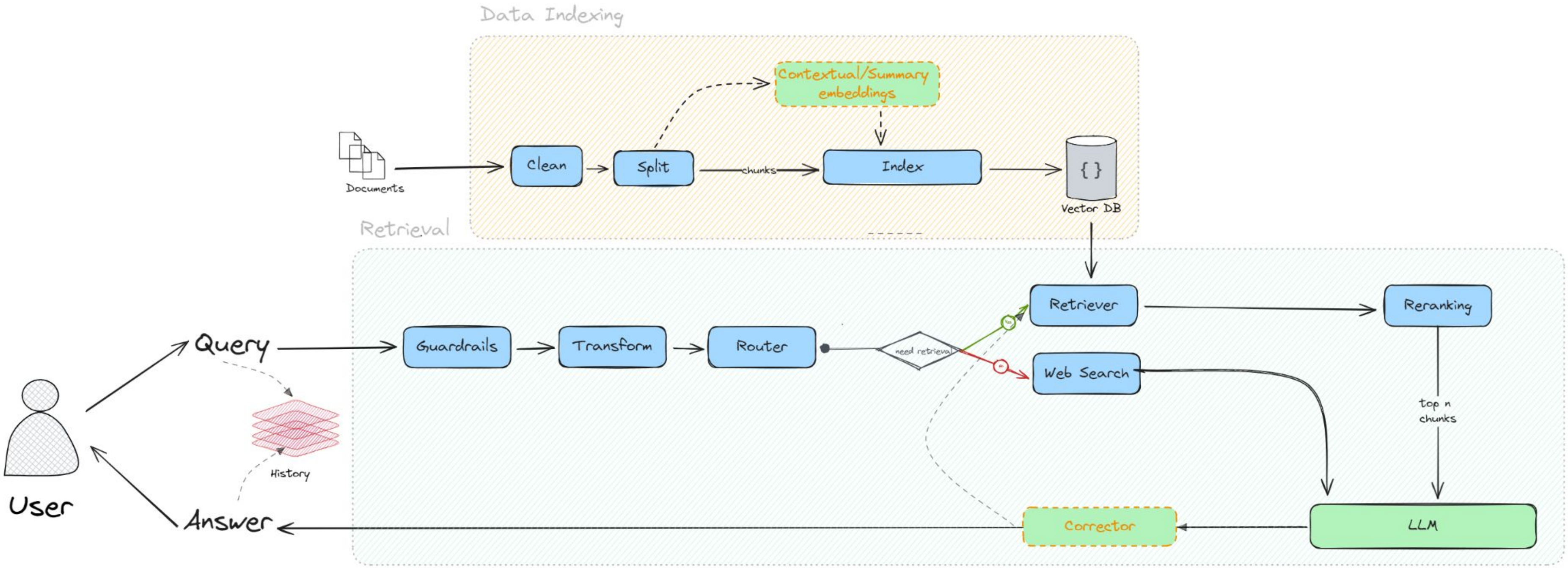
RAG in a nutshell: Retrieval



RAG in a nutshell: Generation



RAG architecture (in real-world applications)

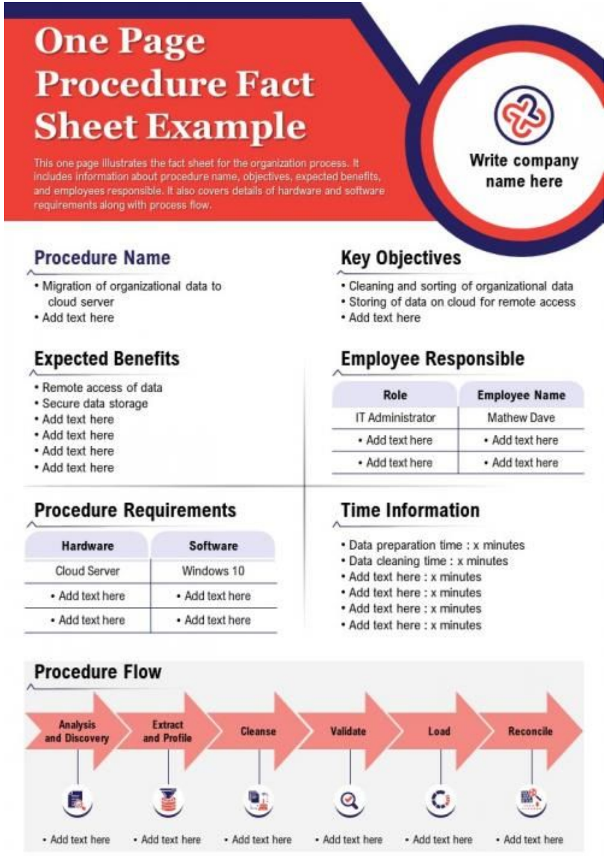


Summary (Part I)

- **RAG**: A system that combines retrieval of relevant information from external sources with generation using an LLM to create contextually enriched outputs.

- **Workflow:**
    - Index external documents (textual content) in the vector store
    - Embed the user’s query
    - Search for the most similar documents
    - Concatenate the relevant documents to the query and generate an answer

How to handle documents with more than text?



One Page Procedure Fact Sheet Example

This one page illustrates the fact sheet for the organization process. It includes information about procedure name, objectives, expected benefits, and employees responsible. It also covers details of hardware and software requirements along with process flow.

Write company name here

Procedure Name

- Migration of organizational data to cloud server
- Add text here

Key Objectives

- Cleaning and sorting of organizational data
- Storing of data on cloud for remote access
- Add text here

Expected Benefits

- Remote access of data
- Secure data storage
- Add text here
- Add text here
- Add text here
- Add text here

Employee Responsible

Role	Employee Name
IT Administrator	Mathew Dave
• Add text here	• Add text here
• Add text here	• Add text here

Procedure Requirements

Hardware	Software
Cloud Server	Windows 10
• Add text here	• Add text here
• Add text here	• Add text here

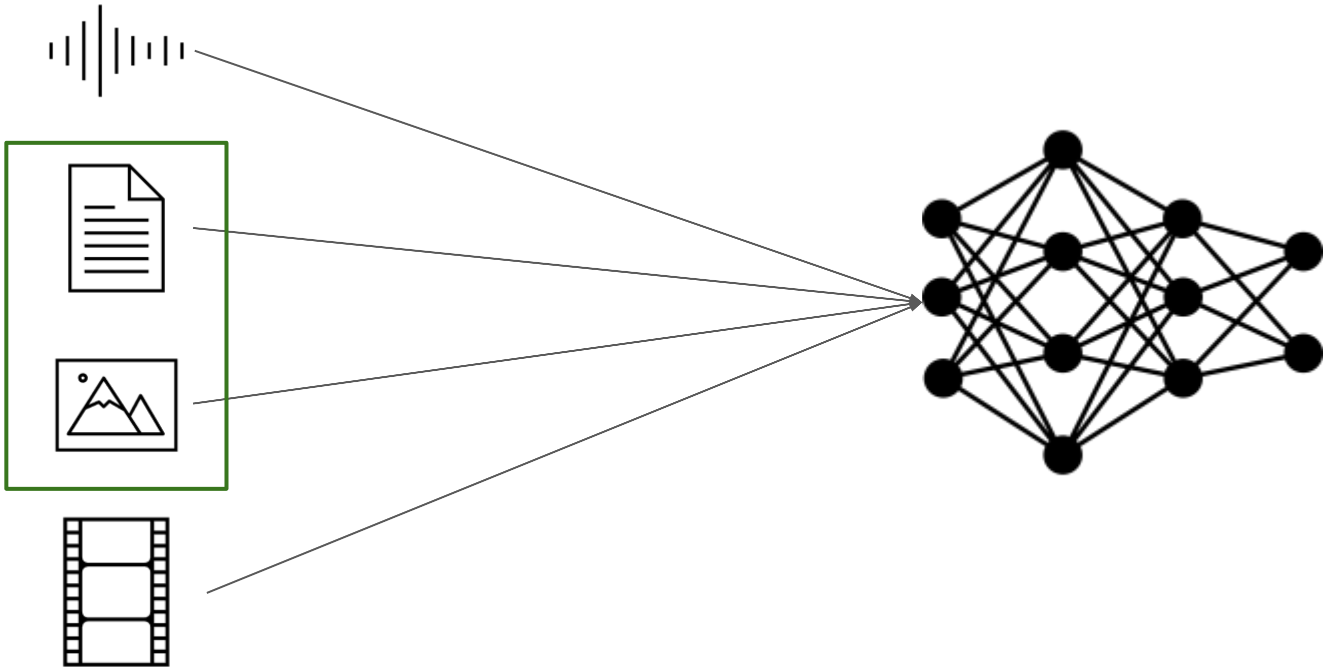
Time Information

- Data preparation time : x minutes
- Data cleaning time : x minutes
- Add text here : x minutes
- Add text here : x minutes
- Add text here : x minutes
- Add text here : x minutes

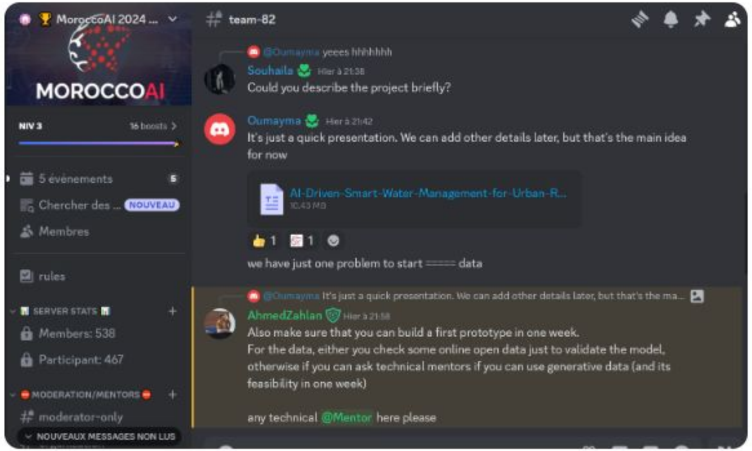
Procedure Flow

Analysis and Discovery	Extract and Profile	Cleanse	Validate	Load	Reconcile
• Add text here	• Add text here	• Add text here	• Add text here	• Add text here	• Add text here

What is multimodality 🤔?



What is multimodality 🤔?



@Oumayma yeees hhhhhhh

Souhaila Hier à 21:38
Could you describe the project briefly?

Oumayma Hier à 21:42
It's just a quick presentation. We can add other details later, but that's the main idea for now

AI-Driven-Smart-Water-Management-for-Urban-R...
10.43 MB

👍 1 1

we have just one problem to start ===== data

@Oumayma It's just a quick presentation. We can add other details later, but that's the ma...

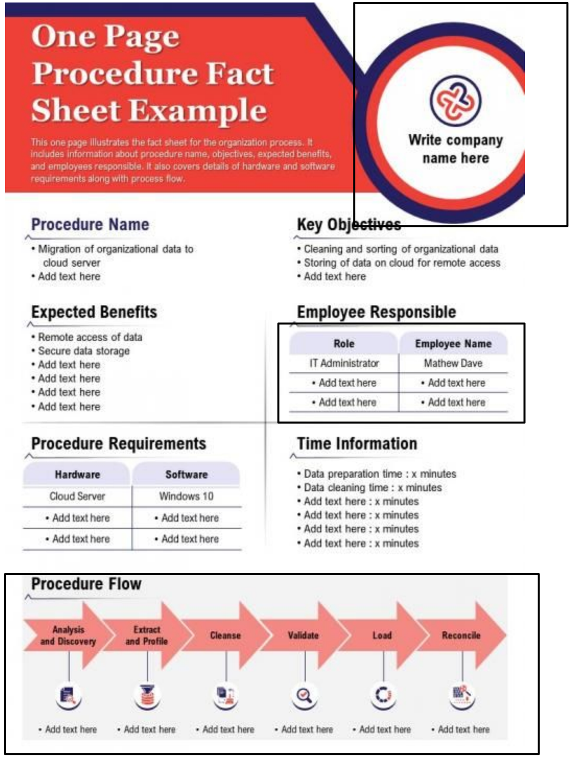
AhmedZahlan Hier à 21:58
Also make sure that you can build a first prototype in one week.
For the data, either you check some online open data just to validate the model, otherwise if you can ask technical mentors if you can use generative data (and its feasibility in one week)

any technical @Mentor here please

What do you think this discord server is about:

Based on the screenshot, the Discord server appears to be focused on **MoroccoAI 2024**, likely a competition, hackathon, or community initiative related to artificial intelligence (AI). The context of the messages suggests a project-based collaboration, where team members are working on AI-

How to build a RAG that handles images and text?



One Page Procedure Fact Sheet Example

This one page illustrates the fact sheet for the organization process. It includes information about procedure name, objectives, expected benefits, and employees responsible. It also covers details of hardware and software requirements along with process flow.

Write company name here

Procedure Name

- Migration of organizational data to cloud server
- Add text here

Key Objectives

- Cleaning and sorting of organizational data
- Storing of data on cloud for remote access
- Add text here

Expected Benefits

- Remote access of data
- Secure data storage
- Add text here
- Add text here
- Add text here
- Add text here

Employee Responsible

Role	Employee Name
IT Administrator	Mathew Dave
• Add text here	• Add text here
• Add text here	• Add text here

Procedure Requirements

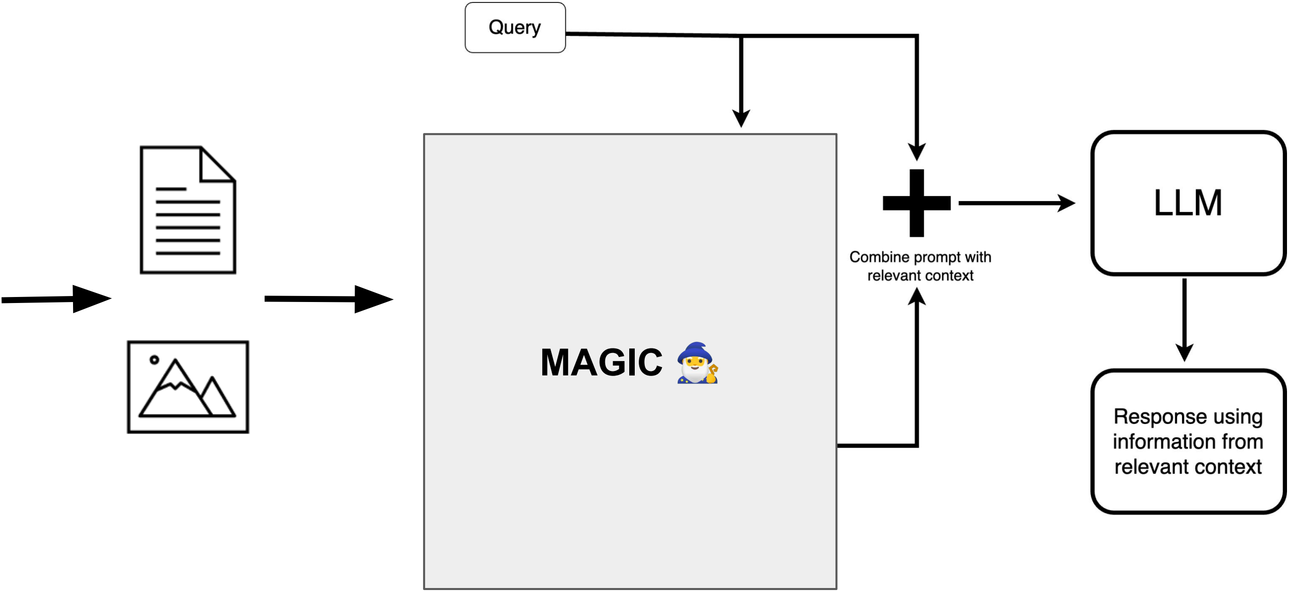
Hardware	Software
Cloud Server	Windows 10
• Add text here	• Add text here
• Add text here	• Add text here

Time Information

- Data preparation time : x minutes
- Data cleaning time : x minutes
- Add text here : x minutes
- Add text here : x minutes
- Add text here : x minutes
- Add text here : x minutes

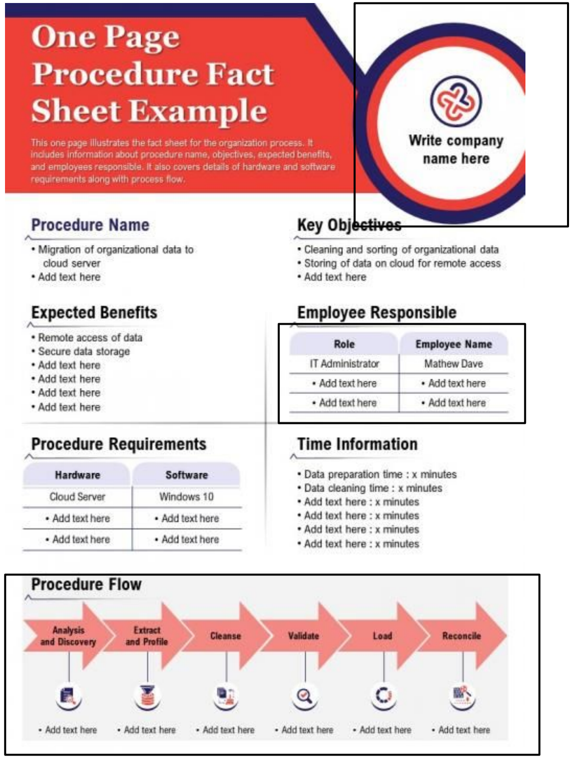
Procedure Flow

Analysis and Discovery	Extract and Profile	Cleanse	Validate	Load	Reconcile
• Add text here	• Add text here	• Add text here	• Add text here	• Add text here	• Add text here



Query
MAGIC 🧙
Combine prompt with relevant context
LLM
Response using information from relevant context

1st approach: Ground all modalities into one primary modality



One Page Procedure Fact Sheet Example

This one page illustrates the fact sheet for the organization process. It includes information about procedure name, objectives, expected benefits, and employees responsible. It also covers details of hardware and software requirements along with process flow.

Write company name here

Procedure Name

- Migration of organizational data to cloud server
- Add text here

Key Objectives

- Cleaning and sorting of organizational data
- Storing of data on cloud for remote access
- Add text here

Expected Benefits

- Remote access of data
- Secure data storage
- Add text here
- Add text here
- Add text here
- Add text here

Employee Responsible

Role	Employee Name
IT Administrator	Mathew Dave
• Add text here	• Add text here
• Add text here	• Add text here

Procedure Requirements

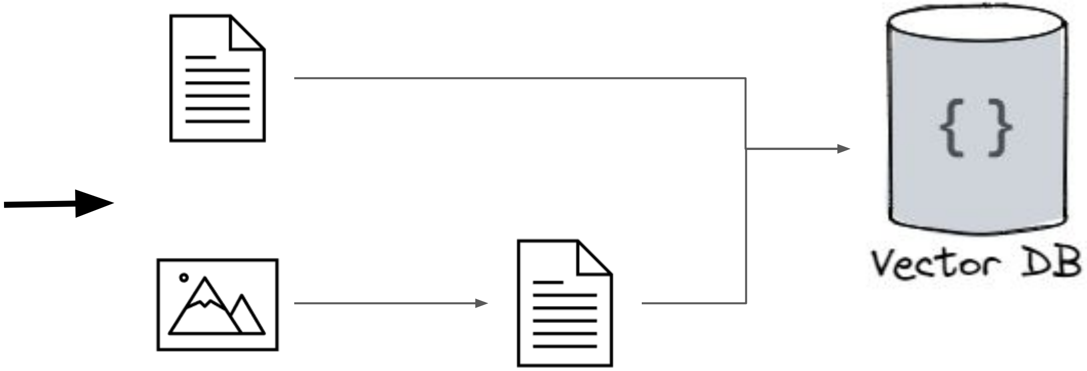
Hardware	Software
Cloud Server	Windows 10
• Add text here	• Add text here
• Add text here	• Add text here

Time Information

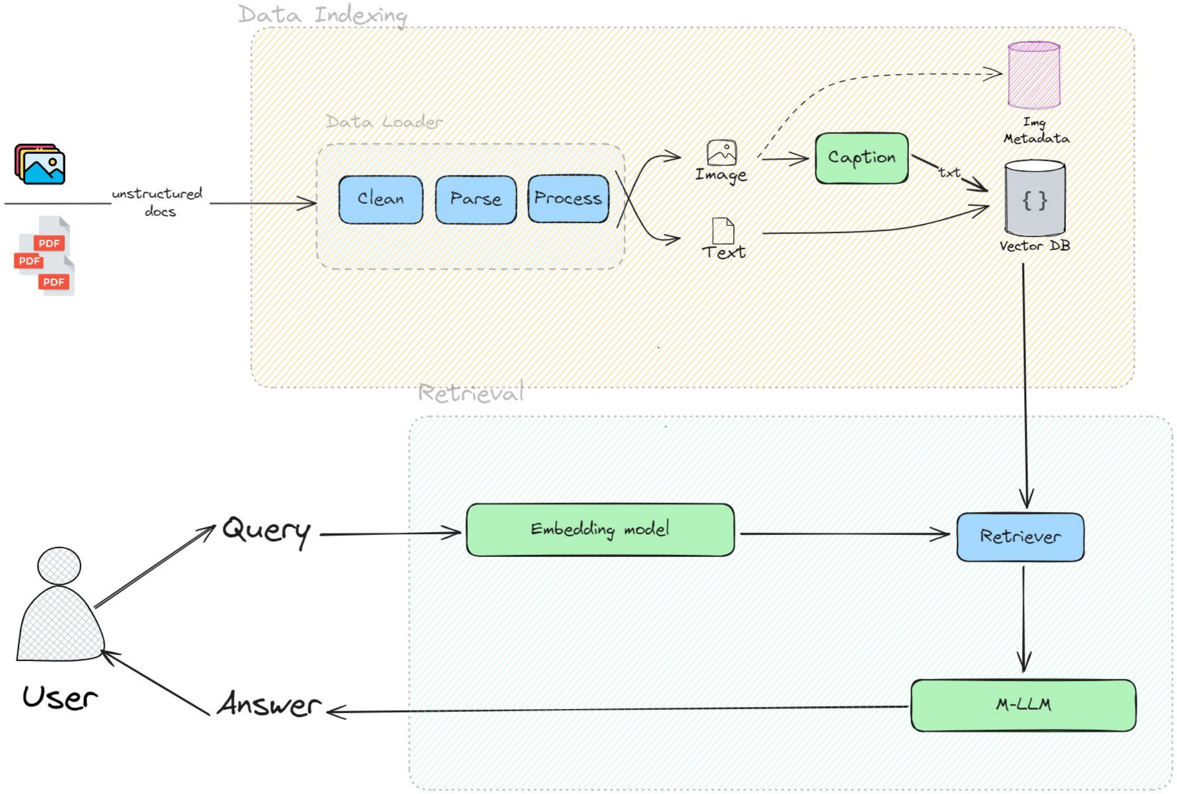
- Data preparation time : x minutes
- Data cleaning time : x minutes
- Add text here : x minutes
- Add text here : x minutes
- Add text here : x minutes
- Add text here : x minutes

Procedure Flow

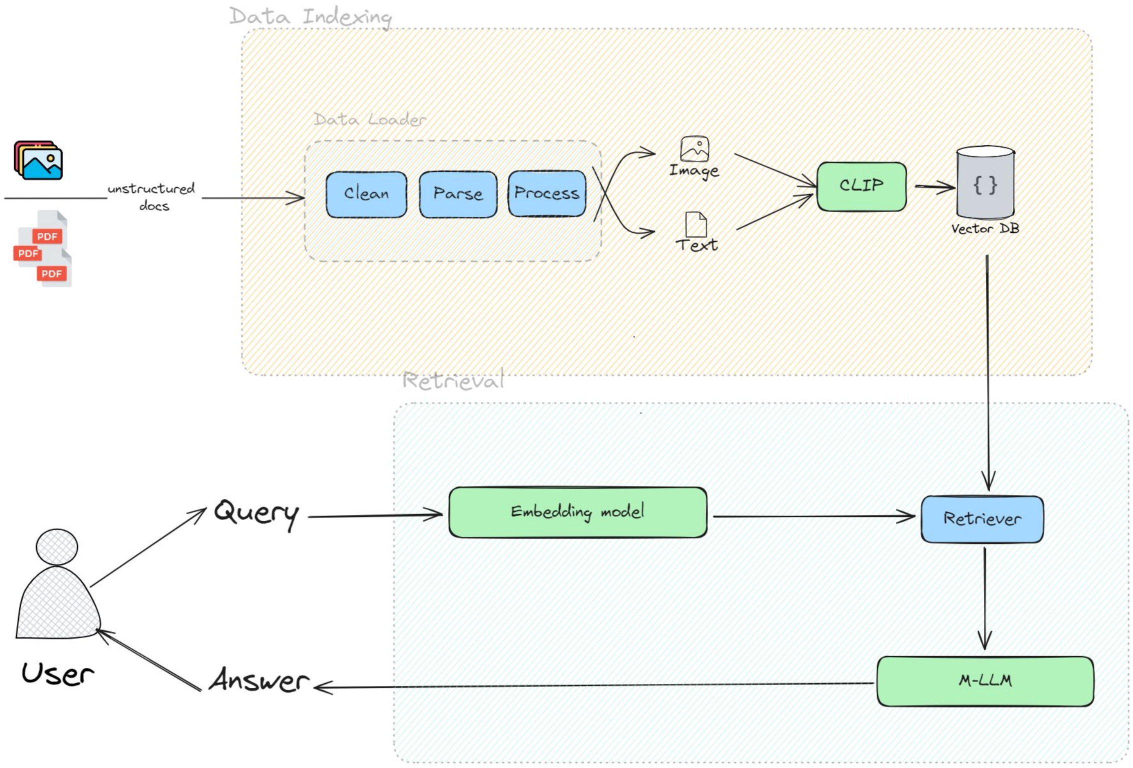
Analysis and Discovery	Extract and Profile	Cleanse	Validate	Load	Reconcile
• Add text here	• Add text here	• Add text here	• Add text here	• Add text here	• Add text here



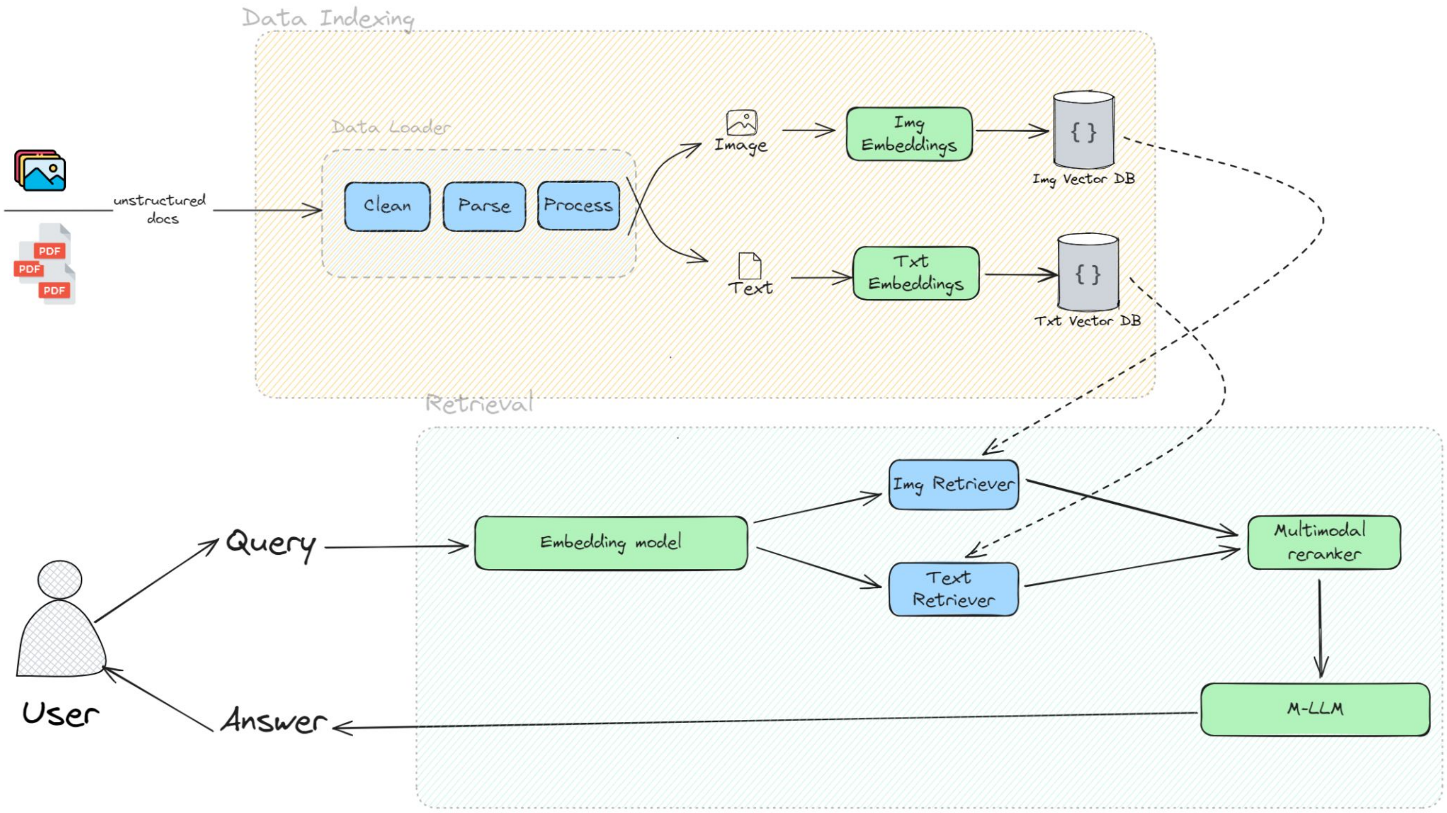
1st approach: Ground all modalities into one primary modality



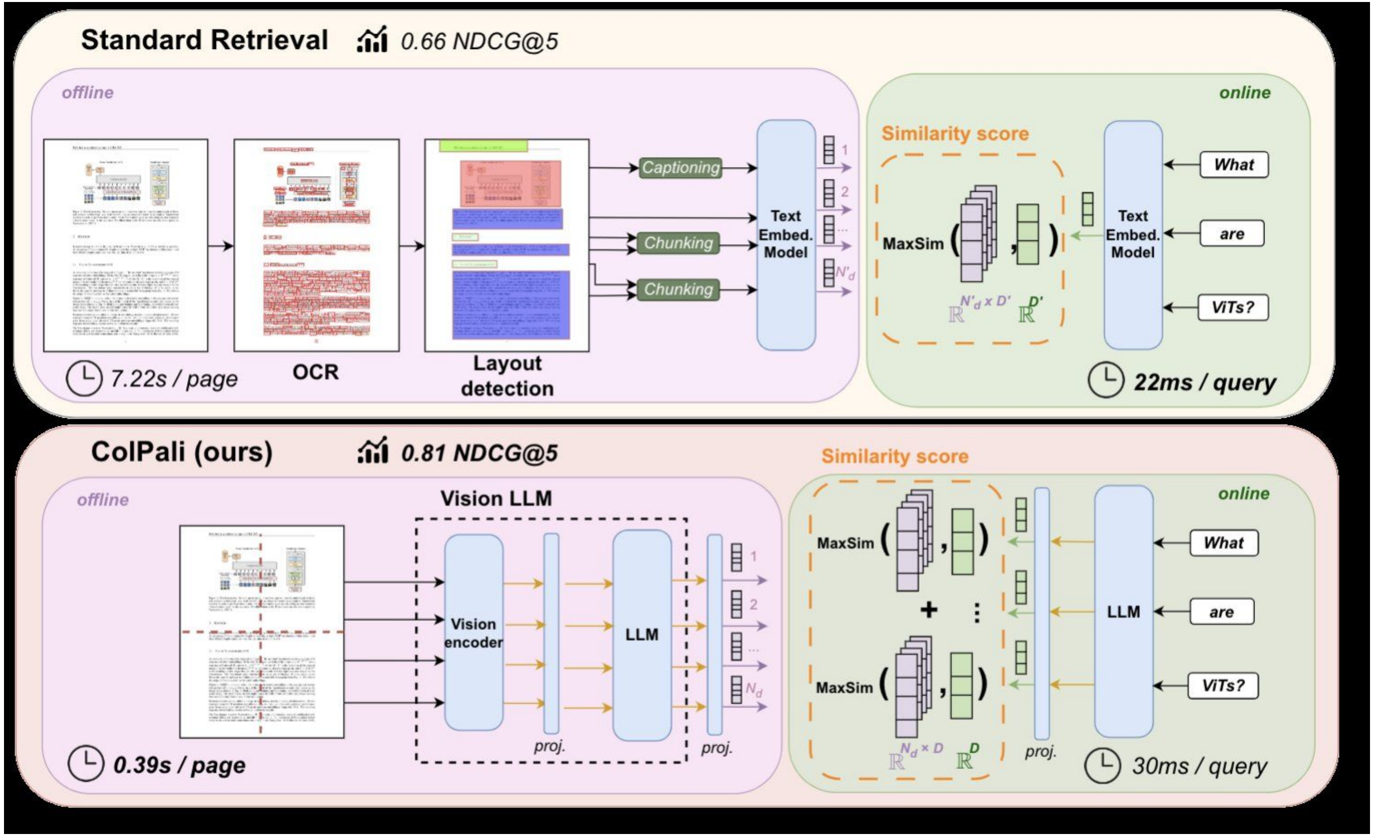
2nd approach: Embed all modalities into one vector space



3rd approach: Have separate stores per modality



4th approach: ColPali

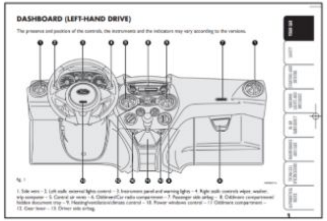


Summary

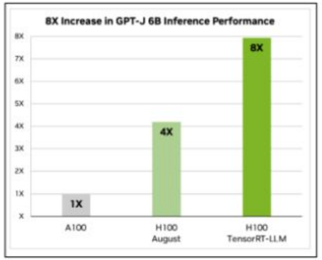
- No one-size fits all method:
  - The choice depends on the use case, type of documents, system constraints etc
  - For eg. CLIP models typically offer only generic insights into objects and shapes without providing detailed explanations and do not allow more than 80 words.

- Challenges:
  - Document parsing
  - Factuality and hallucinations
  - Splitting
  - Latency
  - …

- Hard to capture all information in text
- Very few "key points" of attention. More Focus on "General Imagery"


DASHBOARD (LEFT-HAND DRIVE)
The presence and position of the controls, the instruments and the indicators may vary according to the versions.

- Some details can be captured in Text
- Has both "key points" of attention and "General Imagery"



- Details can be perfectly captured in text.
- "Has key points of attention"