🚀Getting Started Guide

Clinical Intelligence Platform – What to Build, How to Build It, and How to Start

This document is a **starter manual**. You should be able to read this and know: - what models you are building - what data they use - how to create training data - what to implement first (without being a doctor)

---

1. The Core Goal (Read This First)

> **Goal:** Build a system that detects **slow patient deterioration over time** and warns humans early.

The system: - ❌does NOT diagnose - ❌does NOT prescribe treatment - ✅only says: *"Risk is increasing over time — please review"*

Everything else in this document exists to serve that goal.

🛖2. High-Level System Pipeline (Mental Model)

```
Raw visits (rows)
   ↓
Time-indexed patient timeline
   ↓
Trend features (math)
   ↓
Risk scoring (ML)
   ↓
Explanation (rules)
   ↓
UI graphs + alerts
   ↓
Human decision
```

Keep this pipeline in mind while building. Each box is a **separate responsibility**.

---

3. What Data You Start With

Raw input data (example)

```
Patient: Raj
Age: 52

Date          Blood Sugar    BP         Notes
-----------------------------------------
2023-01-01      110         130/85     "Borderline"
2023-07-01      118         135/88     "Monitor"
2024-01-01      126         140/92     "High"
2024-07-01      142         150/95     "Medication started"
```

Important: - Each row = one visit - This data is **not** directly used by ML - It must be converted into a **timeline** first

---

✈️4. Patient Timeline (Your First Backend Task)

Convert visits into time-series format:

```json
{
  "patient_id": "raj_001",
  "metrics": {
    "blood_sugar": [
      {"date": "2023-01-01", "value": 110},
      {"date": "2023-07-01", "value": 118},
      {"date": "2024-01-01", "value": 126},
      {"date": "2024-07-01", "value": 142}
    ]
  }
}
```

This timeline is used by: - UI graphs - Trend detection - Risk scoring

---

5. Model 1 – Trend Detector (NO ML)

Purpose

Turn raw timelines into **trend features**.

Input

Time-series values.

Output (example)

```json
{
  "metric": "blood_sugar",
  "trend_direction": "UP",
  "percent_change": 29,
  "duration_months": 18,
  "persistence": true
}
```

How to implement

- Percent change
- Linear slope
- Count of consecutive increases
- Time duration

☸️This step is deterministic math. No training.

⚓6. Model 2 – Risk Scoring Model (MAIN ML)

Purpose

Answer:

> *Given these trends, how worried should we be?*

Inputs (IMPORTANT)

The model never sees raw values like `110` or `142`.

It only sees **trend features**:

```json
{
  "age": 52,
  "sugar_percent_change": 29,
  "sugar_trend_up": 1,
  "bp_trend_up": 1,
  "trend_duration_months": 18,
```

```
  "medication_delay": 1
}
```

Outputs

```
{
  "risk_score": 0.78,
  "risk_level": "HIGH",
  "confidence": 0.82
}
```

Recommended models (in order)

1. Logistic Regression (start here)
2. Gradient Boosted Trees (XGBoost / LightGBM)
3. Small Neural Network (only later)

☸️Do NOT use transformers or LSTMs early.

---

7. How Training Data Actually Looks

Each training row represents:

> **"What was the patient's trend state at time T, and did something bad happen later?"**

Example training row

```
{
  "age": 55,
  "sugar_percent_change": 30,
  "bp_trend_up": 1,
  "trend_duration_months": 18,
  "medication_delay": 1,
  "label": 1
}
```

- `label = 1` → deterioration happened later
- `label = 0` → no deterioration

Labels are based on **future events**, not diagnoses.

🛕8. How to Create Training Data (Realistic Plan)

Phase 1: Synthetic Data (YOU MUST DO THIS)

- Manually generate realistic feature rows
- Define simple rules for labels
- Goal: test pipeline, not accuracy

☸️Every real healthcare project does this first.

Phase 2: Public Medical Datasets

Once pipeline works: - Use public ICU datasets (e.g. MIMIC, eICU) - Extract timelines - Compute trend features - Label based on outcomes (ICU transfer, death, escalation)

You are learning **patterns**, not deploying these models.

Phase 3: Human Feedback (Future)

In real deployment: - Clinicians acknowledge / dismiss alerts - Feedback becomes labels - Model thresholds are calibrated

No uncontrolled auto-learning.

🛥️9. Model 3 – Explanation Generator (NO ML)

Purpose

Explain *why* risk increased.

Input

Top contributing trend features.

Output

```
Risk increased because:
- Blood sugar rose 29% over 18 months
```

```
- Blood pressure increased across 4 visits
- Medication was delayed
```

Rules

- No diagnosis
- No treatment advice
- Facts + time only

🏫10. What the UI Uses

The UI consumes: - Patient timelines → line graphs - Trend features → arrows / highlights - Risk scores → risk-over-time graph - Explanations → callouts

The UI never plots: - LLM text - Predictions - Diagnoses

---

11. What to Build FIRST (Very Important)

Order of implementation

1. Patient timeline data model
2. Trend detector
3. UI graphs from timelines
4. Synthetic training data
5. Risk scoring model
6. Explanation generator

Do NOT jump to ML first.

Final One-Line Summary

> **This system converts scattered visits into a computable story, detects when the story worsens, and explains why — safely.**

Use this document as your implementation guide. Build one box at a time.