

MEDMAP: Bridging Medical Deserts with Agentic AI

Track: DataBricks sponsored | **Team:** Cempasuchil (Milena Fernanda Rivera Hernandez & Valentina Ayelen Cruz Mendoza)

I. Executive Summary. By 2030 the world will face a shortage of healthcare professionals, a crisis driven not by a lack of expertise, but a failure of coordination. MedMap is an intelligent Document Parsing and Reasoning Agent built on Databricks Data Intelligence Platform. Serving as an intelligence layer for the Virtue Foundation, transforming unstructured field notes into actionable geospatial insights to route life-saving resources effectively.

II. Core Objectives. The main goal of this project was to move beyond simple *search* and create a system that can *think* about healthcare data.

- Structure the Unstructured: Convert messy PDFs and field notes into queryable structured data using IDP
- Integrity verification: Detects incomplete or suspicious claims about hospital capabilities.
- Identify Medical Deserts: Geographically pinpoint regions where critical procedures are unavailable within a safe travel distance to optimize NGO intervention, and map them in a dynamic visualization.

III. Technical Architecture

- **Governance: Unity Catalog** secures all unstructured data (Volumes) and structured verified records (Delta Tables).
- **Reasoning Engine: Llama 3.3 70B Instruct** (via Databricks Model Serving) serves as the core "Brain" for tool-calling and logic, supplemented by **OpenAI GPT-4o** for complex context.
- **Vector RAG: Databricks Vector Search** (using BGE-Large-EN embeddings) enables semantic retrieval, allowing the agent to find facilities based on *capabilities* rather than just keywords.
- **Orchestration: LangGraph** manages the ReAct (Reason+Act) loop, dynamically selecting tools for SQL lookups, math verification, or geospatial distance calculation.
- **User Interface: A Streamlit dashboard**, deployed as a native **Databricks App**, provides a "Command Center" for planners.

IV. Methodology & implementation. Our solution was built in four phases:

1. **Data Ingestion:** We established a Unity Catalog pipeline to ingest raw field reports, enabling **Change Data Feed (CDF)** to keep the knowledge base real-time.
2. **Intelligent Extraction:** Using Pydantic models, we built an IDP pipeline that extracts key entities (Bed Count, Equipment, Staffing) from free text, feeding a "Gold" Delta Table.
3. **The Integrity Agent:** We engineered a Multi-Agent system equipped with custom tools.
4. **Impact Visualization:** The final output is an interactive map where NGO workers can "Click-to-Audit" any region. The system visualizes data in a Traffic Light system powered by **MLflow Tracing** to provide transparent reasoning for every flag.

V. Results and conclusions. MedMap's deployment on the Ghana dataset yielded critical geospatial insights, successfully distinguishing "Surgical Deserts" in the Northern Region (Tamale/Wa) from the saturated services in the South (Accra). The Agentic Integrity Engine flagged approximately 15% of records as high-risk anomalies, automatically detecting "hallucinated capacity" where clinics claimed advanced capabilities without verified infrastructure. This automated IDP pipeline processed field notes 100x faster than manual auditing, transforming months of work into seconds. In conclusion, MedMap proves that Agentic AI is a viable infrastructure for solving planetary-scale coordination failures. By transitioning from static databases to a dynamic reasoning engine, the Virtue Foundation can now direct resources based on verified need rather than unverified claims. The modular Unity Catalog and Vector RAG backbone ensures this architecture is instantly scalable, ready to map medical deserts in future target regions like Honduras or Mongolia.