Global AI Hackathon | Feb 7-8, 2026

in collaboration with MIT Sloan AI Club & MIT Club of Northern California

Bridging Medical Deserts - Building Intelligent Document Parsing Agents for the Virtue Foundation

Sponsored Track by DataBricks

_____

1. Motivation / Goal to Achieve

Motivation: By 2030, the world will face a shortage of over 10 million healthcare workers — not because expertise doesn't exist, but because it is not intelligently coordinated. This is a planetary-scale systems failure and a moonshot opportunity for AI.

Rooted in a real-world challenge faced by the **Virtue Foundation**, this problem is not theoretical: skilled doctors remain disconnected from the hospitals and communities that urgently need them.

This challenge invites you to build an **agentic AI intelligence layer for healthcare** — a system that can reason, decide, and act to connect the right medical expertise with the right hospitals at the right moment.

You are not just building software. You are designing the coordination engine for global healthcare — turning data into action and multiplying human impact through AI.

Ambitious Goal: The goal of this challenge is nothing less than to **reduce the time it takes for patients to receive lifesaving treatment by 100×** using an agentic AI system.

You will build an **Intelligent Document Parsing (IDP) agent** that goes far beyond simple search. This agent will extract and verify medical facility capabilities from messy, unstructured data — and reason over it to understand where care truly exists and where it is missing.

Your system should be able to:

- Identify infrastructure gaps and medical deserts
  Detect incomplete or suspicious claims about hospital capabilities
- Map where critical expertise is available — and where lives are currently at risk due to lack of access

If successful, your AI agent becomes part of a new healthcare intelligence layer: one that helps route patients and doctors faster, guide investment to the right places, and enable organizations like the Virtue Foundation to act with unprecedented speed and precision.

At planetary scale, even small improvements in coordination mean millions of patients treated sooner — and countless lives saved.

2. Core Features (MVP)

1. **Unstructured Feature Extraction:** Process free-form text fields (e.g., procedure, equipment, and capability columns) to identify specific medical data.
2. **Intelligent Synthesis:** Combine unstructured insights with structured facility schemas to provide a comprehensive view of regional capabilities.
3. **Planning system:** Think creatively how you could include a planning system which is easily accessible and could get adopted across experience levels and age groups

3. Stretch Goals (We know that you can do it!)

1. **Citations:** include row-level citations to indicate what data was used to support a claim by an agent.

   Bonus: Provide citations at the *agentic-step level* (e.g., if your agent makes 5 sequential reasoning calls, demonstrate which specific data was used for each step).

   Hint: Consider how experiment tracking tools can "trace" the inputs and outputs of internal agent loops to provide this level of transparency.

2. **Visualize with a Map**: create a map to demonstrate your conclusions visually. [Here](#) is some inspiration.
3. **Real-impact Bonus**: the Databricks team is collaborating with the VF team to ship an agent by June 7th. [Here are the questions being explored](#). Here are some real-world requirements that

unlock impact and we invite you to tackle any of these!

4. Hints and Resources

- **Primary Tech Stack:** Focus on RAG and Agentic workflows, which are top-tier use cases for AI in the medical industry. Here's what the classic base state might be:
    - Agentic orchestrator: langgraph, llama index, crewAI
    - ML lifecycle: MLFlow
    - RAG: Databricks, FAISS, lanceDB
    - Text2SQL: genie
- **Environment:** The challenge is scoped for small amounts of high-impact medical data to ensure compatibility with the Databricks Free Edition.
- **Datasets:** Real-world facility reports and medical notes from a single country provided by the Virtue Foundation.
    - [Virtue Foundation Ghana Dataset](#)
    - [Schema Documentation](#)
    - [Prompts and pydantic models](#) used to create the data

5. Evaluation Criteria

- **Technical Accuracy (35%):** How reliably does the agent handle "Must Have" queries and detect anomalies in facility data?
- **IDP (Intelligent Document Parsing) Innovation (30%):** How well does the solution extract and synthesize information from unstructured "free-form" text?
- **Social Impact (25%):** Does the prototype effectively identify "medical deserts" to aid in resource allocation?
- **User Experience (10%):** Is the interface intuitive for non-technical NGO planners using natural language?

6. Why It Matters

Every data point you extract represents a patient who could receive care sooner. By automating understanding from medical notes — the most critical AI agent use case in healthcare — hackers are creating the intelligence layer that can transform scarcity into coordinated action and bring lifesaving expertise to the world's most underserved regions.