Bridging Medical Deserts: AI-Powered Healthcare Intelligence System

Databricks Sponsored Track | Virtue Foundation Challenge

Team Members: Arturo Favara, Emma Sampietro
Emails: favara@mit.edu, esampi@stanford.edu

1 Challenge and Objective

NGO planners often work with incomplete and inconsistent facility data. Our objective is to turn unstructured hospital records into an explainable decision layer for regional gap detection and faster staffing/funding prioritization.

2 What We Built (Hackathon MVP)

We implemented an end-to-end pipeline tailored to the Virtue Foundation Ghana dataset:

- **Parsing and normalization:** deduplicate facilities, clean sparse fields, and preserve full row context (source links, location, and capability text) for downstream traceability.
- **Evidence-scored hospital agent:** for each facility, an LLM outputs structured CSV files over six capability categories (score, summary, evidence gaps, sources), with schema validation, targeted retries, and model fallback for reliable batch runs.
- **Policy-ready region analytics:** enrich Ghana regional data with demographics, care proxies, and composite gap metrics that support maternal risk ranking, staffing prioritization, and threat-flag triage.
- **Frontend handoff assets:** Lovable-ready CSVs (hospitals + region analytics + indicator glossary) power map visualization and natural-language queries with interpretable fields.

3 Current Artifacts and Scale

- **Input scale:** 987 raw facility rows, 894 unique hospital names, 41 columns.
- **Hospital output bundle:** 477 facilities in the merged frontend dataset, including six category score columns and traceable evidence fields.
- **Regional policy layer:** 21 Ghana regions with 27 columns, including 5 binary threat indicators for prioritization.

This gives planners both macro (region) and micro (facility) views in the same workflow.

4 How This Maps to Judging Criteria

Technical Accuracy: validation, retries, and explicit uncertainty reduce silent failure.
IDP Innovation: free text is converted into auditable capability profiles, not keyword matches.
Social Impact: outputs rank where care capacity is weakest for NGO allocation.
User Experience: map + plain-language Q&A is designed for non-technical planners of all ages.

5 Demo Flow

In the demo, a user (medical volunteer/NGO/policy maker) asks *"Which regions have lower maternal-care coverage?"* and receives ranked regions tied to interpretable metrics and hospital-level evidence fields. The interface then surfaces the top at-risk regions, the underlying policy metrics (e.g., delivery_skilled_pct, policy_composite_gap_score), and the supporting hospital evidence fields used for the ranking.

6 Next Steps

Fine-tune detection of hospital data, services and reviews. Add multi-country ingestion and scheduled refresh for longitudinal monitoring.