AI-Driven Health Intelligence System

Submitted by Shubhra Karmahe

Executive Summary

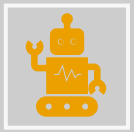
The Problem: Medical data is often trapped in fragmented Excel files with inconsistent formatting and sensitive PII, making it inaccessible for rapid, non-technical analysis.

The Approach: A robust data engineering pipeline coupled with a **Grok-Llama 3** agent to provide secure, natural language clinical recommendations.

The Value: Converts raw "dirty" data into a queryable, HIPAA-compliant knowledge base that supports real-time decision-making.

End-to-End Technical Architecture

The architecture is divided into three distinct layers to ensure data moves from "Raw" to "Insights" securely.



Ingestion Layer: Reads .xlsm files; enforces standard naming conventions and data types.

Transformation Layer Handles missing values, performs PII masking, and executes complex feature engineering.

Intelligence Layer :A **SQL-Agent** that uses Llama 3 to translate user questions into live database queries.

Ingestion Layer

Our approach prioritizes patient privacy and data cleanliness before the AI ever sees the records.

De-identification: The Patient_Number is transformed using deterministic masking (e.g., 12****78). This allows the AI to link clinical data across tables without exposing actual patient IDs.

Missing Data Strategy: * Physical_activity is defaulted to 0 to prevent over-optimistic health reporting.

Pregnancy and other categorical flags are filled with "data not available".

Persistence: All cleaned data is stored in a structured SQLite database with optimized indexes on age, sex, and patient_number.

Transformation Layer

We don't just store data; we add clinical intelligence via a transformation layer:

Metabolic Insights: Automated BMI binning (Underweight, Normal, Overweight, Obese).

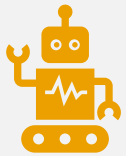
Sex-Aware Logic: Hemoglobin levels are flagged as "Normal" or "Abnormal" using sex-specific biological thresholds ($Male: 14–18$ g/dL vs. $Female: 12–16$ g/dL).

Behavioral Context: Calculation of physical activity metrics (active days vs. missed days) from longitudinal logs.

Intelligence Layer

The intelligence layer uses the **Llama 3 (70B)** model via the **Groq LPU** for sub-second responses.

SQL-Agent Strategy: Instead of training on static data, the model uses **In-Context Learning**. It interprets the SQL schema and generates precise queries to pull live data.



Instruction Tuning: The model is constrained by a "Clinical Persona" system prompt to ensure responses are professional, evidence-based, and devoid of hallucinations.

User Interface & Deployment

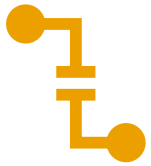
The solution is deployed via a **Streamlit** web application:

Conversational Interface: A chat-based portal where users ask health questions in plain English.

Visualization Engine: The AI automatically triggers bar charts or data tables when it detects "analytic intent."

Auditability: Every interaction and SQL query is logged to agent_ops.log for administrative review.

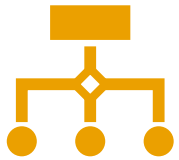
Evaluation & Refinement Framework



To ensure reliability, the system implements a rigorous evaluation loop:

Faithfulness: Does the AI's recommendation match the database values?

Relevancy: Does the AI ignore irrelevant columns to focus on the specific patient query?



Self-Correction: A logic loop where the agent analyzes its own SQL errors and auto-refines the query if the first execution fails.

Challenges & Solutions

Challenge: Protecting patient privacy while maintaining data utility.

Solution: Deterministic masking allows the AI to track a "Patient ID" without knowing the identity.

Challenge: AI hallucinating clinical parameters.

Solution: Strict schema injection and sex-aware feature engineering grounded the AI in hard facts.

Challenge: Latency in traditional LLM APIs.

Solution: Groq's LPU integration reduced response times from 10+ seconds to <1 second.

Tools & Resources Used

- **Core Logic:** Python, Pandas, NumPy.
- **Database:** SQLite, SQLAlchemy.
- **LLM Stack:** LangChain, ChatGroq (Llama 3 70B), Groq API.
- **Front End:** Streamlit.
- **Evaluation: LLM as Judge, Manual Evaluation**

Future Next Steps

01

Longitudinal Prediction: Using Dataset 2 logs to predict future weight/BMI trends using regression.

02

Evaluation Framework:

Use RAGAS for evaluation

03

EHR Integration: Transitioning from static Excel uploads to a live API-based feed for hospital systems.

Thank you