

AI Agent with RAG and Tool Calling

Vikram Kumar

vikramkumarsj@gmail.com

91-9880491190

Contents:

1. Objectives
2. Solution Architecture
3. Technical Modules
4. AI Insights
5. Future Improvements

Objectives:

Technical Solution to build a function-calling AI agent that combines:

- A chat interface for query based Insights generation on USGS Retrieval-Augmented Generation (RAG) dataset containing **USGS Research and daily USGS water status dataset**.
- Structured tool/API calling using a USGS public API for new datasets and save to existing RAG.
- Demonstrate ability to design and implement intelligent systems that combine reasoning, retrieval (latest), and external tool use.

USGS Data Analysis

☐ Turn on function calling

Prompt

In summary what are the essential components of water level monitoring programs.?

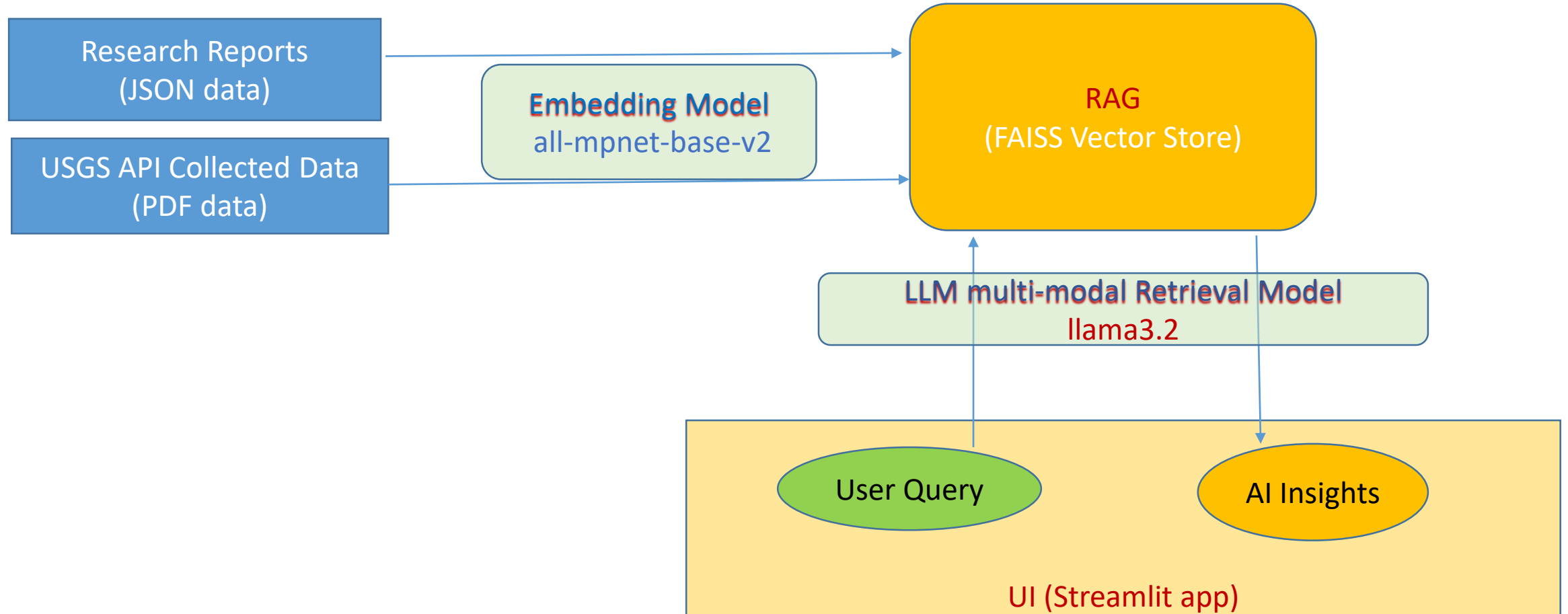
Submit

The essential components of water-level monitoring programs include:

1. Selection of observation wells
2. Determination of the frequency of water-level measurements
3. Implementation of quality assurance
4. Establishment of effective practices for data reporting.

> Chat messages

Solution Architecture:



Core Modules in the Codebase:

Module	Type	Details		
'ApiData' / 'ApiDataNew' / 'research_reports'	Historical and daily USGS Datasets queried through functional calling API Folder to save supporting research reports	Data used to build RAG https://api.waterdata.usgs.gov/docs/ogcapi/keys/ Data used to build RAG https://pubs.usgs.gov/circ/circ1217/pdf/circular1217.pdf		
functioncalling-rag-main.py	code folder	Readme and execution files		
streamlit_function_calling.py	Streamlit AI Chat interface python code	Frontend UI for USGS RAG data analysis.		
function_calling_new_dataset.py	Code to collect data from USGS API	Frontend UI for USGS RAG data analysis.		
vector_data_manager.py / vector_data_manager_for_research_reports.py 'faiss_index_USGS'	Python Module to add/update datasets to RAG FAISS Vector data store	FAISS vector database Embedding Model: sentence-transformers/all-mpnet-base-v2		
RAG_with_RetrievalQA.py	Core Module to query RAG and generate AI output	Gen AI Chat/Query LLM Model (llama3.2) for context based search.		

Insights:

Research Data Query

USGS Data Analysis

☐ Turn on function calling

Prompt

In summary what are the essential components of water level monitoring programs.?

Submit

The essential components of water-level monitoring programs include:

1. Selection of observation wells
2. Determination of the frequency of water-level measurements
3. Implementation of quality assurance
4. Establishment of effective practices for data reporting.

> Chat messages

USGS Data Collections Query

Prompt

publish all the FeatureCollection dataset in organized layout.

Submit

Based on the provided context, I will attempt to organize the FeatureCollection datasets into a structured format.

USGS Daily Water Statistics

The following are the USGS daily water statistics organized by monitoring location:

1. **USGS-02198955**
 - Date: 2025-09-01
 - Parameter Code: 00010 (Temperature in degrees Celsius)
 - Value: 26.1
2. **USGS-03067510**
 - Date: 2025-09-01
 - Parameter Code: 63680 (Flow rate in units per second, FNU)
 - Value: 1.0

AI Insights for Consolidated dataset

RAG Query:

Review all the dataset for daily water statistics (provided in FeatureCollection datasets) by monitoring location and check if it is as per recommendations on essentials components of water level monitoring program mentioned in the research report.

To review all the datasets for daily water statistics (provided in FeatureCollection datasets) by monitoring location and validate if it is as per recommendations on essentials components of water-level monitoring programs mentioned in the research report:

1. Selection of Observation Wells: Unfortunately, I don't know if this information is available in the provided datasets.
2. Determination of the frequency of water-level measurements: I don't have access to real-time data or specific details about the frequency of measurements for each monitoring location.
3. Implementation of quality assurance: The provided context does not mention any quality assurance metrics for the daily water statistics datasets.

However, I can provide an example for a specific monitoring location:

Monitoring Location (example monitoring_location_id "USGS-02198955") provides details on the approval status ("Provisional") and water level value (26). However, I don't have information on whether this location covers details on quality assurance metrics as recommended in the research report on best practices for water level monitoring programs.

Regarding data reporting and other essential components of the water-level monitoring program, I don't have enough information to provide a detailed answer. The provided context only mentions that many agencies lack formalized written plans for the design and operation of ground-water-level networks, but it does not provide specific details on this topic.

Future Improvements:

1. More data collections for RAG adding Research papers / Monitoring Centers.
2. Fine Tune API Query as per data needs.
3. Performance review of embedding models & LLM Q&A models.
4. Source JSON data conversions to nodes to efficient data storage and retrieval.
5. Improve prompt response accuracy with more examples of Q&A in the context setting, Exploring domain specific custom Q&A LLM Model training.
6. Enable better explain output through use of Reasoning LLM Models, enabling parameter 'return_source_documents' = True for FAISS retrievals.
7. More meaningful user queries testing against the dataset.
8. Image data handling and data chunking strategies.