Guardian Alerts (Sprint 2) – Monitoring Pipeline Report

1) Executive summary

I built an end-to-end notebook that flags patient risk using two complementary signals:

1. a time-series anomaly model (LSTM autoencoder; IsolationForest fallback) that looks for unusual behaviour over recent days, and
2. a behavioural-anomaly classifier (Random Forest or MLP) trained on engineered features (deltas, rolling means, z-scores).

On top of the model scores, I added a clinically anchored vitals overlay (SpO₂, temperature, blood pressure, activity, meals skipped). The final risk\_level is the maximum of the model view and the vitals view, and every Medium/High row includes a clear reason.

The pipeline writes artifacts/alerts.csv with:  
user\_id, timestamp, anom\_score, clf\_prob, risk\_level, reason.

2) Data ingestion & normalization

* Auto-discovery: The notebook finds the dataset in the same folder (prefers New AI spreadsheet - Sheet1.csv).
* Schema mapping (→ canonical names):
  + patientId → user\_id
  + observationStart → timestamp (uses observationEnd if Start is absent)
  + Behaviour: stepsTaken → steps, calorieIntake → calorie\_intake, sleepHours → sleep\_hours,  
    waterIntakeMl → water\_intake (mL→L), bathroomVisits → bathroom\_visits
  + Vitals/context: heartRate → heart\_rate, spo2 → spo2, temperature → temperature,  
    bloodPressure "120/80" → bp\_sys, bp\_dia, mealsSkipped → meals\_skipped, exerciseMinutes → exercise\_minutes
* Time handling: Parse timestamps and sort by user\_id, timestamp.
* Missing values: Per-user interpolate → back/forward fill → remaining NaNs to 0.
* Scaling: StandardScaler fit on all model features; saved as artifacts/scaler.pkl.

3) Models & features

3.1 Sequence anomaly (LSTM AE; IF fallback)

* Architecture: LSTM autoencoder (hidden=32, latent=16), teacher-forced reconstruction.
* Window: SEQ\_LEN = 14 days.
* Training: 30 epochs, Adam(1e-3).
* Score: Mean-squared reconstruction error per window → aligned to timestamps.
* Calibration: Compute the dataset’s 80th and 95th percentiles of the raw error (err\_p80, err\_p95) to avoid “everything = High”.

If PyTorch isn’t available, IsolationForest (contamination 0.05) is used and we take -score\_samples as an error-like measure.

3.2 Behavioural anomaly classifier (RF/MLP)

* Models: rf (default) or mlp.
* Engineered features (for each of 12 inputs: 5 behaviour + 7 vitals):  
  value, delta, 7-day rolling mean, 7-day rolling z-score.
* Labels: If no ground truth exists, create weak labels: top 5% by reconstruction error = anomalous.
* Output: clf\_prob (0–1). For interpretability, we use 0.65 / 0.85 as Medium / High hints.

4) Alert logic (how risk\_level is decided)

This is the exact mapping the notebook implements.

4.1 Signals computed first

* anom\_score (0–1): min–max normalization of the reconstruction error (for visibility and plots).
* recon\_error: the unnormalized error used to compare against percentiles err\_p80 / err\_p95.
* clf\_prob (0–1): classifier probability (or scaled decision value).
* Vitals snapshot: spo2, temperature (°C), bp\_sys/bp\_dia, exercise\_minutes/day, meals\_skipped/day.

4.2 Model risk (based on anomaly + classifier)

High if recon\_error ≥ err\_p95 OR clf\_prob ≥ 0.85

Medium if recon\_error ≥ err\_p80 OR clf\_prob ≥ 0.65 (and not High)

Low otherwise

4.3 Vitals risk (direction-aware clinical thresholds)

* SpO₂: Low ≥95% | Medium 90–94% | High <90%
* Temperature (°C): Low <38.0 | Medium 38.0–39.3 | High ≥39.4
* Blood pressure (mmHg): Low <130/<80 | Medium 130–139 or 80–89 | High ≥140 or ≥90 (escalate internally if ≥180/120)
* Exercise minutes/day: Low ≥20 | Medium 10–19 | High <10
* Meals skipped/day: Low 0–1 | Medium 2 | High ≥3

4.4 Final risk & reasons

risk\_level = max(model\_risk, vital\_risk) # High > Medium > Low

* reason (string): for Medium/High, we list every trigger that fired, e.g.
  + “Strong sequence anomaly (≥95th percentile)”
  + “Classifier: strong behavioural anomaly (≥0.85)”
  + “SpO₂ 89% (<90)”, “High fever 39.5 °C (≥39.4)”, “Stage 2 HTN 165/102”, “Very low activity (6 min)”, “Meals skipped: 3”
* For Low, we leave reason blank (as requested) to keep the CSV clean.

5) Outputs & visualizations

* Primary CSV: artifacts/alerts.csv — entire dataset, columns:  
  user\_id, timestamp, anom\_score, clf\_prob, risk\_level, reason
* Saved models: lstm.pt (if LSTM) or iforest.pkl, plus clf.pkl and scaler.pkl.
* Thresholds meta: thresholds.json (min/max error, p80, p95).
* Notebook visuals:
  1. Anomaly score distribution with p80/p95 markers
  2. anom\_score vs clf\_prob (scatter) colored by final risk
  3. Risk counts (bar)
  4. Example patient timeline (key features + risk overlay)

6) Values and sources (provenance)

* Exercise target: WHO adults’ guideline 150–300 min/week moderate (≈21–43 min/day).
* Blood pressure categories: American Heart Association (Normal/Elevated/Stage 1/Stage 2; crisis ≥180/120).
* Temperature: NHS fever in adults ≥38.0 °C; we treat ≥39.4 °C as high fever.
* SpO₂: Typical “normal” ~95–100%; <90% concerning at rest (Cleveland Clinic style guidance).
* Meals skipped: heuristic (no formal standard).

(In the code, these are embedded as comments near the thresholds for auditability.)

7) Configuration knobs

* Engine: ENGINE = "lstm" | "iforest"; Classifier: CLF = "rf" | "mlp".
* Window length: SEQ\_LEN (default 14).
* Sensitivity: tweak err\_p80/err\_p95 or classifier cutoffs (0.65/0.85).
* Vitals thresholds: adjust SpO₂/Temp/BP/activity/meals to match clinical guidance or site policy.

8) Limitations & next steps

* Not a diagnostic tool: Alerts are for monitoring/triage.
* Weak labels: Until we have ground truth, the classifier learns from anomaly tails.
* Context variance: Vitals can depend on altitude, chronic conditions, or orders; future work: per-patient baselines & clinician-tuned thresholds.
* Enrichment: Add NLP on nursingNote, behaviourTags, emotionTags to improve reasons.

9) TL;DR

* We map anomalies → Low/Medium/High via calibrated sequence error and a behavioural classifier, then apply clear clinical cutoffs for vitals.
* Output is one final risk\_level plus a readable reason explaining what fired.
* Thresholds are transparent, tunable, and sourced (WHO/AHA/NHS/Cleveland Clinic where applicable).