**FX Anomaly Model — Two Tracks (Databricks + Azure Blob)**

We operate the autoencoder‑based FX anomaly model in **two tracks**:

* **Track A — Existing clients (scheduled retraining)**
* **Track B — New client onboarding**

**Track A — Existing Clients (Scheduled Retraining)**

**Inputs**

* New parquet files under: <client\_name>/retraining/<ingest\_timestamp>/\*.parquet
* Holdout for acceptance under: <client\_name>/test\_set/<version\_or\_asof>/\*.parquet
* Bulk history under: <client\_name>/History/<ingest\_timestamp>/\*.parquet (Available Already)

**Assumptions (Track A)**

* Upstream drops data under retraining/ every ~60 days with a valid <ingest\_timestamp>; paths are append‑only.
* Schema remains compatible with our current **featureset v1** (additive columns allowed; no breaking renames/types).
* Effective rows after de‑dup/filtering ≥ min\_rows\_per\_window (client‑configurable; default 50k).
* A stable, representative **test\_set/** exists for the client (no leakage; labels or silver‑labels are maintained).
* Credentials, cluster image, and MLflow registry access are available at run time (reproducible runtime).
* **Automation baseline:** All automated **model creation and retraining** use the **last best known model** (current MLflow *Production* version) for architecture, hyperparameters, and preprocessing choices.

**Automation & human intervention (Track A)**

* **Automated:** polling/event trigger → validation → training → evaluation → registry → canary gate.
* **Human needed when:**
  + FAILED\_SANITY/FAILED\_VOLUME to coordinate data fixes with provider.
  + Two consecutive REJECTED\_NO\_IMPROVEMENT runs → review architecture/HPs with DS lead.
  + Canary shows material precision drop or alert volume surge → business sign‑off before proceeding.

**How a run starts**

* **Event‑driven**: upstream notifies (message/email) that new data is available.
* **Time‑driven**: every **60 days** since last successful retrain; if no event, poll for new data up to **4 attempts** every **6 hours**. If still nothing → SKIPPED\_NO\_DATA and alert.

**What the job does**

1. **Intake & sanity**: read → merge → de‑duplicate → schema/type checks → basic completeness checks; log file list, sizes, hashes.
2. **Training window & features**: use a **rolling 4‑year window** ending at the newest txn\_ts; fit fresh preprocessors (scalers, encoders with <UNK>), date features, clipping/log as configured.
3. **Model training**: train an autoencoder using the **last best known model configuration** (architecture & HPs from MLflow *Production*). Early stopping on validation; choose threshold to optimize **F1** on validation.
4. **Evaluate & accept**: score the **fixed holdout**; if **ΔF1 ≥ 0.01**, accept; else reject.
5. **Register & deploy**: register artifacts in MLflow; MINOR version bump; Staging → Production (canary if required).: register artifacts in MLflow; MINOR version bump; Staging → Production (canary if required).

**Exit points**

* FAILED\_SANITY → stop before training; notify provider with failed checks.
* FAILED\_VOLUME → stop before training; notify provider.
* FAILED\_TRAINING → log, retry once on clean cluster; alert if persistent.
* REJECTED\_NO\_IMPROVEMENT (ΔF1 < 0.01) → no deployment; production unchanged.
* SKIPPED\_NO\_DATA → no run beyond logging/alerting.

**Outputs**

* MLflow run with metrics (F1, precision/recall at threshold, PR‑AUC), confusion matrices, validation stats, and a signed training manifest.

**Track B — New Client Onboarding**

**Inputs**

* Bulk history under: <client\_name>/History/<ingest\_timestamp>/\*.parquet
* Holdout (if supplied) under: <client\_name>/test\_set/<version\_or\_asof>/\*.parquet

**Assumptions (Track B)**

* Provider can supply ≥ **12 months** or **≥25k rows** of history with mandatory columns (IDs/time, core categorical and numeric features). (
* Column names and dtypes either match **featureset v1** or a **data dictionary**/mapping table is provided.
* Units and conventions are standard.
* Either a labeled **test\_set/** exists, or the client accepts a rules‑based proxy and a canary‑first acceptance.
* PII handling and access controls are confirmed before any training artifacts are registered.
* **Automation baseline:** The **first‑cut automated model** will use **one of the last best models’ parameters** (architecture & hyperparameters) as the starting configuration.

**What can be automated (green path)**

* Compatibility scoring against featureset v1 (column presence, dtypes, simple synonyms) and Great‑Expectations checks.
* Auto‑mapping using a provided data dictionary + common synonyms; standardization of missing markers and timestamp parsing.
* First‑cut training with the existing pipeline/HPs; evaluation on test\_set/ if available.
* Auto‑registration to **Staging** and a **capped canary** (e.g., small % of traffic, fixed duration) with automatic rollback on guardrail breach.

**Human‑in‑the‑loop checkpoints (why & when)**

* **Ambiguous or missing mappings** (multiple plausible column matches, non‑standard units) → requires SME + DS to confirm.
* **INCOMPATIBLE\_FEATURES** (mandatory fields absent, unexpected category semantics like Buy/Sell direction) → define mapping or a new featureset before training.
* **Low volume or poor quality** (fails FAILED\_VOLUME/sanity) → negotiate data refresh or scope adjustments.
* **No labeled holdout** → agree on proxy evaluation method and canary acceptance thresholds with stakeholders.
* **Canary review** → confirm precision/alert budget is acceptable before full Production.

**What the job + analyst do**

1. **Intake & sanity**: read, merge, de‑duplicate, and run checks:
   * Required columns present with correct dtypes; timestamps parseable to UTC
   * 5% duplicates after de‑dup → fail
   * Effective rows ≥ min\_rows\_per\_window (post sanity & filters); otherwise fail
   * Standardize missing markers to null
2. **EDA & mapping (human‑in‑the‑loop)**: quick profiling (distributions, missingness, categorical coverage), domain checks (e.g., BuyAmount vs SellAmount by currency), and mapping to **featureset v1**. If not compatible, define a new **featureset v2**.
3. **First‑cut model**: where compatible, train with the **same pipeline and HPs** used for existing clients.
4. **Evaluate**: use labeled sample if available, or a rules‑based backtest to estimate precision/recall. Acceptance uses the **same ΔF1 ≥ 0.01** rule if a test set exists; otherwise, proceed with a controlled canary and manual review.
5. **Register & deploy**: register artifacts; Staging → limited canary; expand to Production once metrics are confirmed.

**Exit points**

* FAILED\_SANITY → stop; send gap report.
* FAILED\_VOLUME → stop; request additional data (e.g., ≥12 months or ≥25k rows) .
* INCOMPATIBLE\_FEATURES → pause; require mapping updates or define a new featureset before training.
* REJECTED\_NO\_IMPROVEMENT (if test set available and ΔF1 < 0.01) → no deployment.