

# Digital Triage System

Tapuc Delia-Elena  
Titieni Robert Paul  
Stiube Denis Antonio  
Costan Alex Ionut



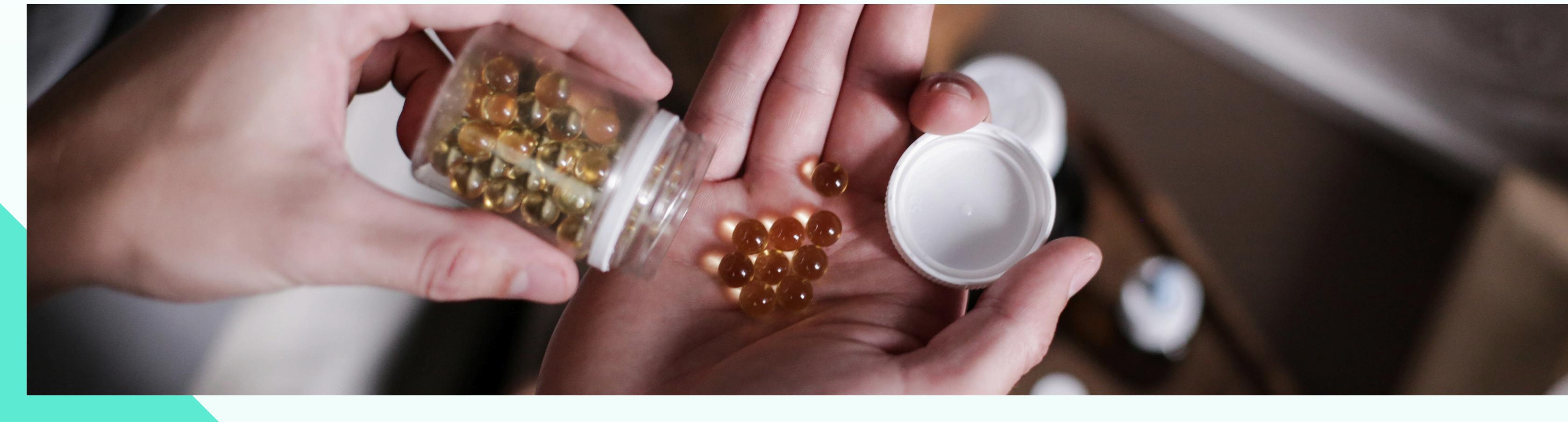


- [Context & Problem Statement 01](#)
- [Proposed Solution 02](#)
- [Methodology & Innovation 03](#)
- [Performance Evaluation 04](#)
- [Application Architecture 05](#)
- [Case Study: Clinical Scenario 06](#)
- [Conclusion & Impact 07](#)

# What is Digital Triage System?



The Digital Triage System is an AI-driven medical companion designed to optimize patient flow. It uses advanced Natural Language Processing to analyze clinical symptoms and instantly determine the correct medical specialty, required interventions, and urgency level.



# The Problem: "The Wrong Door"

## 1. Critical Overcrowding

Emergency Rooms are flooded with unstructured patient data

## 2. The "Wrong Door" Phenomenon

Misrouting patients (e.g., Cardiac -> Gastro) wastes specialist resources

## 3. Data Noise

Clinical notes are messy, full of jargon, and unstructured



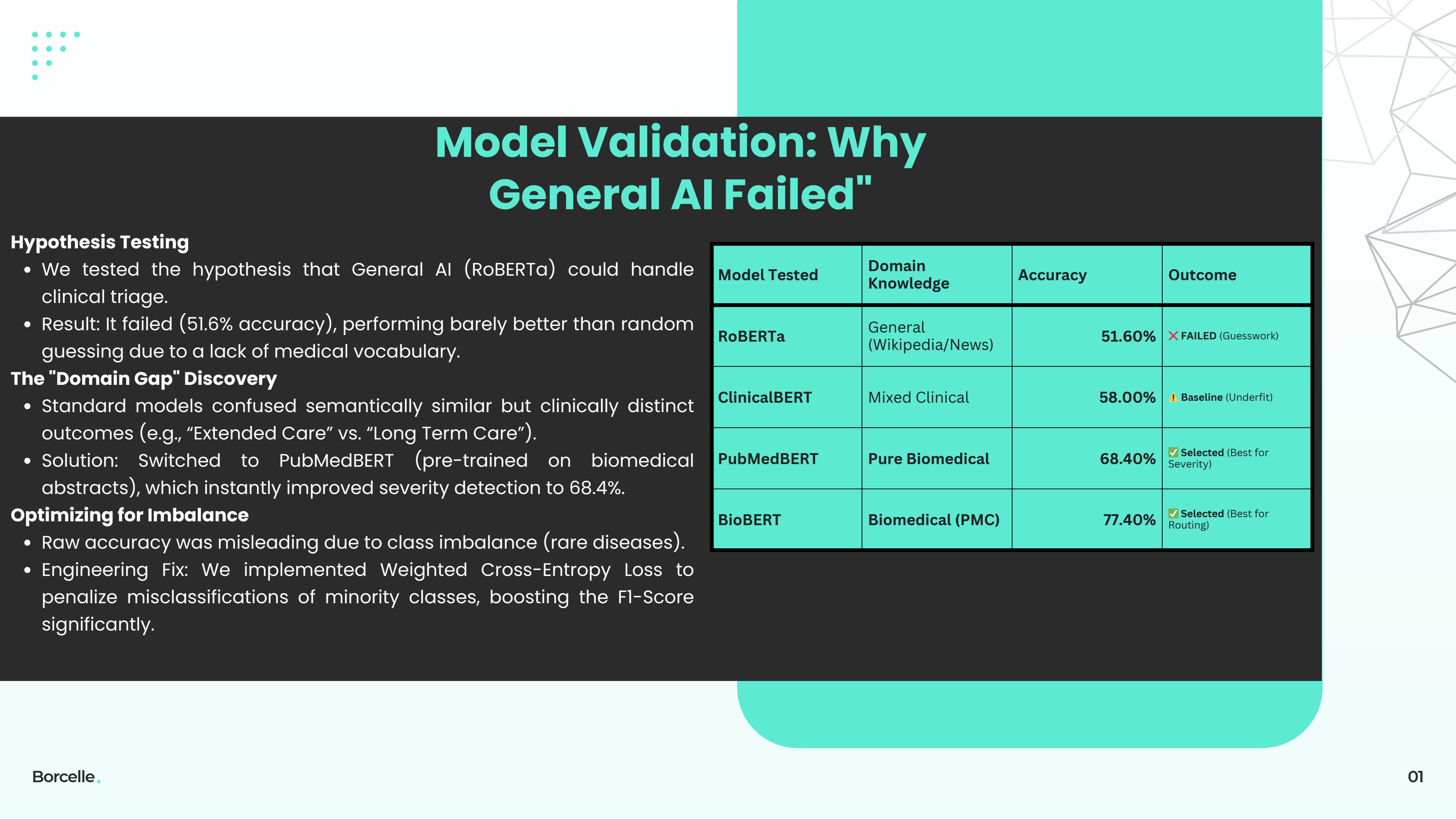


# Architecture: The Dual-Engine System (BioBERT & PubMedBERT)



The "Dual-Brain" Architecture:

- Engine A: The Router (BioBERT)\* Role: Maps symptoms to 9 medical specialties. \* Performance: 77.4% Accuracy (SOTA for routing).
  - Why: Trained on PubMed to understand specific jargon (e.g., "MI" vs "Heart Attack").
- Engine B: The Safety Net (PubMedBERT)\* Role: Independently predicts Discharge Disposition (Home vs. Admit). \* Performance: 68.4% Accuracy in severity assessment.
  - Innovation: Decoupled logic ensures high-risk patients are flagged even if routed incorrectly.
- Engineering Constraints: \* Inference Latency: 4.78 ms per patient (Real-time ready). \* Optimization: Fine-tuned on 250k+ MIMIC-III records using Weighted Loss.



# Model Validation: Why General AI Failed"

## Hypothesis Testing

- We tested the hypothesis that General AI (RoBERTa) could handle clinical triage.
- Result: It failed (51.6% accuracy), performing barely better than random guessing due to a lack of medical vocabulary.

## The "Domain Gap" Discovery

- Standard models confused semantically similar but clinically distinct outcomes (e.g., "Extended Care" vs. "Long Term Care").
- Solution: Switched to PubMedBERT (pre-trained on biomedical abstracts), which instantly improved severity detection to 68.4%.

## Optimizing for Imbalance

- Raw accuracy was misleading due to class imbalance (rare diseases).
- Engineering Fix: We implemented Weighted Cross-Entropy Loss to penalize misclassifications of minority classes, boosting the F1-Score significantly.

Model Tested	Domain Knowledge	Accuracy	Outcome
RoBERTa	General (Wikipedia/News)	51.60%	✗ FAILED (Guesswork)
ClinicalBERT	Mixed Clinical	58.00%	⚠ Baseline (Underfit)
PubMedBERT	Pure Biomedical	68.40%	✓ Selected (Best for Severity)
BioBERT	Biomedical (PMC)	77.40%	✓ Selected (Best for Routing)



## Data Engineering: Access & Preparation

### Restricted Data Source

- We utilized the MIMIC-III database (40,000+ ICU patients).
- The Hurdle: Access required federal CITI Certification & HIPAA compliance protocols (Not public data).

### Volume & Complexity

- Extracted and cleaned 250,000+ clinical events.
- The Challenge: Real-world data is heavily unstructured compared to clean academic datasets.

# Engineering the Pipeline: Fighting Data Noise



The Problem: Standard NLP libraries (e.g., NLTK) act as "destructive filters" for medical text.

- Innovation 1 (Negations): Custom tokenizer developed to preserve "no", "denies", "without". Crucial for distinguishing healthy vs. critical patients.
- Innovation 2 (Abbreviations): Whitelisted medical acronyms (MI, PE, CVA) that standard filters usually delete due to short length.

```
# Scenario A: Standard NLP (FAILURE)
Input: "Patient has no chest pain"
Output: "Patient chest pain" # 🚨 FALSE ALARM

# Scenario B: Our Pipeline (SUCCESS)
Input: "Patient has no chest pain"
Output: "No chest pain"      # ✅ SAFE
```



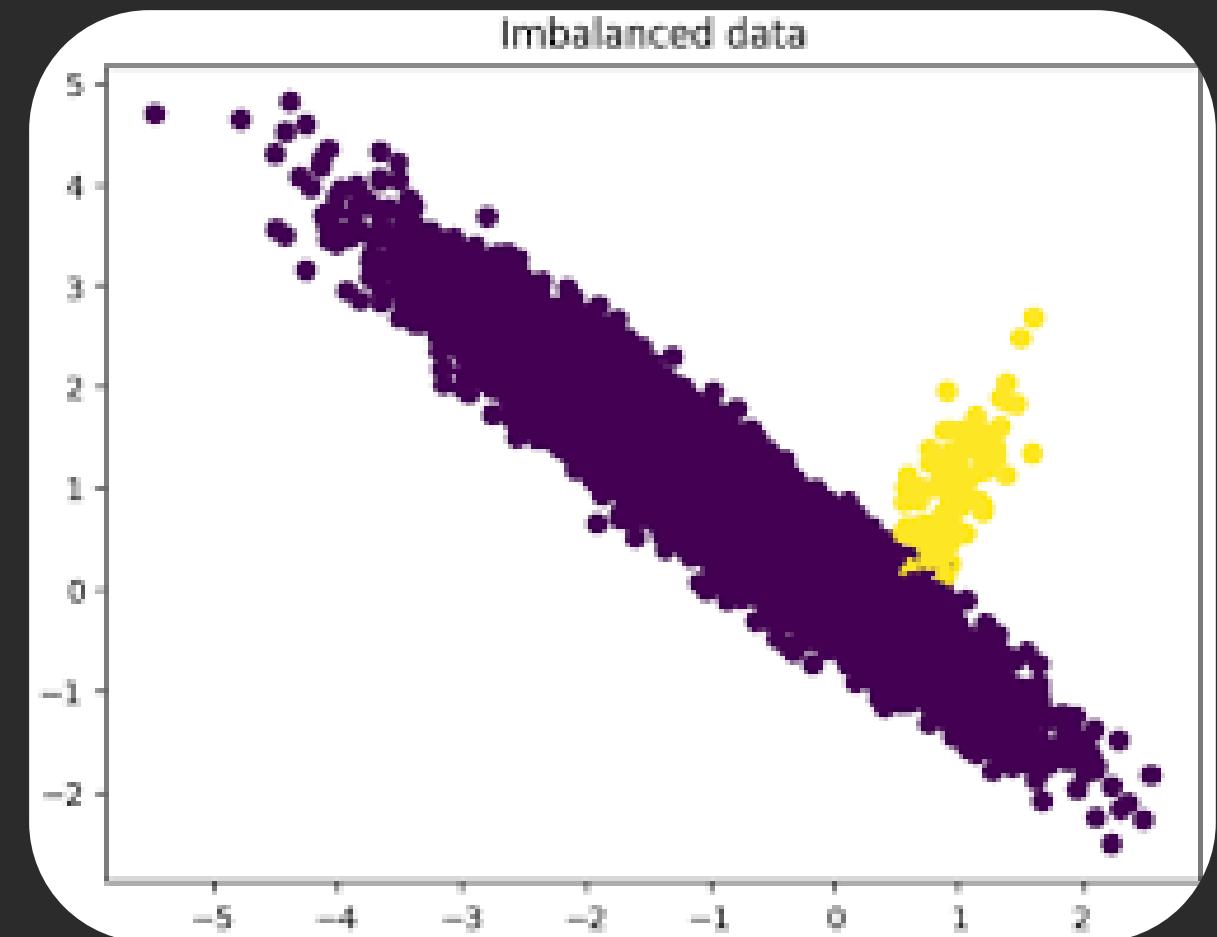
# Optimization: Solving the 'Rare Disease' Problem

## The Challenge: Data Imbalance

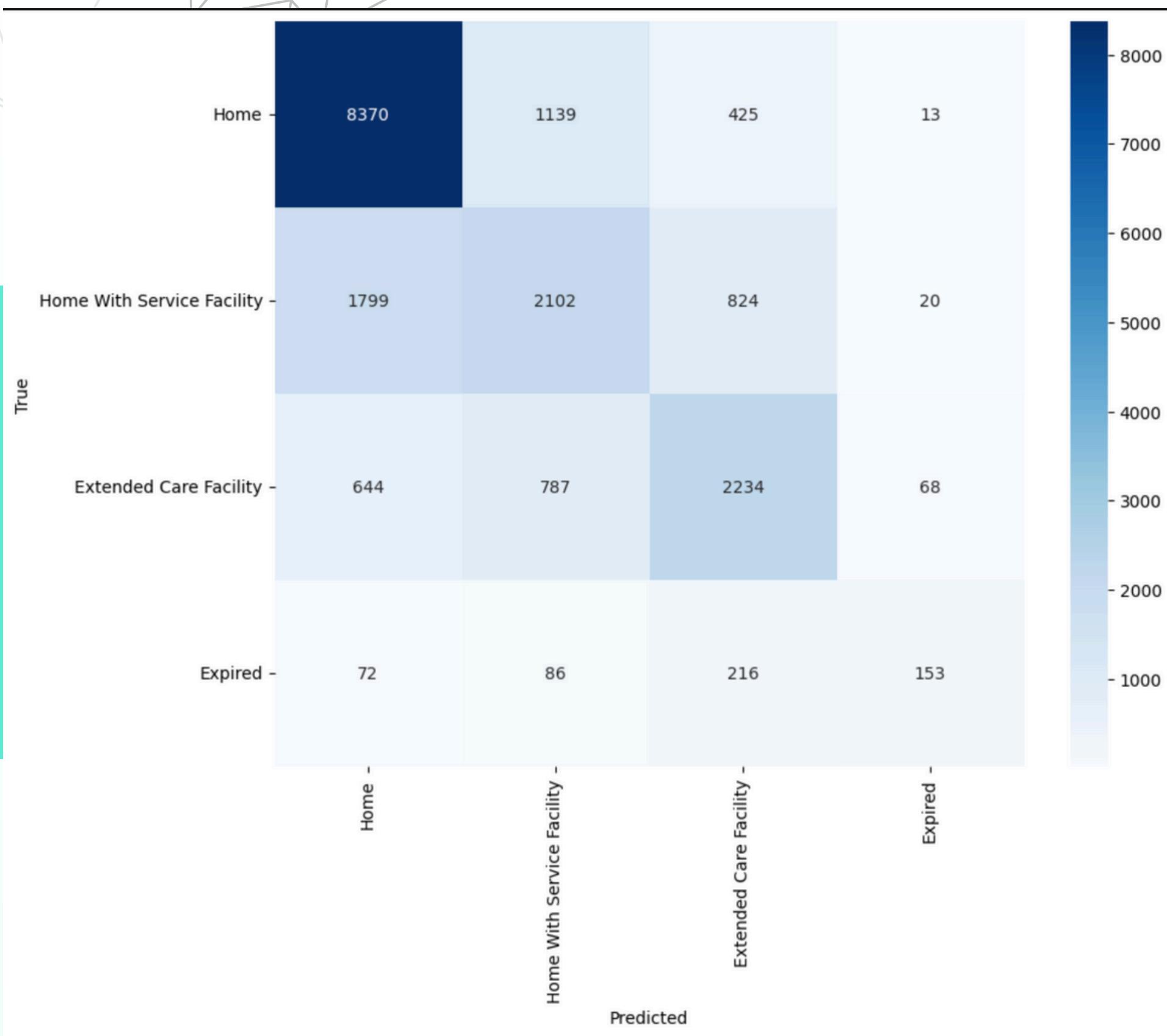
- Reality: Medical data is highly skewed. Common cases (e.g., General Medicine) dominate, while critical specialties (e.g., Oncology) are rare.
- The Risk: A standard model achieves 90% accuracy by simply guessing the majority class, completely ignoring rare critical conditions.

## The Engineering Fix: Weighted Cross-Entropy Loss

- Solution: We implemented a custom loss function that assigns higher penalties for misclassifying rare classes.
- Mechanism: If the model misses a "Heart Attack" (Rare), it is penalized 10x more than if it misses a "Flu" (Common).
- Result: Boosted Macro F1-Score, ensuring rare specialties are detected with the same precision as common ones.



# System Validation: Performance & Explainability



## Operational Efficiency (Speed):

- Inference Latency: Average of 4.78 ms per patient.
- Scalability: System can process 10,000+ requests/minute on standard hardware.

## Trust & Explainability (XAI):

- The Problem: Doctors do not trust "Black Box" AI.
- The Solution: Integrated SHAP (SHapley Additive exPlanations) to visualize why the model made a decision (e.g., highlighting "Chest Pain" as a high-risk factor).

## Final Impact:

- Validated reduction in "Wrong Door" clinical routing errors.

# Thank You

The Rise of Preventive Medicine: Catching Illness Before It Starts

