AI ACCELERATOR PROGRAM
PROBLEM SCOPING ASSIGNMENT
By
Houzibe Denis
dhouzibe1@gmail.com

Submitted on 25 February, 2026

PROBLEM SCOPING

1- **Problem:**
   Anomaly Detection for Suspicious Claims in Health Insurance

2- **Problem Statement:**
   Health Insurance Fraud represents billions in annual losses, but the traditional detection systems fail to adapt to evolving fraudulent claims patterns.

3- **Stakeholder Analysis:**
   a. **Primary stakeholder:**
   - Health Maintenance Organizations (HMOs): responsible for claims verification and payment. They are the direct financial risk bearer.
   b. **Secondary stakeholder:**
   - Healthcare Providers (Hospitals, Pharmacies, Labs, Opticals, Dentals, etc): They submit claims.
   - Policyholders/patients: They are the beneficiaries and are impacted by rising premiums and potential denial of care if funds are mismanaged.
   c. **Tertiary Stakeholder (the regulator & industry body):**
   - The National Health Insurance Authority (NHIA) sets the rules and standards. Also, they define the regulatory framework and reporting requirements.

4- **Current Solutions & Limitations**
   a. **Current Solution**
      ● Manual Auditing: Claims are reviewed by human auditors after payment or on a random sample basis. This is a slow, expensive, and highly subjective process.
      ● Simple Rule-Based Systems: Basic software flags claims that exceed a certain amount, are for a specific high-cost procedure, or come from a provider with a history of issues. For example, "Flag any single claim over ₦500,000."

b. **Limitations of Current Solutions:**

- Reactive, Not Proactive: Most detection happens *after* the money has been paid, making recovery difficult and costly. The goal is prevention, not just detection.
- Inability to Detect Complex Schemes: Simple rules fail to catch sophisticated fraud, such as:
  - Billing for non-existent beneficiaries ("Ghost Patients").
  - Upcoding: Billing for a more expensive service than was actually performed.
  - Unbundling: Billing for the individual components of a single procedure to get a higher total payment.
  - Collusion**:** Patients and providers working together to file false claims.

5- **Proposed AI Opportunity:**

I proposed to develop and deploy an **AI-Powered Anomaly Detection Platform for Suspicious Claims** in Health Insurance. This system will learn normal patterns of behavior and automatically flag outliers for investigation, making the process proactive, intelligent, and efficient.

The platform will consist of two key AI/ML components:
a. **Supervised Learning Model for Known Fraud:** This model will be trained on historical claims data that have been previously labeled as "fraudulent" or "legitimate." It will learn the complex signatures of known fraud types (upcoding, phantom billing) and automatically flag new claims that closely resemble these past patterns.
b. **Unsupervised Learning Model for Anomaly Detection:** This is the core innovation. Instead of looking for known fraud, this model will learn what "normal" looks like for a specific provider, patient demographic, or procedure. It will then flag statistical anomalies that deviate significantly from the norm.

6- **Success Metrics**

The success of the AI solution will be measured by its impact on efficiency, financial savings, and deterrence.

Metric Category	Metric	Baseline (Estimated)	Target (After 1 Year)	Measurement Method
EFFICIENCY	Claims Audit Turnaround Time	15 days per batch	5 days per batch (67% reduction)	Time tracking within the claims management system.
	Investigator Cases per Day	5 random claims	25 high-risk claims (5x increase)	User activity logs in the new dashboard.
FINANCIAL IMPACT	Value of Fraud/Error Detected	₦50 million identified/year	₦150 million identified/year	Value of claims flagged and confirmed as fraudulent/erroneous by investigators using the tool.
	Prevention of Improper Payments	N/A (reactive only)	₦75 million prevented	Value of high-risk claims intercepted *before* payment.
ACCURACY	Precision Rate (Hit Rate)	< 10% (from random audits)	> 40% (of flagged claims are fraudulent/erroneous)	Ratio of confirmed fraudulent flags to total flags investigated.
	False Positive Rate	Very High (from rules)	< 5%	Percentage of high-risk flags that are investigated and found to be legitimate.

7- **Potential Data Sources**

The primary data source is partnering HMO. Historical Claims Data is the most critical dataset