

Azure text classification: governed decisions

Baseline: labeled data + TF-IDF / linear model; promote a transformer only if its measured improvement justifies cost.

LEARNING PLANE

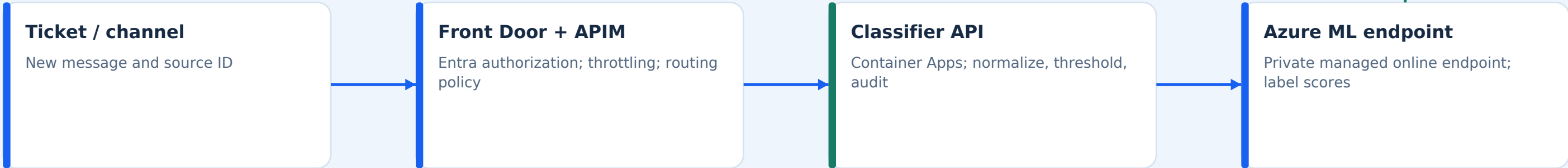
Separate model development from production inference



promote

DECISION PLANE

Single synchronous classification call; queue-based bulk intake is an optional adapter



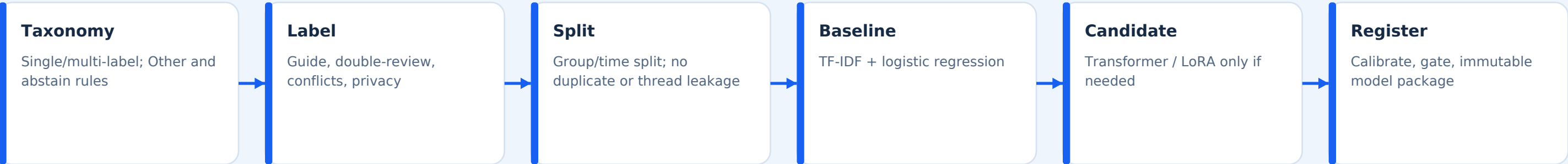
Closed-loop operation
API returns label(s), calibrated score, model/taxonomy versions and review flag. Human corrections flow back to the curated label set.

From labels to a deployable model

Use a simple model as a controlled benchmark. Keep the test set untouched until final promotion.

A. CURATE AND TRAIN

Azure ML pipeline components create reproducible artifacts with data, code and environment versions.



B. SCORE AND LEARN

The API enforces taxonomy and decision policy; delayed human labels enable real-world performance measurement.



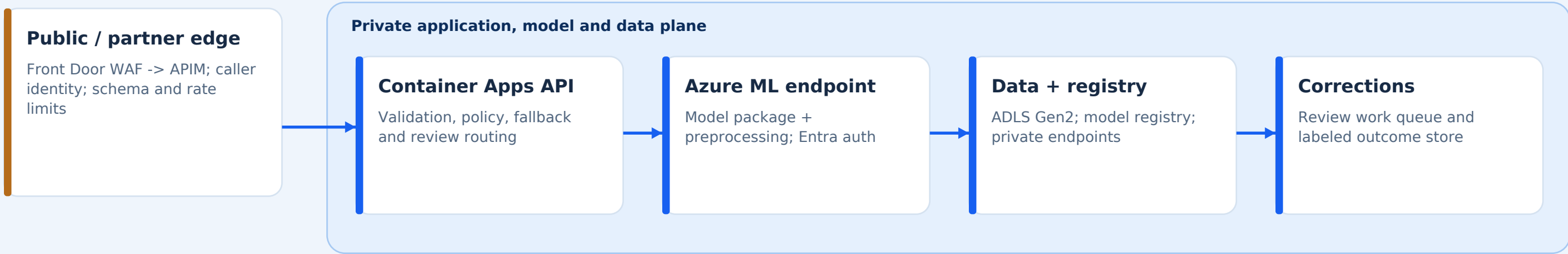
Prediction contract request_id, taxonomy_version, model_version, labels[{name, score, threshold}], decision, review_reason, timestamp; do not log raw text by default.

Secure inference and monitored feedback

Class drift is not the same as accuracy drift. Join reviewed outcomes to predictions before claiming production quality.

PRODUCTION BOUNDARY

Keep training assets and inference endpoints isolated by environment and least privilege.



Model governance

- Lineage**
Version taxonomy, label guide, dataset split, code, environment and model.
- Promotion**
Compare against baseline by class; require owner review before release.
- Rollback**
Blue/green or traffic-split endpoint deployment; preserve prior model.

Production signals

- Quality**
Per-class precision/recall/F1, confusion and calibration on reviewed labels.
- Drift**
Text length, language, embedding/topic shift, class mix and review volume.
- Service**
p95 latency, errors, throughput, cost, PII and fairness slices where relevant.

Implementation decisions and release gates

A label is a business decision. Document its meaning and the cost of each type of classification error.

Workstream	Build actions	Acceptance evidence
Phase 1: define	Agree taxonomy, routing actions, multi-label rules, minimum confidence and review ownership.	Label guide; edge-case decisions; privacy review.
Phase 2: data	Collect representative history; adjudicate labels; group/time split with untouched test set.	Leakage check; class coverage and inter-annotator agreement.
Phase 3: baseline	Train TF-IDF + linear classifier in Azure ML; package preprocessing with model.	Per-class metrics, calibrated thresholds, error analysis.
Phase 4: compare	Only then evaluate embeddings or transformer/LoRA against baseline on validation.	Meaningful gain on hard cases within latency/cost budget.
Phase 5: operate	Deploy endpoint, review queue, corrections, monitoring and staged rollback via IaC/CI.	Canary acceptance; ground-truth join; retrain/rollback runbooks.

Azure ML is the primary flexible model path

Custom text classification is a managed alternative

LoRA is optional, not a baseline dependency

Microsoft documentation used for service choices

[1] MLOps v2

<https://learn.microsoft.com/azure/architecture/ai-ml/guide/machine-learning-operations-v2>

[3] Model monitoring

<https://learn.microsoft.com/azure/machine-learning/concept-model-monitoring>

[2] Secure endpoints

<https://learn.microsoft.com/azure/machine-learning/concept-secure-online-endpoint>

[4] Custom classification

<https://learn.microsoft.com/azure/ai-services/language-service/custom-text-classification/overview>