

# Retrieval-augmented generation (RAG) for sustainability regulations

**Quint Kleinsman**

*Information Science Student, UvA*

*“LLMs are Fancy Autocomplete  
with a dash of Magic”*

- Tobias Zwingmann

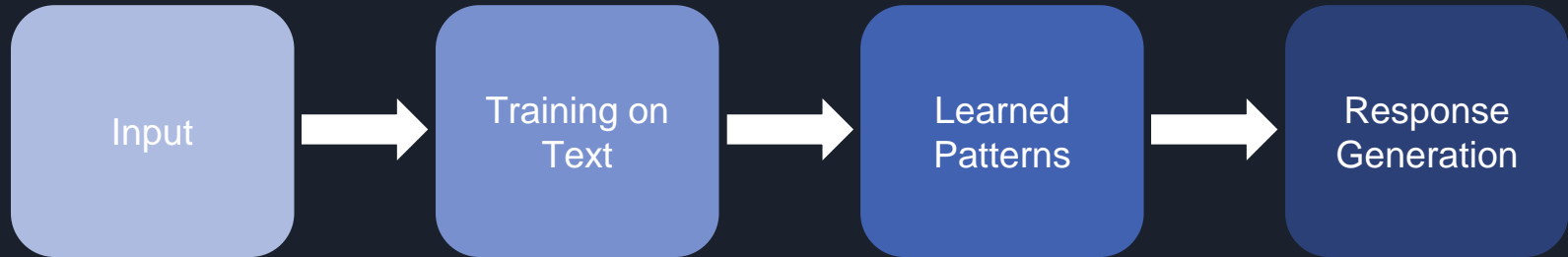
# Large language Models Fundamentals

- Trained on millions/billions of documents.
- Base model
- Only knows training data - cutoff date

What is the capital of France?

What is the capital of Germany?

What is the capital of Italy?



# Domain-Specific Knowledge in LLMs

## External knowledge (RAG) :

Leverages external documents for responses

### Benefits:

- Keeps responses current with low cost

### Challenges:

- Quality depends on document retrieval accuracy

## Internal (finetuning) :

Tailors the model to specific topics

### Benefits:

- Highly tailored model for specific tasks

### Challenges:

- Requires significant data, time and compute; risks overfitting

	Finetuning	RAG
Cost	High	Low
Scalability	Limited	High
Flexibility	Static (retraining)	Dynamic
Ideal Use Case	Specialized domains	Evolving domains

# Steps to Build a RAG Model



## Initial Setup:

- Preprocess data into manageable chunks
- Create dense/sparse embeddings to represent data
- Store embeddings in a vector database



## Retrieval:

- Embed the user query using the same model
- Match the query with relevant stored embeddings



## Generation:

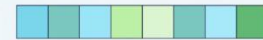
- Combine retrieved chunks with the query (prompt template)
- Use LLM to generate the final response

## Sparse embeddings



Sparse vectors are often high-dimensional with **many zero values**. They are generated from algorithms like BM25 and SPLADE and are used in **keyword-based search**.

## Dense embeddings



Dense embeddings contain mostly **non-zero values** and are generated from machine learning models. These vectors capture the semantic meaning of text and are used in **semantic search**.

# Showcase

The background features a series of dark grey, parallel lines that create a sense of depth and perspective, receding towards the right. A prominent green parallelogram is positioned in the upper right area, and a blue parallelogram is located below it, both appearing to be part of the layered structure.

CoC No	Name	Description	Expected	Value	Framework (latest version of the standard where not specified)	Evidence
3.3.1	Build resilience to business requirements	Only the level of resilience and therefore availability actually justified by business requirements and impact analysis should be built or purchased in the case of a collocation customer. 2N infrastructures are frequently unnecessary and inappropriate. If only a single level of resilience is available in the data centre an increased resilience or availability for critical services might be obtained by splitting the IT platform across multiple sites and making applications resilient to the loss of an individual site.	New build or retrofit	3	The organisation shall consider and document the level of resilience and availability actually justified by business requirements and impact analysis that should be built or purchased in the case of a collocation customer. Note 1: EN 50600-1 mandates a risk analysis to determine the level of redundancy of DC infrastructures to properly reflect the level of desired resilience and availability.	Review of Design Drawings/Project Scope
3.3.2	Consider multiple levels of resilience	It is possible to build a single data centre to provide multiple levels of power and cooling resilience to different floor areas. Many co-location providers already deliver this, for example, optional 'grey' power feeds without UPS or generator back up.	New build or retrofit	3	The organisation shall consider and document the provision of multiple levels of power and cooling resilience to different floor areas. Note: for example, optional 'grey' power feeds without UPS or generator back up.	Review of Design Drawings/Project Scope/Service Charging Models

JRC assessment  
framework  
ISO 50001  
ISO 14001  
ISO 30134

#### 4 Context of the organization

##### 4.1 Understanding the organization and its context

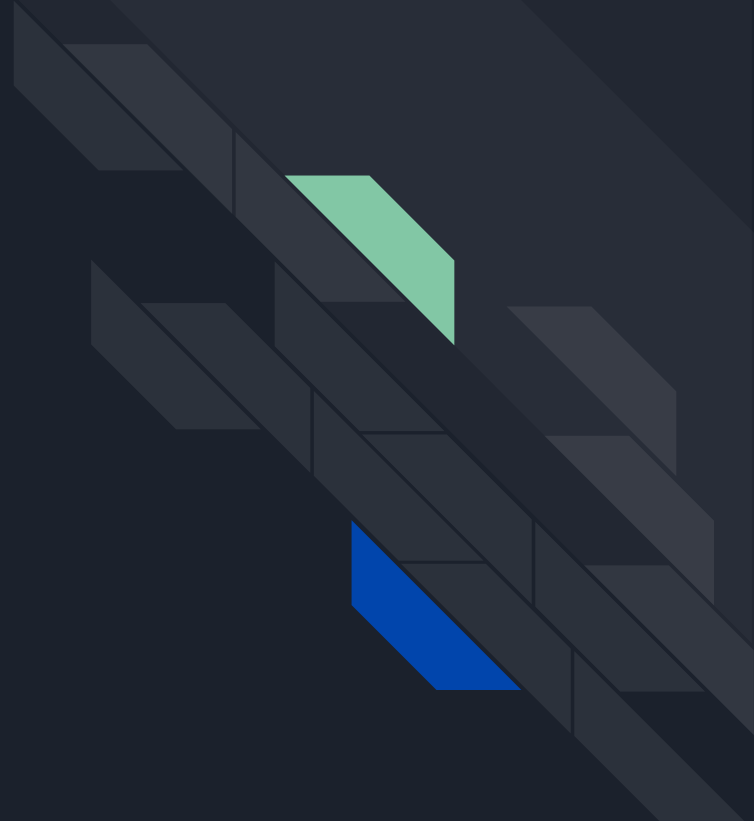
The organization shall determine external and internal issues that are relevant to its purpose and that affect its ability to achieve the intended outcomes of its environmental management system. Such issues shall include environmental conditions being affected by or capable of affecting the organization.

##### 4.2 Understanding the needs and expectations of interested parties

The organization shall determine:

- the interested parties that are relevant to the environmental management system;
- the relevant needs and expectations (i.e. requirements) of these interested parties;
- which of these needs and expectations become its compliance obligations.

Thank you for  
your attention





```
custom_template = PromptTemplate(
    template="""
```

```
You are a highly knowledgeable assistant specializing in the JRC
assessment framework (JRC code of conduct) and ISO standards
(50001, 14001, 30134).
```

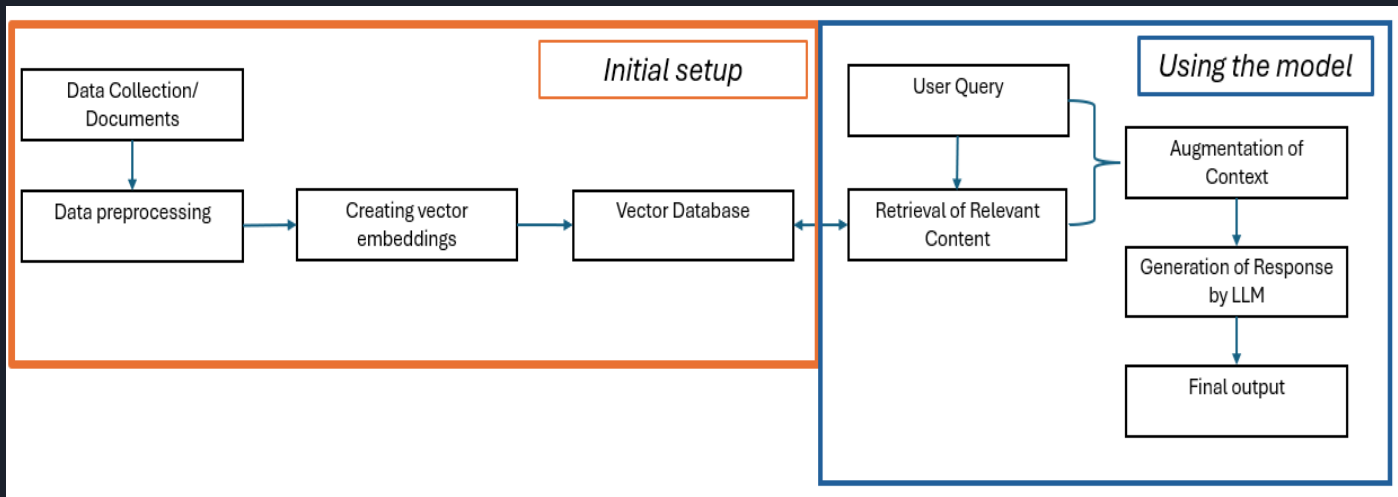
```
Your task is to generate a developer-focused framework by
interpreting relevant chunks from the provided content. Follow
these strict instructions:
```

1. **Focus on Developer Needs**:

- Provide actionable technical requirements for developers.
- Ensure the requirements are framed in plain, developer-friendly language.

2. **Structure**:

- Include clear headings for each requirement category.
- Break down requirements into:



```
5 Your query: How do I monitor and evaluate energy performance effectively?
6
7 Debug: conversational_system took 15.01 seconds
8
9 Bot: # Developer-Focused Framework for Monitoring and Evaluating Energy Performance
10
11 This framework is designed to help developers implement systems compliant with ISO 50001 standards for energy management. It provides actionable technical requirements focused on monitoring and evaluating energy performance effectively, based on the
12 provided content chunks.
13
14 ## 1. Monitoring and Measurement of Energy Performance
15
16 ### 1.1 Monitoring Energy Performance
17 - **Technical Details**:
18   - The system shall determine what needs to be monitored and measured, including:
19     - Effectiveness of action plans.
20     - Energy performance indicators (ENPIs).
21     - Operation of significant energy uses (SEUs).
22   - Actual vs. expected energy consumption.
23   - The methods for monitoring, measurement, analysis, and evaluation shall ensure valid results.
24   - Monitoring and measurement shall be performed at specified intervals.
25 - **Evidence Required**: Documented information on monitoring and measurement results must be retained.
26 - **Reference Frameworks**: ISO 50001, Section 9.1.1
27 - **Type of Requirement**: Mandatory
28
29 ### 1.2 Use of Monitoring Equipment
30 - **Technical Details**:
31   - Calibrated or verified monitoring and measurement equipment shall be used and maintained to ensure accurate data collection.
32 - **Evidence Required**: Documented information on calibration and verification.
33 - **Reference Frameworks**: ISO 14001, Section 9.1.1
34 - **Type of Requirement**: Mandatory
35
36 ## 2. Data Collection and Evaluation
37
38 ### 2.1 Data Collection Plan
39 - **Technical Details**:
40   - The organization shall define and implement an energy data collection plan, specifying necessary data, collection methods, and frequency.
41   - Measurement equipment must provide accurate and repeatable data.
42 - **Evidence Required**: Retain documented information on the data collection plan and measurement system.
43 - **Reference Frameworks**: ISO 50001, Sections 6.6 and 7.5
44 - **Type of Requirement**: Mandatory
45
46 ### 2.2 Evaluation and Analysis of Energy Performance
47 - **Technical Details**:
48   - The system shall evaluate energy performance improvements by comparing ENPI values against corresponding energy baselines.
49   - Limitations of data (e.g., accuracy, precision, measurement uncertainty) shall be considered in analysis.
50 - **Reference Frameworks**: ISO 50001, Section A.9.1
51 - **Type of Requirement**: Advisory
52
53 ## 3. Continual Improvement and Management Review
54
55 ### 3.1 Integration with Plan-Do-Check-Act Cycle
56 - **Technical Details**:
57   - The system shall be part of a continual improvement framework, implementing actions from data analysis and management review cycles.
58 - **Reference Frameworks**: ISO 50001, Section 0.3
59 - **Type of Requirement**: General
60
61 ### 3.2 Management Review Inputs
62 - **Technical Details**:
63   - Energy performance inputs to management review shall include extent objectives and energy targets met, energy performance improvement, and status of action plans.
64 - **Reference Frameworks**: ISO 50001, Section 9.3.3
65 - **Type of Requirement**: Mandatory
66
67 ## 4. Developer Notes
68 - Ensure configuration supports automatic data analysis for real-time insights.
69 - Maintain flexibility in system design to accommodate energy performance improvements through updates in technology and operational adjustments.
70
```