

# **Deep Learning: RAG System using Milvus & Hugging Face**

Jiani XU



# Project Overview

**Dataset: 《The Future of Sustainability: Navigating Trends and Innovations for a Sustainable Tomorrow》 (89 pages)**

- **Source:** VTT Technical Research Centre of Finland
- **System Functions:** User input → retrieve relevant context → generate answer
- **Tools used:** Hugging Face, Milvus, Sentence Transformers



FEBRUARY 2025 | MICHAEL HANF, LEAD SUSTAINABLE BUSINESS, VTT

## The future of sustainability

Navigating trends  
and innovations for  
a sustainable tomorrow

beyond the obvious

# Model use

## 1.Embedding & Milvus

### Current

**Embedding model:** BAAI/bge-base-en-v1.5  
**Vector size:** 768 dimensions

### Previous

**Embedding model:** all-MiniLM-L6-v2 &  
intfloat/e5-small-v2  
**Vector size:** 384 dimensions

## 2.LLM for answer generation

### the Flan-T5 series:

G google/flan-t5-xxl
Text2Text Generation • Updated Jul 27, 2023 • 302k • 1.24k
G google/flan-t5-base
Text2Text Generation • Updated Jul 17, 2023 • 1.56M • 848
G google/flan-t5-large
Text2Text Generation • Updated Jul 17, 2023 • 1.28M • 734
G google/flan-t5-xl
Text2Text Generation • Updated Nov 28, 2023 • 293k • 489
G google/flan-t5-small
Text2Text Generation • Updated Oct 10, 2023 • 624k • 331



# Prompt

```
PROMPT = """"  
### ROLE ###  
You are a sustainability expert AI assistant that ONLY responds based on verified information.  
  
### INSTRUCTION STEPS ###  
1. STRICTLY ANALYZE the user's question type:  
    - Factual question (what/where/when/who)  
    - Explanatory question (why/how)  
    - Significance question (importance/impact)  
  
2. CONTEXT PROCESSING:  
    Read the provided context EXACTLY as written.  
    Identify SPECIFIC sentences that DIRECTLY relate to the question.  
  
3. RESPONSE GENERATION RULES:  
    [REQUIRED] Answer MUST be grounded in explicit context statements  
    [REQUIRED] For "why" questions:  
        - Only explain reasons EXPLICITLY stated in context  
        - If no causation mentioned, use predefined response  
    [REQUIRED] When information is:  
        a) Fully available → Concise 1-3 sentence answer  
        b) Partially related → "I cannot find..." response  
        c) Missing → "I cannot find..." response  
    [PROHIBITED] Never:  
        - Assume unstated connections  
        - Combine information from different sections  
        - Use examples not in context  
  
### FORMAT CONTROL ###  
- No markdown of any kind  
- Avoid transitional phrases ("However", "Additionally")  
- Use bullet points ONLY when listing explicit items from context  
  
### CONTEXT ###  
{context}  
  
### QUESTION ###  
{question}
```

```
### SAFETY PROTOCOL ###  
If uncertain about ANY part of the response, immediately fallback to:  
"I cannot find relevant information in the provided documents."  
  
### RESPONSE ###  
.....  
  
def build_prompt(context: str, question: str) -> str:  
    sanitized_context = " ".join(context.strip().splitlines())  
    sanitized_question = question.strip().replace("\n", " ")  
  
    max_context_length = 3000  
    truncated_context = (sanitized_context[:max_context_length] + "...") if len(sanitized_context) > max_context_length else sanitized_context  
  
    return PROMPT.format(context=truncated_context, question=sanitized_question)
```

# The 1st question: What is the purpose of the “Future of Sustainability” report?



## Sustainability Whitepaper Q&A

Ask questions about the *Future of Sustainability* 2025 whitepaper.

⟳ Rebuild vector DB (for dev/debug)

Enter your question:

What is the purpose of the “Future of Sustainability” report?

### Answer

The primary purpose of this study is to identify emerging trends and megatrends while offering forward-looking perspectives for businesses, policymakers, and organisations.

Debug info ▾

# The 2nd question: What are some examples of green finance mentioned in the report?



## Sustainability Whitepaper Q&A

Ask questions about the *Future of Sustainability 2025* whitepaper.

Rebuild vector DB (for dev/debug)

Enter your question:

What are some examples of green finance mentioned in the report?

### Answer

Green bonds, impact investing, and sustainability-linked loans are examples of how finance is evolving to support businesses that contribute to climate resilience, social inclusion, and environmental restoration.

Debug info



**Thank  
you very  
much!**

