Retention Agent -Documentation

Overview

The **Retention Agent** helps Customer Success Managers (CSMs) proactively identify customers at risk of churn by analyzing product usage patterns over time. Its purpose is to reduce manual monitoring of usage dashboards while ensuring that **all decisions remain under human control**.

The agent does **not** contact customers directly. It produces **early-warning insights and recommendations** that help humans decide when and how to intervene.

Functional Scope

The Retention Agent operates in **batch mode** using mock telemetry data. Its responsibilities are intentionally limited:

- Analyze historical product usage per customer
- Detect negative engagement trends
- Estimate churn risk as a probability (0–1)
- Recommend a safe, human-reviewed next step
- Log human feedback for observability

Explicit Non-Goals

The agent does **not**:

- Automatically message customers
- Make retention promises
- Replace CSM judgment

- Operate in real time

Inputs and Outputs

Inputs (Mock Data)

- Weekly product usage telemetry (usage.csv)
    - customer_id
    - week
    - logins
    - leads_created
    - features_used

This mock data structure mirrors real-world SaaS telemetry commonly found in churn datasets.

Outputs

For each customer, the agent produces a **CSM Card** containing:

- Churn risk score (0–1)
- Risk band (low / medium / high)
- Top churn reasons (human-readable)
- Recommended action (suggestion only)
- Next-step guidance for the CSM
- Debug metadata (signals, thresholds, risk points)

Outputs are written to:

mock-data/outputs/retention_agent/daily_output.json

Processing Flow

Each daily batch follows a clear, deterministic intelligence flow:

1. **Load telemetry data**

2. **Group usage by customer**

3. **Sort usage by time (week order)**

4. **Extract churn signals**

5. **Compute churn risk score**

6. **Route action using rules**

7. **Format human-readable output**

8. **(Optional) Log human feedback**

This mirrors the shared backbone pattern:

Input → Understand → Decide → Suggest → Log

Signal Extraction Logic

The Retention Agent extracts explainable churn signals, including:

- **Login drop percentage** (early vs late usage)

- **Engagement drop** (leads created)

- **Feature usage drop**

- **Inactive streak length**

- **Low-usage periods count**

Signals are calculated using simple, auditable statistics (averages and percentage drops), ensuring transparency.

Churn Scoring (ML-Lite)

The agent uses an **ML-lite scoring approach**:

- Signals are converted into weighted risk points (0–100)

- Risk points are mapped to a probability using a sigmoid function

- The output is a **churn_score between 0 and 1**

Why ML-Lite?

- Fully explainable

- Easy to calibrate

- No black-box behavior

- Suitable for evaluation and safety-first environments

The ML score **never directly triggers actions**.

Action Routing (Rule-Based)

Churn scores are converted into actions using explicit thresholds:

Churn Score	Confidence	Action
≥ 0.80	≥ 0.60	Draft re-engagement (human reviewed)
$0.50 - 0.79$	any	Alert CSM
< 0.50	any	No action

This separation ensures:

- ML estimates risk

- Rules enforce business safety

Human-in-the-Loop Design

Human control is mandatory:

- All outputs are **recommendations only**

- CSMs can accept, ignore, or modify suggestions

- No penalties exist for ignoring AI output

- Feedback actions are logged explicitly

A dedicated feedback_logger records:

- Action taken (accepted / ignored / edited)

- Contextual metadata (score, confidence)

Observability and Feedback

The system is observable by design:

- Churn scores and thresholds are logged

- Human responses are recorded in JSONL format

- Debug metadata allows root-cause analysis

- False positives and override rates can be measured

This enables continuous improvement **without changing system behavior**.

Safety and Governance

Safety is enforced structurally:

- **Global kill switch**

- **Per-agent kill switch (retention_agent)**

- No outbound customer communication

- Deterministic, auditable decisions

- Explicit human review points

When the kill switch is enabled:

- No analysis is performed

- No outputs are generated

- System degrades safely

Automation Level

The Retention Agent operates at:

Level 1 – Shadow Mode

- AI provides insights only
- Humans make all decisions
- No automated execution

Higher automation levels are intentionally out of scope.

Summary

The Retention Agent is a complete, safety-first early warning system for customer churn. It combines explainable analytics, controlled ML scoring, explicit human oversight, and robust governance mechanisms. The agent adheres strictly to all required phases and is suitable for demonstration, evaluation, and future extension.