Introduction to RAG & Fine-Tuning

Building Smarter AI with Your Own Data

Thomas Torku, Ph.D.

Ghana AI Research Network (GAIN)
Data Professional — Part-Time Lecturer, The Ohio State University (OSU)

GAIN Dialogue Series
January 2026 Edition

Outline

Outline

The Problem with Generic AI

LLMs Don't Know Your Data

- GPT doesn't know Ghana's 2021 census results
- Doesn't understand your company's data
- No access to your research papers

The Problem with Generic AI

LLMs Don't Know Your Data

- GPT doesn't know Ghana's 2021 census results
- Doesn't understand your company's data
- No access to your research papers

No Local Context

- Doesn't understand "Ashanti Region"
- Doesn't know "GLSS" surveys
- Missing Ghana-specific terminology

Hallucinations

- Makes up statistics
- Sounds plausible but wrong
- No source verification

The Problem with Generic AI

LLMs Don't Know Your Data

- GPT doesn't know Ghana's 2021 census results
- Doesn't understand your company's data
- No access to your research papers

Hallucinations

- Makes up statistics
- Sounds plausible but wrong
- No source verification

No Local Context

- Doesn't understand "Ashanti Region"
- Doesn't know "GLSS" surveys
- Missing Ghana-specific terminology

The Solution

Combine two powerful techniques:

- **RAG** - Retrieval-Augmented Generation
- **Fine-Tuning** - Domain Adaptation

Outline

What is RAG?

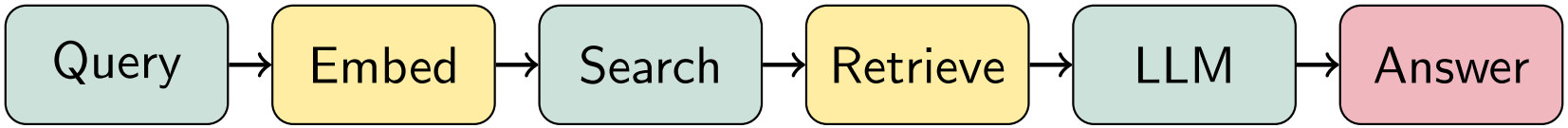
Definition

RAG (Retrieval-Augmented Generation) = Fetch relevant documents first, then generate answers based on those documents.

What is RAG?

Definition

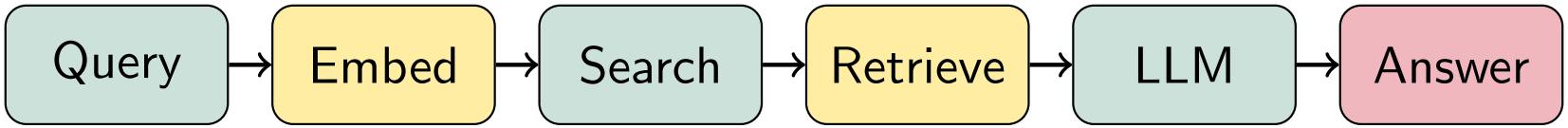
RAG (Retrieval-Augmented Generation) = Fetch relevant documents first, then generate answers based on those documents.



What is RAG?

Definition

RAG (Retrieval-Augmented Generation) = Fetch relevant documents first, then generate answers based on those documents.



Grounded	No Hallucinations	Citable	Updatable
Answers backed by real documents	Can't make up facts	Shows sources	Add new data easily

Outline

What is Fine-Tuning?

Definition

Fine-Tuning = Train an existing model on your data to improve domain performance.

Before Fine-Tuning

Query: "Kumasi region population"

- ✗ Retrieves wrong documents
- ✗ Doesn't know Kumasi = Ashanti
- ✗ Generic keyword matching

After Fine-Tuning

Query: "Kumasi region population"

- ✓ Returns Ashanti Region: 5.4M
- ✓ Learned Ghana terminology
- ✓ Semantic understanding

Today: Fine-Tuning all-MiniLM-L6-v2

Teaching: "Kumasi" → "Ashanti", "GLSS" → "Living Standards"

Outline

RAG + Fine-Tuning = Better Together

RAG Alone	Fine-Tuning Alone	Combined
✓ Grounds responses	✓ Domain understanding	✓ Accurate retrieval
✗ May miss relevant docs	✗ No data access	✓ Domain-aware
		✓ Grounded answers

RAG + Fine-Tuning = Better Together

RAG Alone	Fine-Tuning Alone	Combined
✓ Grounds responses	✓ Domain understanding	✓ Accurate retrieval
✗ May miss relevant docs	✗ No data access	✓ Domain-aware
		✓ Grounded answers

Our Architecture

Query → **Fine-Tuned Embeddings** → **ChromaDB** → **Census Docs** → **OpenAI GPT** → **Answer + Sources**

RAG + Fine-Tuning = Better Together

RAG Alone	Fine-Tuning Alone	Combined
✓ Grounds responses	✓ Domain understanding	✓ Accurate retrieval
✗ May miss relevant docs	✗ No data access	✓ Domain-aware
		✓ Grounded answers

Our Architecture

Query → **Fine-Tuned Embeddings** → **ChromaDB** → **Census Docs** → **OpenAI GPT** → **Answer + Sources**

✓
Better Retrieval for Ghana Terms

✓
Cited Sources with Every Answer

2 hrs
Build & Deploy Today

Outline

What We'll Build Today

Workshop Modules

1. **Data Preparation** *20 min*
   Load & chunk Ghana census data
2. **RAG Pipeline** *25 min*
   Embeddings + ChromaDB + GPT
3. **Fine-Tuning** *25 min*
   Train on Ghana-specific pairs
4. **Streamlit Deploy** *25 min*
   Build & launch chat interface

Tech Stack

- **Python** 3.10+
- **all-MiniLM-L6-v2**
- **Sentence-Transformers**
- **ChromaDB**
- **OpenAI GPT-4o-mini**
- **Streamlit**

What We'll Build Today

Workshop Modules

1. **Data Preparation** *20 min*
   Load & chunk Ghana census data
2. **RAG Pipeline** *25 min*
   Embeddings + ChromaDB + GPT
3. **Fine-Tuning** *25 min*
   Train on Ghana-specific pairs
4. **Streamlit Deploy** *25 min*
   Build & launch chat interface

Tech Stack

- **Python** 3.10+
- **all-MiniLM-L6-v2**
- **Sentence-Transformers**
- **ChromaDB**
- **OpenAI GPT-4o-mini**
- **Streamlit**

Data Source

Ghana Statistical Service
2021 Population & Housing Census

Outline

Let's Build!

Ghana Census Q&A Agent

with RAG + Fine-Tuned Embeddings