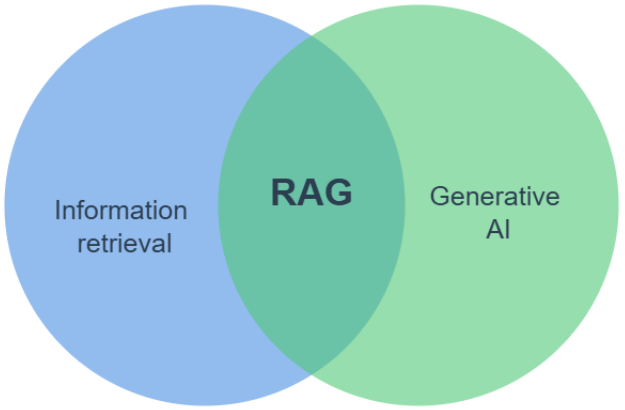
AI Assisted Literature Review: Query Search with RAG

Presenter: Shabnam Barbhuiya

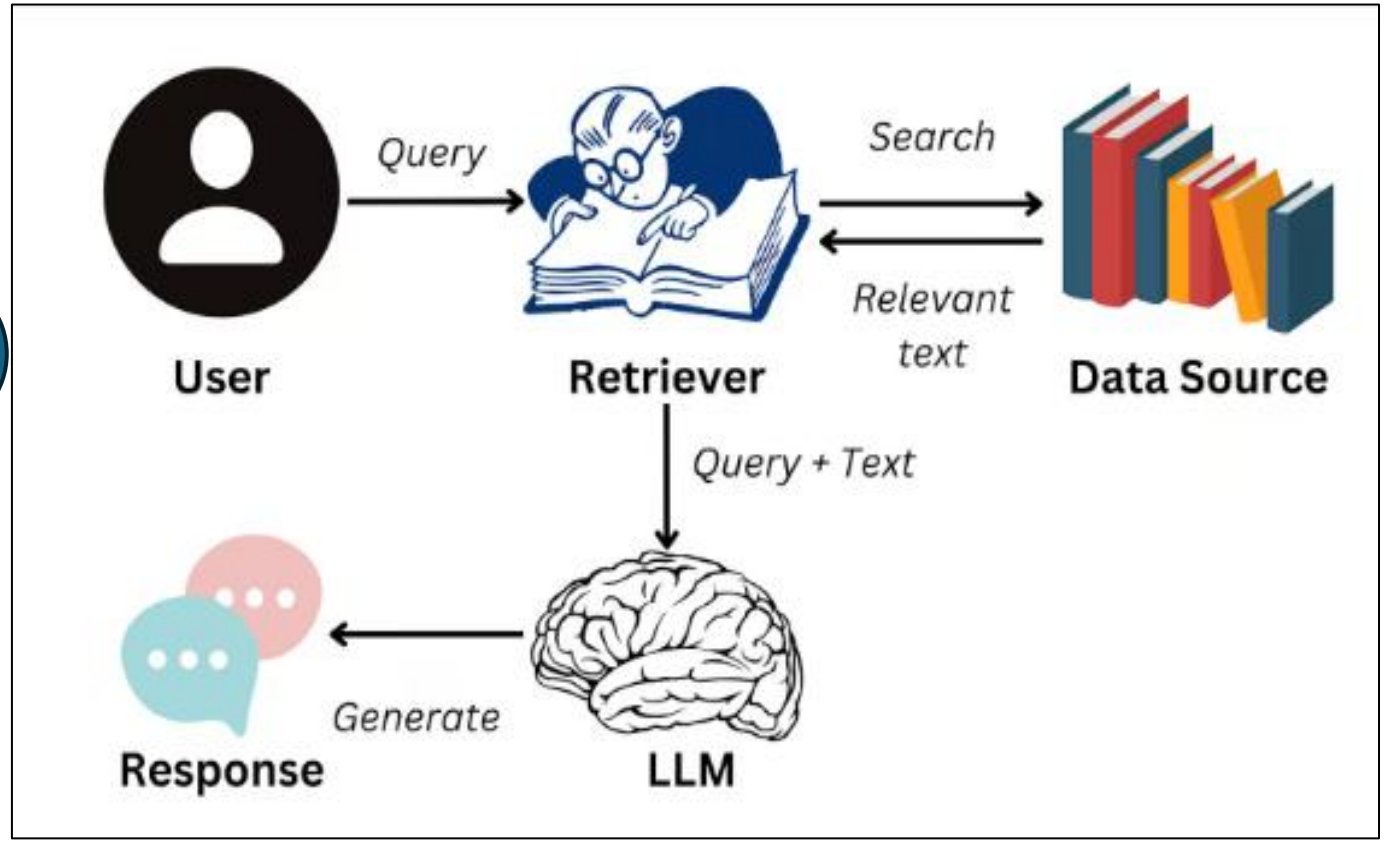
From Summarization to Smart Querying with RAG

Retrieval-Augmented Generation (RAG)

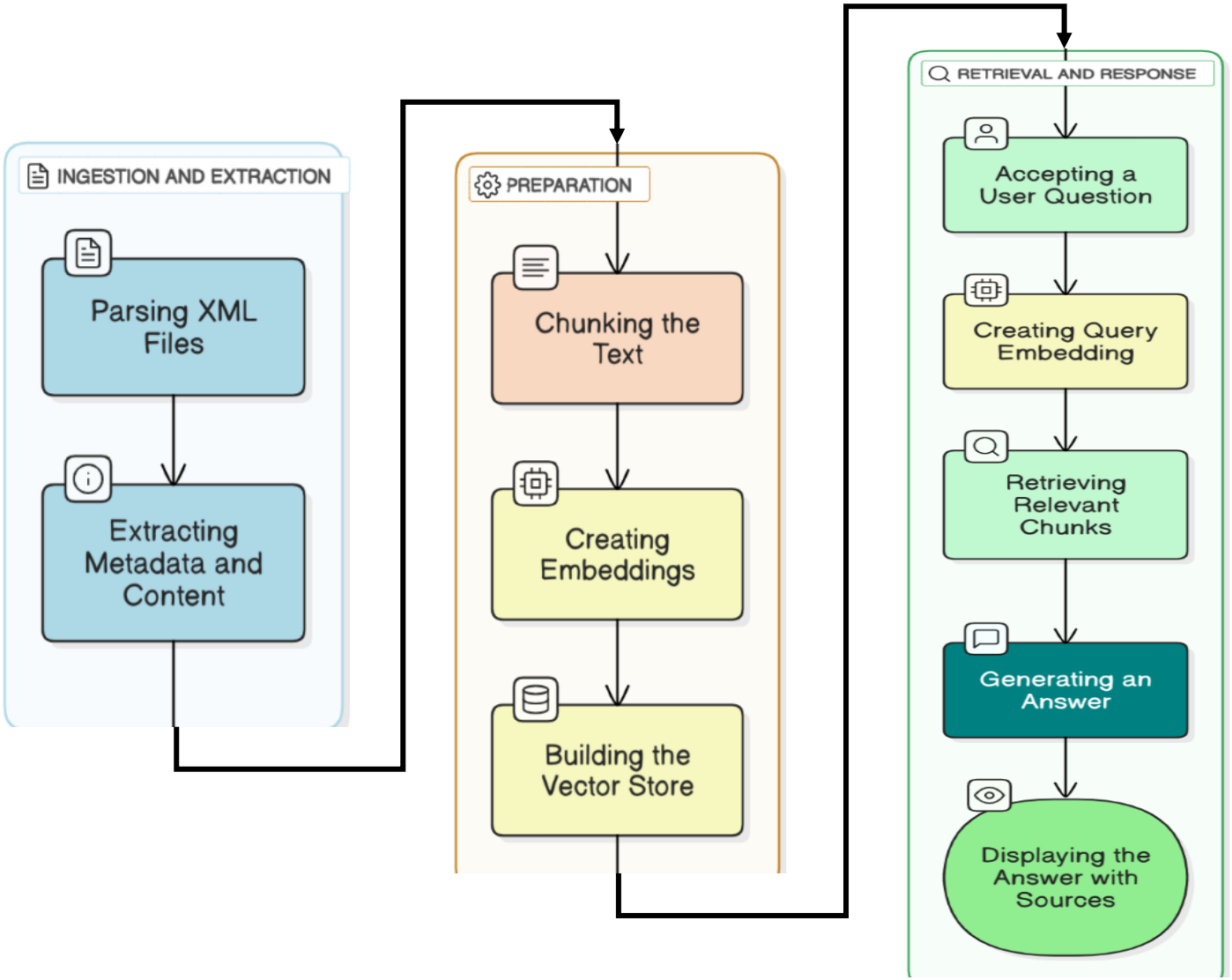


A hybrid framework that combines:
Retriever – pulls relevant data from documents
Generator – answers your question using that data

How RAG Delivers Grounded Answers



Query Search / RAG Workflow



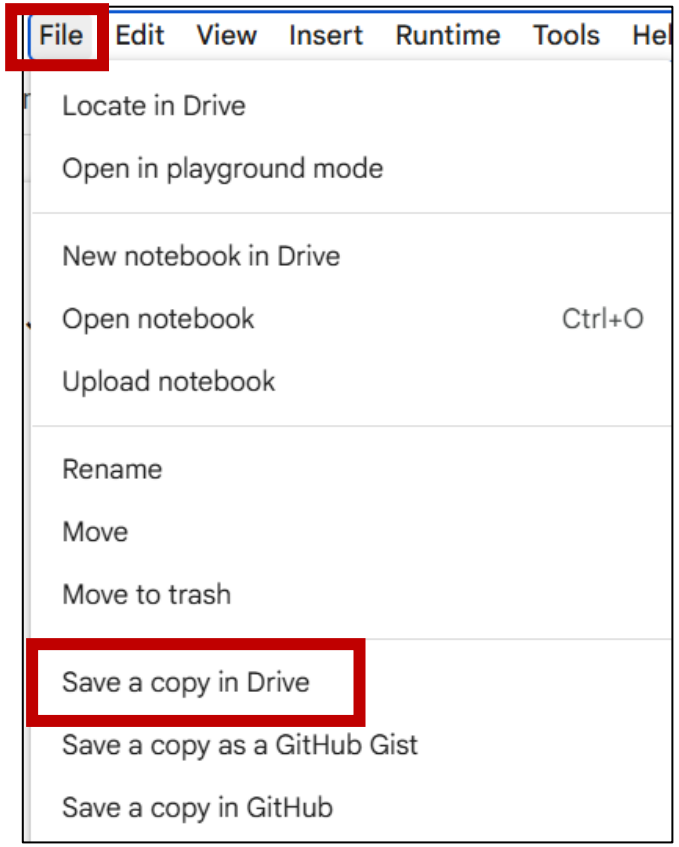
Library Used:

- **Sentence-BERT** – Converts text and queries into semantic vectors
- **FAISS** – Fast Approximate Nearest Neighbor Search
- **Markdown Conversion** – XML → Markdown before embedding
- **LLaMA 3 via Groq** – Generates the final answer
- **LangChain** – Orchestrates the end-to-end workflow

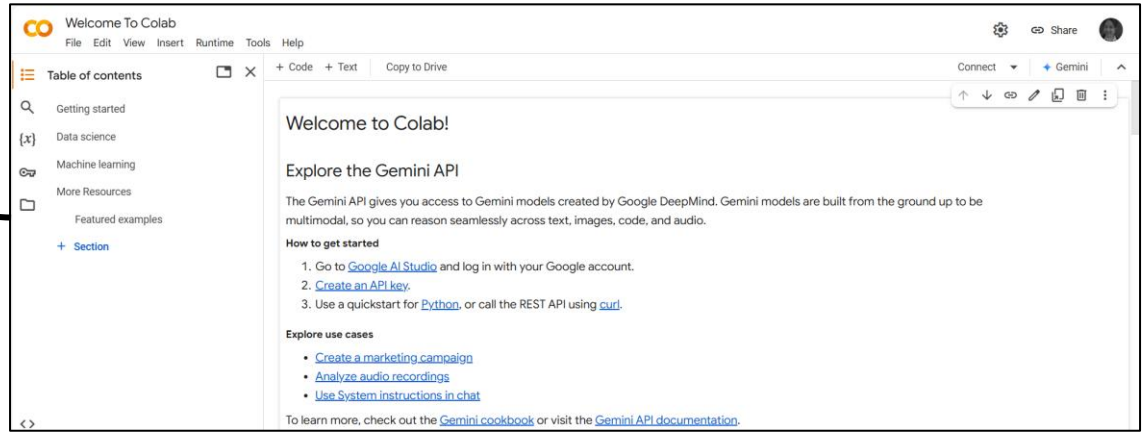
Colab Link

INSTRUCTION

Setting Up Your Colab Notebook



Home Page



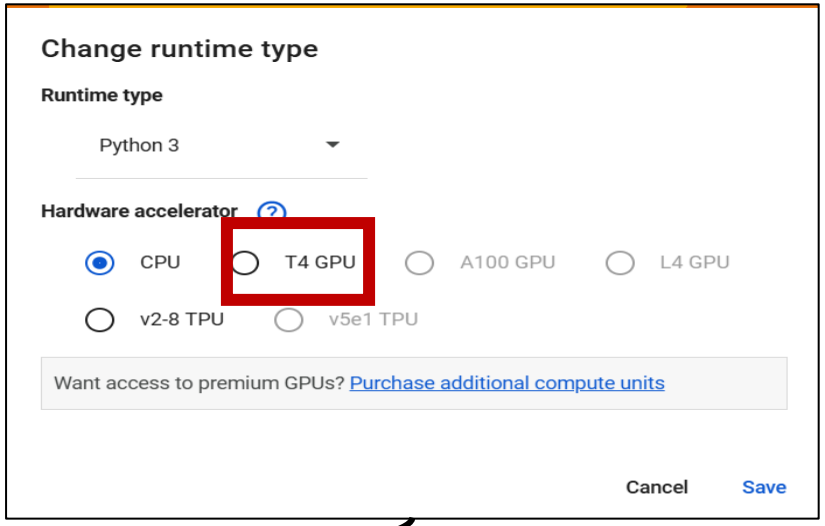
1. Copy the Colab Notebook to Your Own Google Drive Go to the menu and click:

File → Save a copy in Drive

2.Change the Runtime to GPU

In the Colab menu, select:

Runtime → Change runtime type



Query Search Colab Overview(DIY)

Step 1 – Install Required Libraries

```
[ ]  # Install dependencies
     !pip install pygetpapers
     !pip install lxml
     !pip install langchain chromadb sentence-transformers
     !pip install -U langchain-community langchain-groq
     !pip install pymupdf markdown2 weasyprint
```

- Go to https://console.groq.com/
- Create an account if you don't have one
- Generate your API token
- Copy and paste it when prompted below
- CLICK **ENTER** once done.

Step 2-Set Groq API Key

```
#  Set API Key
import os, getpass
os.environ["GROQ_API_KEY"] = getpass.getpass("🔐 Enter your Groq API Key: ")
```

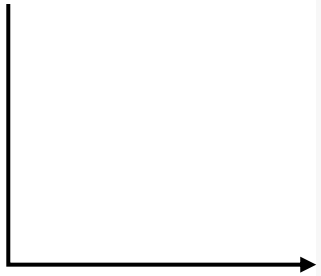
```
🔐 Enter your Groq API Key: ··········
```

Query Search Colab Overview(DIY)

Step 3-Download Research Papers

```
[ ]  # Download papers from EuropePMC
     !pygetpapers --query '"Climate change"' --xml --limit 2 --output /content/data_climate --save_query
```

Step 4-Parse XML/PDF Files to Markdown



```
#  Parse XMLs to Markdown and extract metadata
import pathlib
import re
from lxml import etree
import fitz  # PyMuPDF
from datetime import datetime

def sanitize_filename(name):
    return re.sub(r'[\/:"*?<>|]+', "_", name)

def extract_text_from_pdf(pdf_path):
    try:
        doc = fitz.open(pdf_path)
        text = ""
        for page in doc:
            text += page.get_text()
```

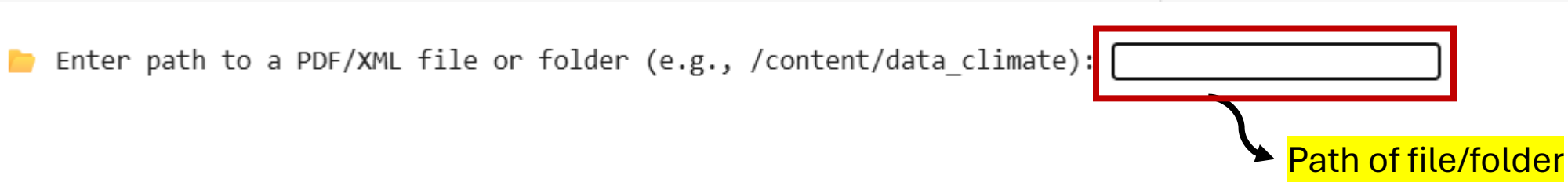
Query Search Colab Overview(DIY)

Step 5-Create Vector Database

```
def load_markdown_documents_with_metadata(metadata_records):
    documents = []
    for md_path, metadata in metadata_records:
        text = md_path.read_text(encoding="utf-8")
        if text.strip():
            doc = Document(page_content=text, metadata=metadata)
            documents.append(doc)
    return documents
```

Step 7-Query your newly created RAG/LLM

```
user_path = input("📂 Enter path to a PDF/XML file or folder (e.g., /content/data_climate): ").strip()
markdown_dir = "/content/markdowns"
os.makedirs(markdown_dir, exist_ok=True)
```

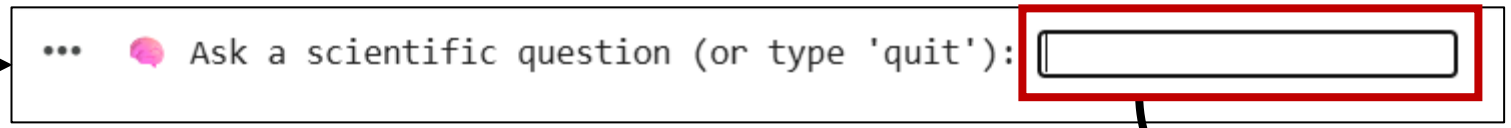


Query Search Colab Overview(DIY)

Step 7- Query your newly created RAG/LLM

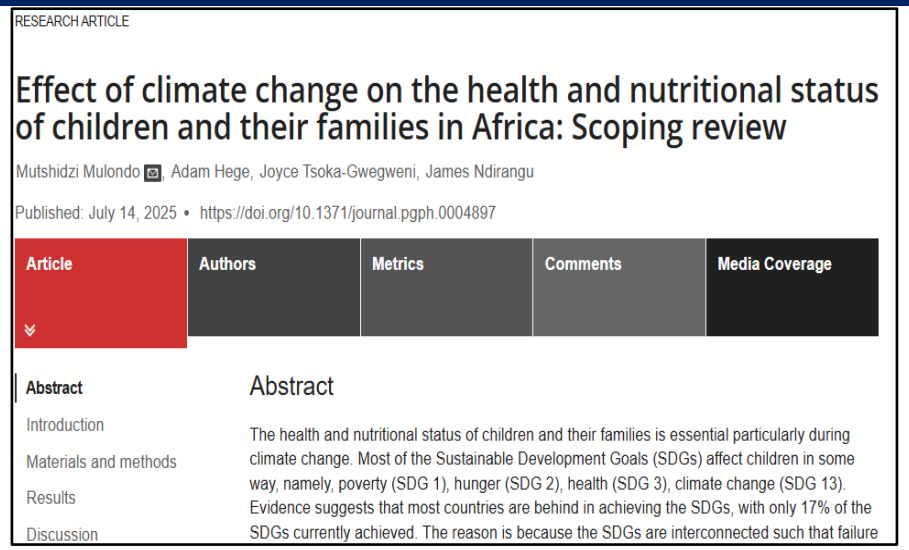
```
qa_log_md = "/content/QA_Log.md"
with open(qa_log_md, "w", encoding="utf-8") as log_file:
    log_file.write(f"# Q&A Log - {datetime.now().strftime('%Y-%m-%d %H:%M:%S')}\n\n")

    while True:
        query = input("🧠 Ask a scientific question (or type 'quit'): ").strip()
        if query.lower() == "quit":
            print(f" Q&A Markdown log saved to {qa_log_md}")
            break

        result = qa_chain.invoke({"query": query})
        answer = result.get("result", "")
```

Interface



Write your query

OUTPUT OVERVIEW



RESEARCH ARTICLE

Effect of climate change on the health and nutritional status of children and their families in Africa: Scoping review

Mutshidzi Mulondo, Adam Hege, Joyce Tsoka-Gwegweni, James Ndirangu

Published: July 14, 2025 • https://doi.org/10.1371/journal.pgph.0004897

Article | Authors | Metrics | Comments | Media Coverage

Abstract
Introduction
Materials and methods
Results
Discussion

Abstract

The health and nutritional status of children and their families is essential particularly during climate change. Most of the Sustainable Development Goals (SDGs) affect children in some way, namely, poverty (SDG 1), hunger (SDG 2), health (SDG 3), climate change (SDG 13). Evidence suggests that most countries are behind in achieving the SDGs, with only 17% of the SDGs currently achieved. The reason is because the SDGs are interconnected such that failure

Scientific Paper

```
Ask a scientific question (or type 'quit'): what is the effect of climate on nutritional status?
```

Answer:

Based on the provided context, climate change events can have a negative impact on nutritional status, particularly among vulnerable populations such as children and adults. The effects of climate change on nutritional status can be influenced by personal and socio-demographic, economic, and environmental factors.

While the context does not provide a direct answer to the question, it suggests that climate change can lead to poor nutritional status, which can further exacerbate health problems. The references provided also support this notion, highlighting the impact of climate change on child health and nutrition.

For example, Reference 3 (Helldén et al., 2021) mentions the importance of considering the effects of climate change on child health, including nutrition. Similarly, Reference 1 (Bhutta et al., 2019) emphasizes the need for paediatricians to address the impacts of climate change on child health, which likely includes nutritional status.

In summary, while the context does not provide a direct answer, it implies that climate change can have a negative impact on nutritional status, particularly among vulnerable populations.

Sources:

- Effect of climate change on the health and nutritional status of children and their families in Africa: Scoping review
- Plos Global Public Health | Https://Doi.Org/10.1371/Journal.Pgph.0004897 July 14, 2025

Output:

```
Ask a scientific question (or type 'quit'): effects of climate change in africa
```

Answer:

Based on the provided context, the effects of climate change in Africa include:

1. Malnutrition
2. Infectious diseases
3. Respiratory diseases in children and adults
4. Adverse pregnancy and birth outcomes
5. High child and maternal morbidity and mortality
6. Mental health problems

These health conditions are associated with various climatic change phenomena or events, such as:

1. High temperatures
2. Drought
3. Floods
4. Wildfires
5. Air pollution

Additionally, the review highlights that personal and socio-demographic, economic, and environmental factors increase the risk of people to the effects of climate change events on poor nutritional and health status.

!!!!Try It Yourself!!!!

COLAB LINK:
https://colab.research.google.com/drive/17J9wEvkQvdaeOihN3N13u_ln5Oez8ssd#scrollTo=i5iXgh2ZMVUp

JOIN OUR DISCORD COMMUNITY:

THANK YOU

#semanticClimate
Transforming information into actionable knowledge