

# Expanding RAG with Multimodal capabilities

Praveen Mohan Prasad

Analytics Specialist  
AWS

Hajer Bouafif

Analytics Solutions Architect  
AWS

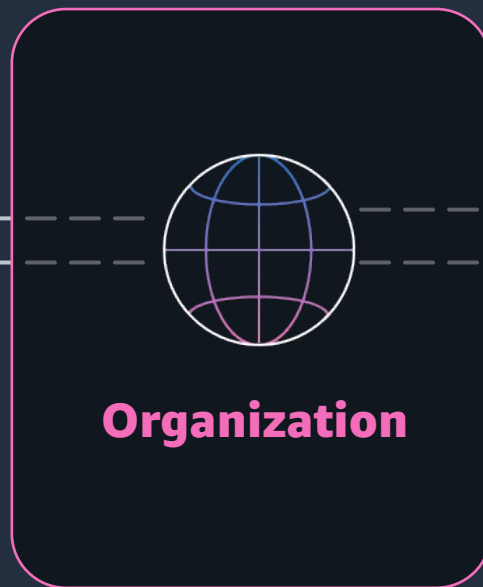
# Documents are ubiquitous



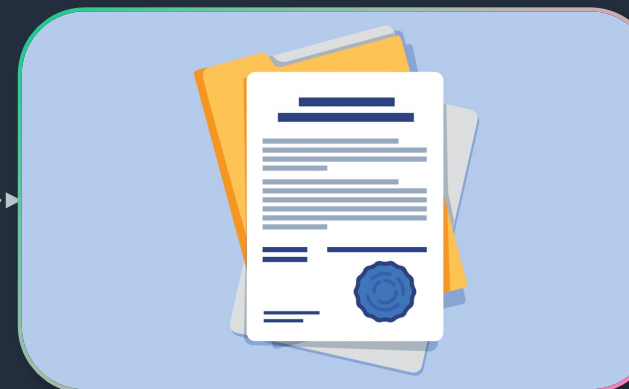
**Business reports (Internal and external)**



**Internal communications (Emails, meeting transcripts)**



**Invoices and Financial contracts**



**Policy and administrative papers**

# Chatbots and question answering



Ask: ?

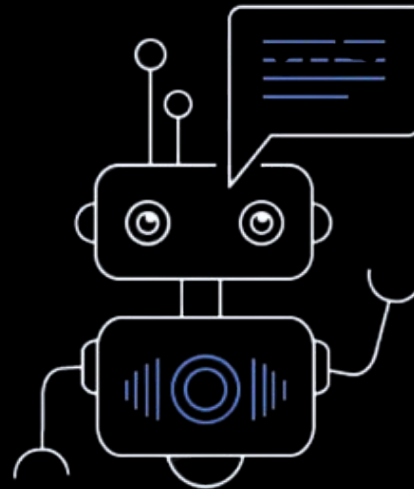
I

Press Enter to apply

GO

## Retrieval chatbots

- + Consistent
- + Less risky
- + Less cost
- Less intelligent
- More rules



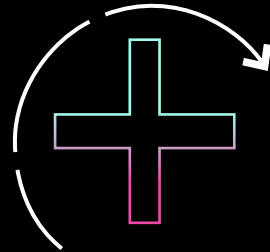
## Generative chatbots

- + Intelligent
- + Conversational
- More cost
- Unpredictable
- Re-training

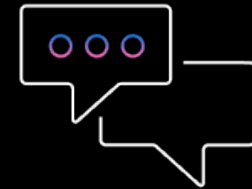
# RAG: Combining the two worlds



**Retrieval**



**Augmentation**



**Generation**

# Is RAG only text ?

Ask: ?

What is the trend in house prices in London?

13x

Current house price to earnings ratio in London

### A turn of fortunes for the London housing market?

London and the East of England have led the rebound in new buyer demand in the first weeks of 2024. Most other areas recorded below-average increases in demand, typically rising in line with last year or only ahead by single digits.

The rebound in London is uniform across the market segments - inner-London, suburban outer-London and the core commuter areas around London. This could reflect a turn of fortunes for the London housing market. Over the last seven years, the city has lagged behind the rest of the UK in terms of sales volumes and house price inflation.

Our house price index shows that London house prices have risen just 13% since the start of 2016. Meanwhile, they are 34% higher across the UK and almost 50% higher in Wales. The average value of a flat in London is just 2% higher over the same period.

Text

Low house price inflation since 2016 and rising earnings means housing affordability in London, measured on a house price to earnings ratio basis, is at its lowest since 2014. However, London housing prices remain expensive by UK standards at 13x earnings.

Slowly improving housing affordability in London is positive news but home buyers still face a sizable affordability challenge with mortgage rates doubling since 2021. We expect market conditions in London to continue to improve over 2024, with earnings rising faster than house prices. This will continue to improve affordability and support levels of housing sales rather than boost house prices.

-1.1%

London annual house price inflation

### Don't get carried away - it remains a buyers' market

The rebound in activity in the first weeks of 2024 is positive news but it's important not to over-interpret what this might mean for 2024. Mortgage rates fell to 4.2% over Q1 2023, which supported sales volumes and led to firmer pricing and modest price falls over 2023. We expect lower mortgage rates to do the same in 2024 - supporting sales volumes rather than having any impact on prices.

We believe that house prices will be kept in check by several factors. Firstly, a greater supply of homes for sale will provide buyers with more choice, especially for larger family homes. Secondly, half of those with a mortgage are yet to refinance onto higher rates. This is important as many would-be buyers are upsizers who will need a larger mortgage to move to a bigger home. Higher repayments will ensure buyers remain price sensitive and focused on value for money.

Thirdly, our data shows we are still in a buyers' market. A small but not insignificant number of sellers continue to cut asking prices to make

Graph (Images)



### Zoopla House Price Index, city summary, December 2023

Source: Zoopla House Price Index. Sparklines show last 12 months trend in annual and monthly growth rates - red bars are a negative value - each series has its own axis settings providing a more granular view on price development.

	Average price	%YoY Dec-23	%YoY Dec-22	Monthly trend	Annual trend
United Kingdom	£264,400	-0.8%	6.1%		
20 City Composite	£304,400	-0.7%	5.4%		
Belfast	£170,200	4.0%	3.2%		
Glasgow	£146,300	1.5%	4.4%		
Edinburgh	£269,900	1.3%	3.8%		
Liverpool	£157,400	0.9%	7.6%		
Leeds	£209,000	0.6%	6.9%		
Newcastle	£153,000	0.5%	6.4%		
Manchester	£223,000	0.4%	7.8%		
Sheffield	£172,000	0.4%	7.5%		
Cardiff	£252,600	0.1%	6.8%		
Birmingham	£208,100	-0.1%	7.9%		

Tables

Aberdeen	£139,900	-2.9%	-0.7%		
----------	----------	-------	-------	--	--

Source: Zoopla House Price Index. Sparklines show last 12 months trend in annual and monthly growth rates - red bars are a negative value - each series has its own axis settings providing a more granular view on price development.

### Contacts

If you have any questions about our research please do get in touch

Richard Donnell  
Director of Research & Insight

Theo Brewer  
Director of Innovation

# Multimodal RAG: Approaches

- Rely on **text as primary** modality
- **Shared embedding** space across modalities
- Use latest **(M)LLMs**
- **Ensemble** approach



# Multimodal RAG

RAG that understands  
**Images and Tables** too!



© 2024, Amazon Web Services, Inc. or its affiliates.

## Chat with your data

CLEAR

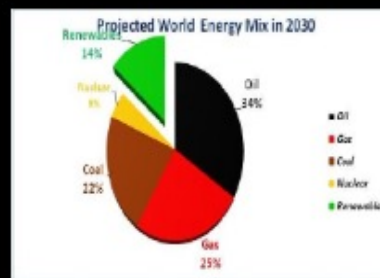
what is the projected energy from renewable sources by 2030 ?

According to the pie chart shown in the image, the projected contribution of renewable energy sources to the world energy mix in 2030 is 14%.

0:00 / 0:08

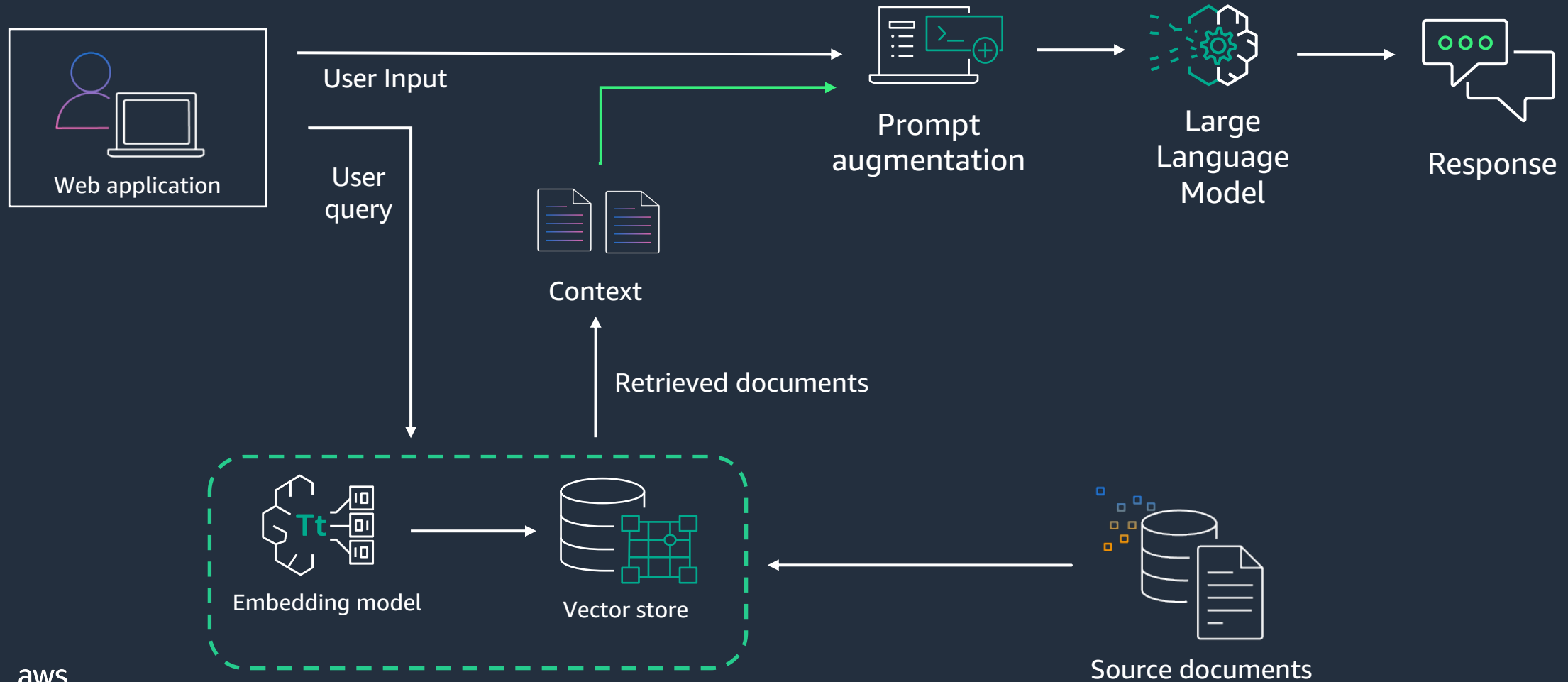
Relevant Sources:

Images:



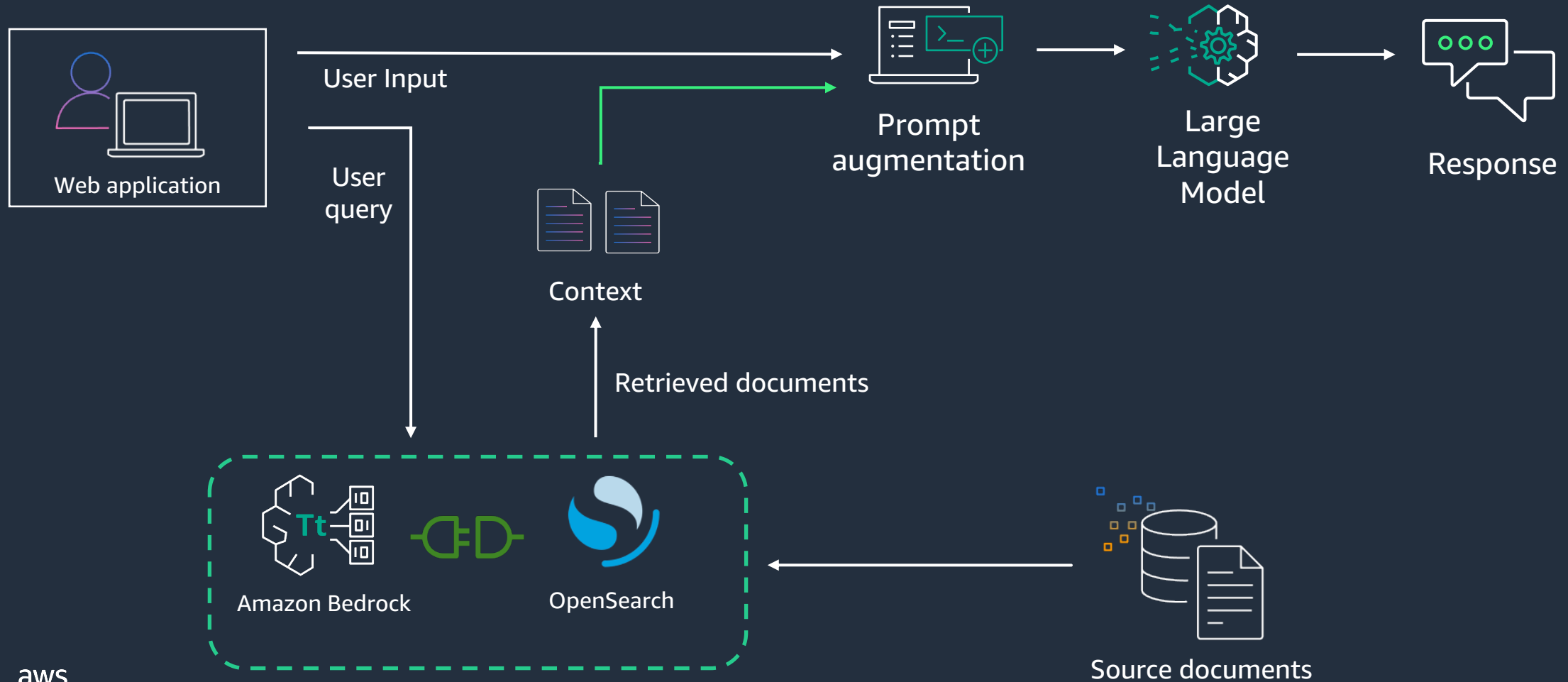
Raw sources:

# RAG application: High level architecture



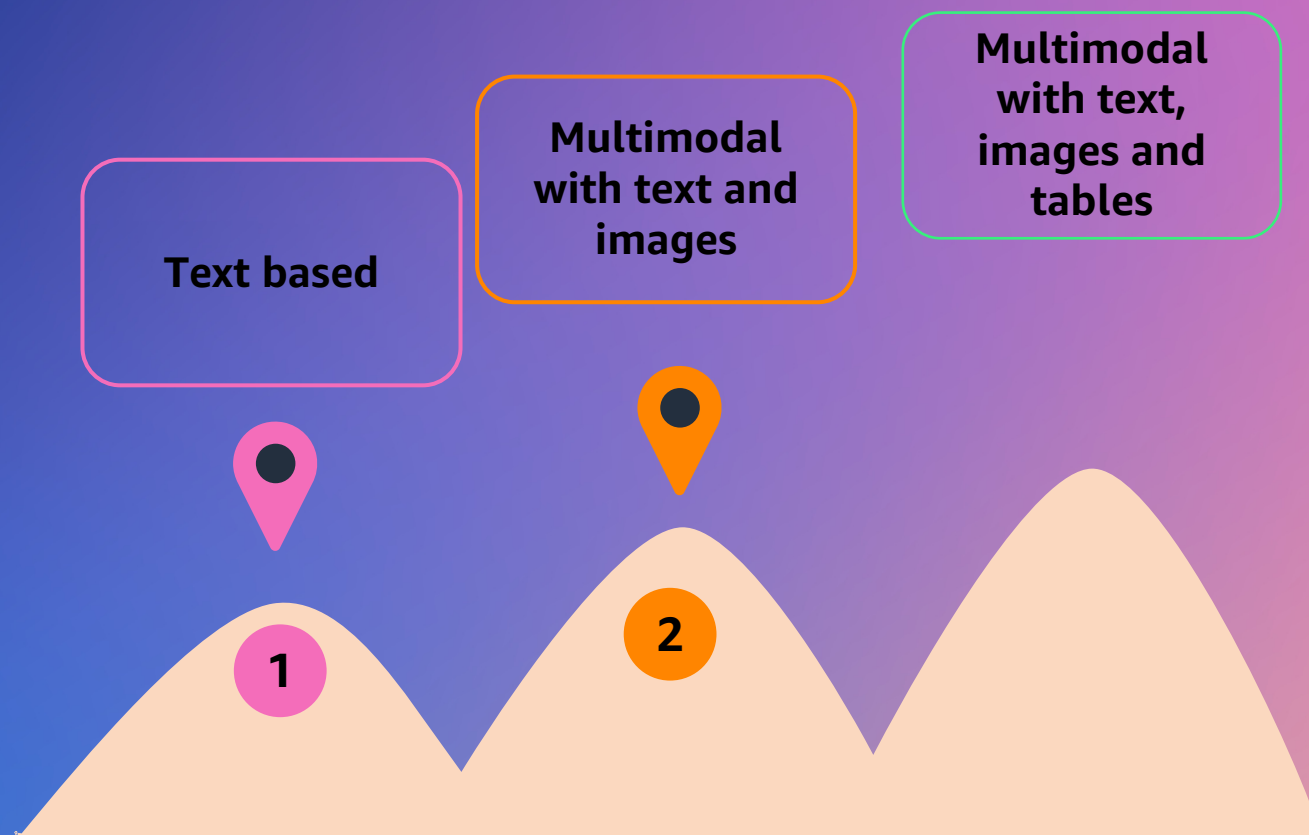


# RAG application: High level architecture



# Demo

RAG with text and images



# Handle Images in RAG

**Grounding images to captions**

# Image captioning : Graphical images

*Question: what is the average of percentage homes selling in London between December 2020 and December 2022?*



PDF



Graph

Extract images from PDF

# Image captioning : Graphical images

*Question: what is the average of percentage homes selling in London between december 2020 and december 2022?*



Input image: Graph



vit-gpt2-image-captioning

A drawing of a person with a computer

Graph generated captions – V0



Salesforce/blip-image-captioning-base

a graph showing the price of a home in london

Graph generated captions – V1



Does not provide analytics information

# Observations

- Classify the images (graphical, non graphical)
- Use specialized ML models (ex. *deplot*) to generate captions for graphical images

<https://huggingface.co/google/deplot>

# Multimodal LLM (MLLMS) to the rescue

**Question: what is the average of percentage homes selling in London between December 2020 and December 2022?**



""This image is a line graph that shows the percentage of homes selling for more than 10% below the asking price over time in two different regions - London & South East, and the Rest of the UK.

The **vertical axis** represents the percentage of homes selling below the asking price by more than 10%, **ranging from 0% to 25%**. The **horizontal axis** shows the time period from **December 2017 to December 2023 (projected)**.

There are **two lines plotted**, one for **London & South East (in black)** and another for the **Rest of the UK (in purple)**. The black line shows a peak around **December 2018**, indicating a higher percentage of homes in London & South East sold below the asking price during that time. The purple line for the Rest of the UK has an upward trend, particularly from **December 2021 onwards**, suggesting an increasing number of homes selling below the asking price in those areas in recent years.

The data is sourced from Zoopla Research analysis of Hometrack data, as mentioned at the bottom of the image."

Input image: Graph

LLM prompt: "What is in the image?"

Image captioning using Claude 3 (Sonnet)



# Retrieval types

RAG with text and images

# Text based retrieval



""This image is a line graph that shows the percentage of homes selling for more than 10% below the asking price over time in two different regions - London & South East, and the Rest of the UK.

....  
The data is sourced from Zoopla Research analysis of Hometrack data, as mentioned at the bottom of the image."

```
"properties": {  
  "processed_element": <image caption>  
  "processed_element_embedding":  
    [0.1,0.4,.....],  
  "element_type": "image"  
  "src_url": "/images/./image1.jpg"  
}
```

Input image: Graph

Generate image captions

Store **embeddings** for captions



Amazon Bedrock  
Titan Embeddings G1



OpenSearch

# Multimodal retrieval

% homes selling for more than 10% below the asking price



Source: Zoopla Research analysis of Hometrack data

""This image is a line graph that shows the percentage of homes selling for more than 10% below the asking price over time in two different regions - London & South East, and the Rest of the UK.

....  
The data is sourced from Zoopla Research analysis of Hometrack data, as mentioned at the bottom of the image."

```
"properties": {  
  "processed_element": <image caption>  
  "processed_element_embedding_bedrock-multimodal": [0.1,0.4,.....]  
  "raw_element_type": "image",  
  "raw_element": "image_binary"  
  "src_doc": "/images/./image1.jpg"  
}
```

Input image: Graph

Generate image captions

Store embeddings for **both image and captions**



Amazon Bedrock  
(Amazon Titan Multimodal  
Embeddings model)



OpenSearch

# Handle Tables in RAG

## **Approach 1:** Grounding tables to text

# STEP 1: Extract tables from pdf

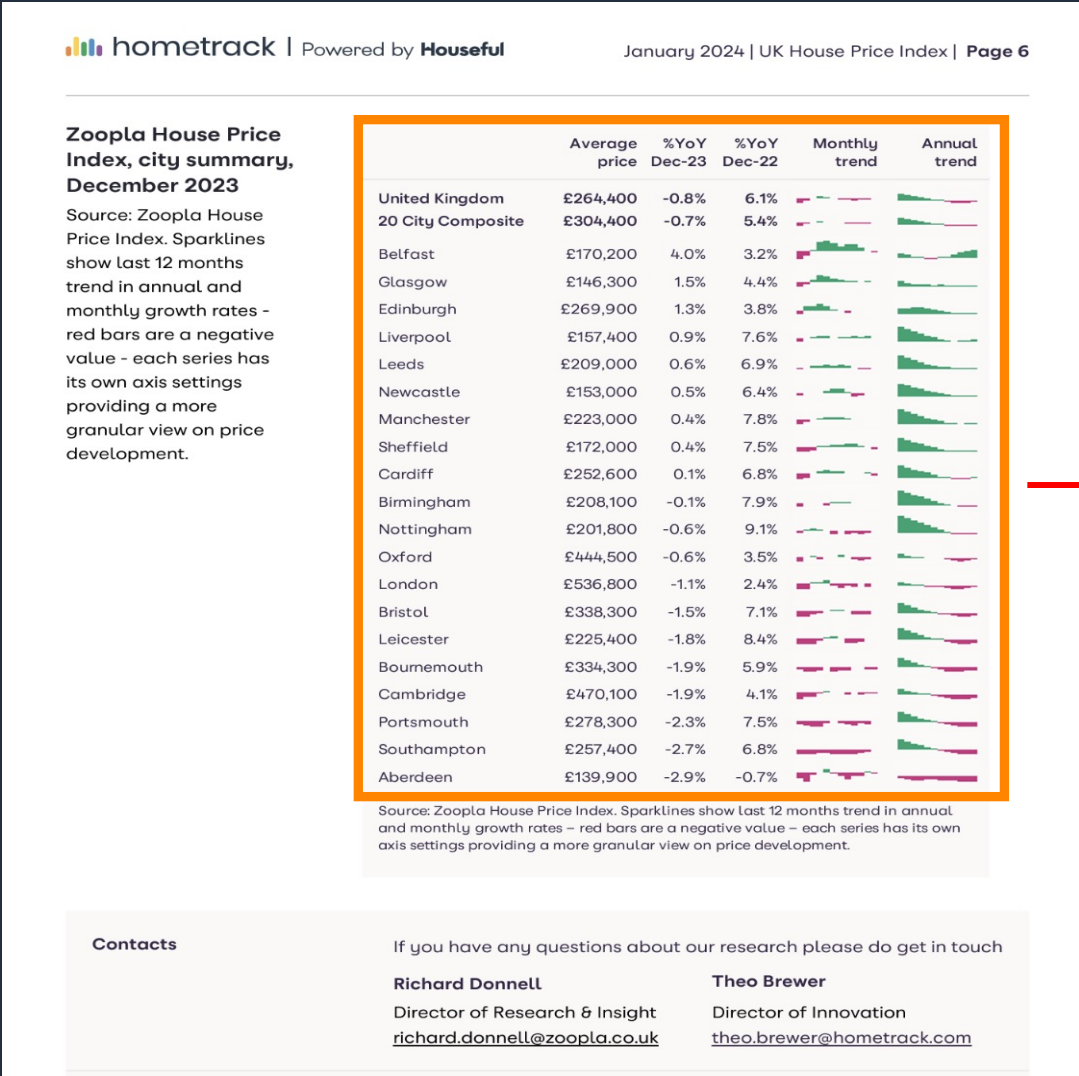


Table1.csv

	Average price	%YoY Dec-23	%YoY Dec-22
20 City Composite	£304,400	-0.70%	5.40%
Belfast	£170,200	4.00%	3.20%
Glasgow	£146,300	1.50%	4.40%
Edinburgh	£269,900	1.30%	3.80%
Liverpool	£157,400	0.90%	7.60%
Leeds	£209,000	0.60%	6.90%
Newcastle	£153,000	0.50%	6.40%
Manchester	£223,000	0.40%	7.80%
Sheffield	£172,000	0.40%	7.50%
Cardiff	£252,600	0.10%	6.80%

## Tools

Python: Camelot, tabula-py

Cloud: Amazon Textract, Google Document AI, Azure Form Recognizer



# STEP 2: Summarise tables to text

"""You are an assistant tasked with summarizing tables and text. \nGive a detailed summary of the table or text. Table or text chunk: {element} """



## CSV string

```
`Average price` `%YoY Dec-23` `%YoY Dec-22` Monthly trend` Annual trend`
United Kingdom `£264,400` -0.8% `6.1%` ``
20 City Composite `£304,400` -0.7% `5.4%` ``
Belfast `£170,200` 4.0% `3.2%` ``
Glasgow `£146,300` 1.5% `4.4%` ``
Edinburgh `£269,900` 1.3% `3.8%` ``
Liverpool `£157,400` 0.9% `7.6%` ``
Leeds `£209,000` 0.6% `6.9%` ``
Newcastle `£153,000` 0.5% `6.4%` ``
Manchester `£223,000` 0.4% `7.8%` ``
Sheffield `£172,000` 0.4% `7.5%` ``
Cardiff `£252,600` 0.1% `6.8%` ``
```

## Summary

""The table presents data on average house prices, year-over-year (YoY) percentage changes, and monthly and annual trends for various cities and regions in the United Kingdom as of December 2023.

1. City-wise trends:

- Belfast, Glasgow, Edinburgh, Liverpool, Leeds, Newcastle, Manchester, Sheffield, and Cardiff saw positive year-over-year price growth, ranging from 0.1% (Cardiff) to 4.0% (Belfast).

4. Monthly trends:

- Most cities showed positive monthly price changes compared to December 2022

5. Annual trends:

- All cities, except Aberdeen, exhibited positive annual price growth compared to the previous year

## Store in OpenSearch k-NN Index

```
"properties": {
  "processed_element": <summary>
  "processed_element_embedding":
    [0.1,0.4,.....]
  "raw_element": <csv_string>
  "raw_element_type": "table"
  "src_path": "/tables/table1.csv"
}
```



# Limitations

- Loss of information and structure
- Lack of deep analytical query capabilities
- Problem with complex tables
- Complex prompting



# Approach 2: Handling tables as Data frames

*Question: Which city in UK has the highest housing price ?*

- Convert Question to SQL using LLM
- Hit SQL on dataframe
- Need for **Chain of Thoughts** in your LLM (Langchain, Llama Index)

**Thought:** To find the city with the highest average housing price, I need to look at the **Average price** column in the dataframe and find the maximum value.

**Action:** python\_repl\_ast

**Action Input:** `print(df["Average price"].max())`  
£536,800

**Thought:** The output shows that the maximum value in the "Average price" column is £536,800. However, this does not directly tell me which city corresponds to that price. To find that out, I need to filter or subset the dataframe to only include the row with that maximum price value.

**Action:** python\_repl\_ast

**Action Input:**

```
max_price_row = df.loc[df["Average price"] ==  
df["Average price"].max()]  
print(max_price_row)
```

14	London	£536,800
----	--------	----------

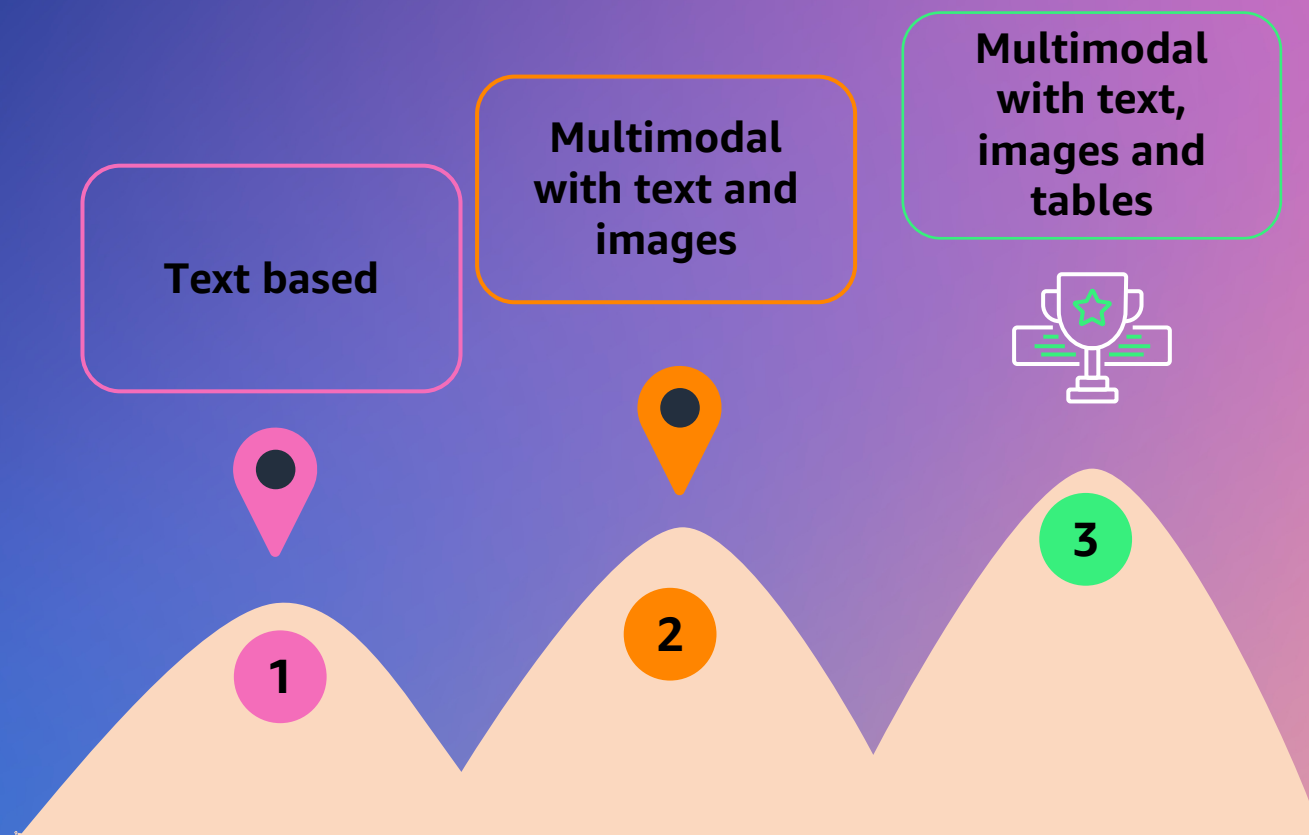
[1 rows x 7 columns]

**Thought:** The filtered dataframe shows that the row with the maximum "Average price" value of £536,800 corresponds to the city "London".

**Final Answer:** The city in the United Kingdom with the highest average housing price is London.  
Finished chain.

# Demo

RAG with text and images and tables



# Complex tables

Model (→)	Lexical	Sparse			Dense				Late-Interaction	Re-ranking
Dataset (↓)	BM25	DeepCT	SPARTA	docT5query	DPR	ANCE	TAS-B	GenQ	ColBERT	BM25+CE
MS MARCO	0.228	0.296 <sup>‡</sup>	0.351 <sup>‡</sup>	0.338 <sup>‡</sup>	0.177	0.388 <sup>‡</sup>	0.408 <sup>‡</sup>	0.408 <sup>‡</sup>	0.401 <sup>‡</sup>	0.413 <sup>‡</sup>
TREC-COVID	0.656	0.406	0.538	0.713	0.332	0.654	0.481	0.619	0.677	0.757
BioASQ	0.465	0.407	0.351	0.431	0.127	0.306	0.383	0.398	0.474	0.523
NFCorpus	0.325	0.283	0.301	0.328	0.189	0.237	0.319	0.319	0.305	0.350
NQ	0.329	0.188	0.398	0.399	0.474 <sup>‡</sup>	0.446	0.463	0.358	0.524	0.533
HotpotQA	0.603	0.503	0.492	0.580	0.391	0.456	0.584	0.534	0.593	0.707
FiQA-2018	0.236	0.191	0.198	0.291	0.112	0.295	0.300	0.308	0.317	0.347
Signal-1M (RT)	0.330	0.269	0.252	0.307	0.155	0.249	0.289	0.281	0.274	0.338
TREC-NEWS	0.398	0.220	0.258	0.420	0.161	0.382	0.377	0.396	0.393	0.431
Robust04	0.408	0.287	0.276	0.437	0.252	0.392	0.427	0.362	0.391	0.475
ArguAna	0.315	0.309	0.279	0.349	0.175	0.415	0.429	0.493	0.233	0.311
Touché-2020	0.367	0.156	0.175							
CQADupStack	0.299	0.268	0.257							
Quora	0.789	0.691	0.630							
DBPedia	0.313	0.177	0.314							
SCIDOCS	0.158	0.124	0.126							
FEVER	0.753	0.353	0.596							
Climate-FEVER	0.213	0.066	0.082							
SciFact	0.665	0.630	0.582							
Avg. Performance vs. BM25		- 27.9%	- 20.3%							

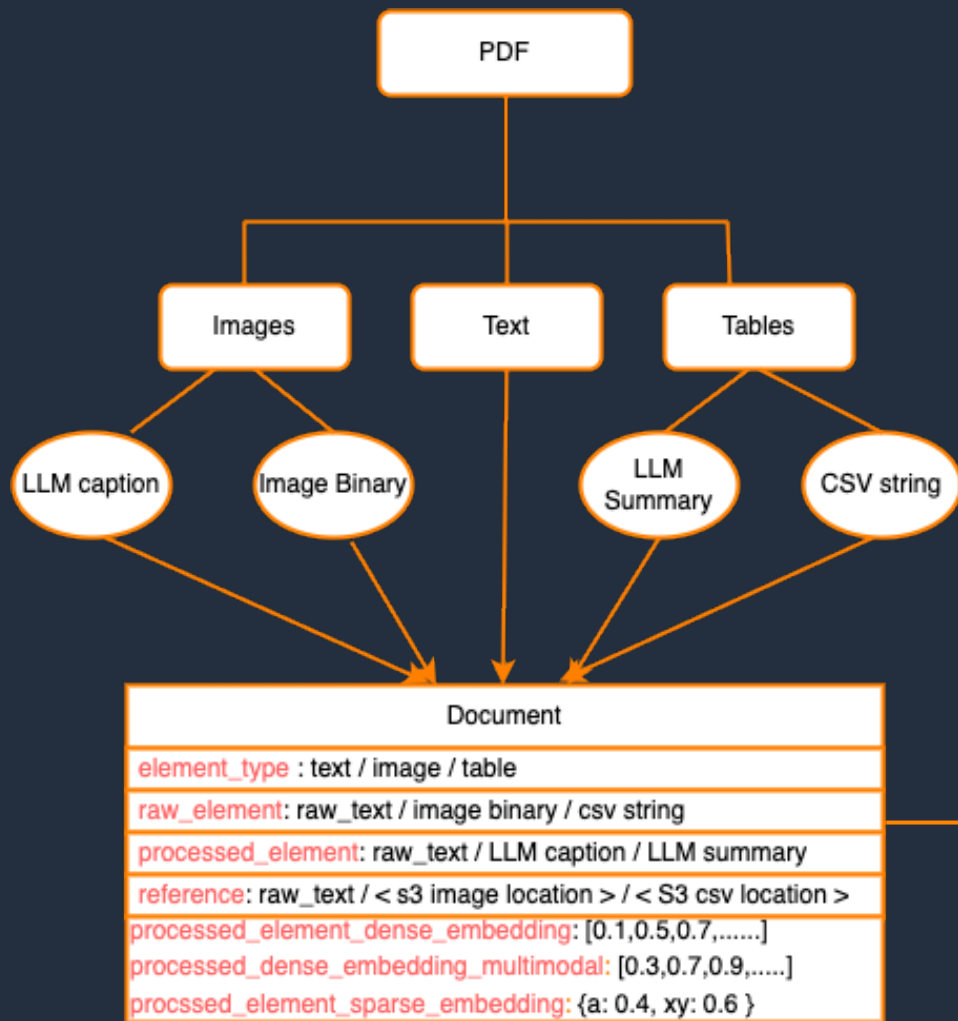
Member Name	Alias Default	Consolidation	Data Storage	Time Balance	Skip Option	Variance Reporting
▲ Accounts		^ (Never)	Auto	None	None	Non-Expense
▲ IncomeStatement	Income Statement	+ (Addition)	Auto	None	None	Non-Expense
▲ 340000	Net Income	+ (Addition)	Auto	None	None	Non-Expense
▲ 310000	Total Pretax Income	+ (Addition)	Auto	None	None	Non-Expense
▶ 311000	Pretax Income From Operations	+ (Addition)	Auto	None	None	Non-Expense
312000	Other Exp (Inc)	+ (Addition)	Auto	None	None	Expense
320000	Provision for Income Tax	- (Subtraction)	Auto	None	None	Expense
330000	Minority Interest Income	+ (Addition)	Auto	None	None	Non-Expense
▲ BalanceSheet	Balance Sheet	~ (Ignore)	Auto	Last	None	Non-Expense
▶ 100000	Total Assets	+ (Addition)	Auto	Last	None	Non-Expense
▶ 200000	Total Liabilities and Equity	- (Subtraction)	Auto	Last	None	Non-Expense
▲ Statistics		~ (Ignore)	Auto	None	None	Non-Expense
▶ LeadsDrivers		~ (Ignore)	Auto	None	None	Non-Expense
▶ Headcount		~ (Ignore)	Auto	Average	Missing	Expense

Date	Details	Debit €	Credit €	Balance €
27 Aug 2021	BALANCE FORWARD			469.98
	N WEB SERVICE21082		4418.37	
	IE21082713468835			4888.35
30 Aug 2021	PRAVIN BHAI		75.00	
	SKYDIVING		205.00	
	VDP-FREENOW	24.80		
	VDP-IRISH PARACHUT	820.00		
	*MOBI JAYASURIYA	700.00		
	*MOBI 49348024	17.00		
	SAVINGS MOVE	3500.00		
	VDC-GOVINDAS	1.95		
	VDC-ORIENTAL PANTR	75.00		29.60
1 Sep 2021	*MOBI ONLINE SAVER		500.00	
			5.00	
		312.50		222.10
		5.00		217.10
			300.00	517.10
			500.00	
			21.00	
			300.00	

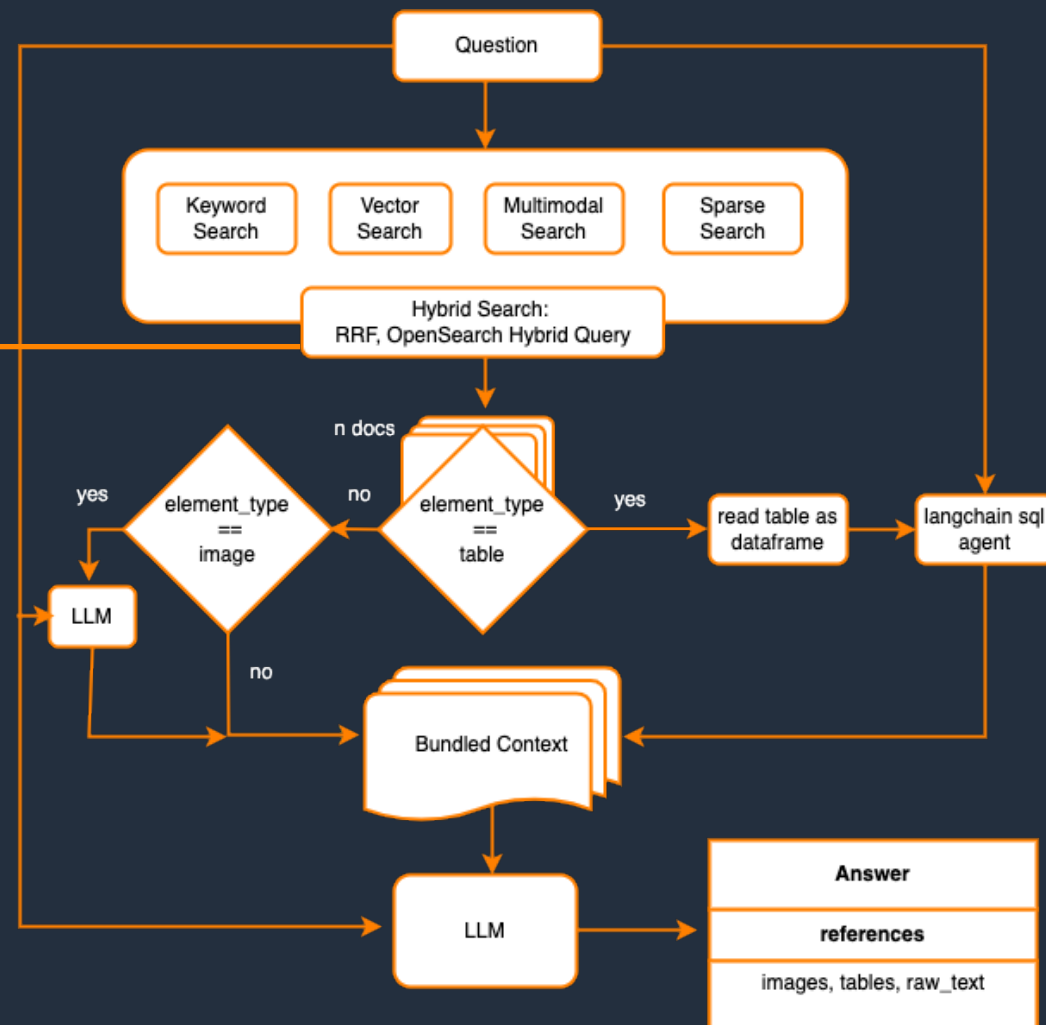
Domain understanding and additional custom processing !

# Multimodal RAG workflow

## Ingestion



## Query



# Key learnings



- Custom pre-processing
- **Chunking and indexing:** Nested vectors
- **Choose the right retriever** (A/B testing)
  - Lexical Search
  - Vector Search
  - Neural Sparse Search
  - Multimodal Search
  - Hybrid Search
- **Apply re-ranking:** Cross encoder models
- **Multi-stage Evaluation:** Retrieval and Generation

# Next steps

- Multimodal queries
- Generation of images to augment the answers
- Evaluation of retriever and LLM outputs: Self RAG and CRAG
- Agents framework: to automate the retriever selection
- Simplify using conversational search plugin of **OpenSearch 2.13**

# Build your own RAG system



AWS Samples – AI Search with  
OpenSearch Service



Conversational Search with  
OpenSearch



ML Search with OpenSearch





# Thank you!

Hajer Bouafif  
bouhajer@amazon.fr

Praveen Mohan  
prasadnu@amazon.com