**Enhancing Medication Safety with Retrieval-Augmented Clinical Question Answering**

This project explores whether grounding large language models (LLMs) in structured FDA drug label data (via the OpenFDA API) improves factual accuracy and faithfulness in clinical question answering (QA), particularly for high-risk medications.

**📖 Project Overview**

Medication errors remain a critical issue in clinical care. While LLMs can assist with medication-related QA, they are prone to hallucinations. We implemented and evaluated Retrieval-Augmented Generation (RAG) pipelines using FDA drug label data to enhance answer reliability and accuracy.

🔗 Live Demo: [medguardrx.com](https://www.medguardrx.com/)

**🏗️ Architecture Summary**

A screenshot of a computer program

Description automatically generated

Query (Clinical Question)

↓

Drug Name Extraction (LLM)

↓

OpenFDA Drug Label API Lookup

↓

Chunking (300 / 600 tokens)

↓

Embedding + Vector Store (ChromaDB)

↓

Context Retrieval (Cosine Similarity / MMR)

↓

LLM Response (gpt-4o / gemini-2.5-pro)

* **Hard grounding prompts:** "Answer the question based only on the following context..."
* **Fallback:** None (future work: retrieval confidence fallback)
* **Baselines**: LLMs with no grounding

**📊 Evaluation**

| **Tool** | **Metric** | **Notes** |
| --- | --- | --- |
| RAGAs | answer\_relevancy |  |
| RAGAs (F1) | factual\_correctness | Low scores (0.16–0.26), context-sensitive |
| RAGAs | context\_recall | Faithfulness improved with grounding |
| RAGAs | faithfulness | Faithfulness improved with OPENFDA grounding |
| LLM-as-judge | quality, correctness, clarity | Less useful than RAGAs |
|  |  |  |

**📁 Repository Structure**

* /notebooks/：Contains all code related to running and evaluating RAG pipeline

- [OpenFDA\_RAGAs\_v1.1.ipynb](https://github.com/nkabram/capstone/blob/main/notebooks/OpenFDA_RAGAs_v1.1.ipynb)

* README.md: Project overview and usage instructions

**✅ Key Findings**

* OpenFDA grounding improves faithfulness but does not necessarily always improve accuracy, especially when source coverage is incomplete
* MMR retrieval occasionally retrieved more diverse or contextually aligned chunks compared to similarity-only retrieval
* Baseline (ungrounded) responses were often accurate and relevant in our sample set
* A flexible prompting strategy that encourages context use while not disabling model prior knowledge is likely to lead to more clinically useful answers

**🚀 Future Directions**

* Expand the dataset with more diverse, clinician-submitted queries
* Combine OpenFDA + other clinical data sources (e.g., MIMIC\_IV)
* Design RAG prompts with fallback behavior based on retrieval confidence.

**📦 Requirements and Setup**