

v3.4-SEC-004 – Anomaly Detection, AI Bias Guard & Runtime Watchdogs

Document Title	Anomaly Detection, AI Bias Guard & Runtime Watchdogs
Version	v3.4
Document ID	v3.4-SEC-004
Date	2025-03-22
Author	Take Back Your Data – Secure Inference Monitoring
Document Type	Public / Certification / Internal

1. Purpose & Scope

This document defines the mechanisms for runtime anomaly detection, AI bias monitoring and watchdog enforcement within MaxOneOpen. It ensures certified forks include active safeguards against unexpected, unsafe or discriminatory execution patterns.

2. Runtime Anomaly Detection

- All forks must implement runtime monitors for behavior deviation
- Detection metrics include: latency spike, memory drift, I/O entropy
- Anomalies must trigger ZK-logged alerts and policy halt if needed
- Forks must support live audit mode for anomaly replay

3. AI Bias Guard Logic

Bias Domain	Detection Mechanism	Action Trigger
Demographic	Prompt-pair result diff	User-flag + watchdog raise
Contextual	Instruction leakage or overreach	Runtime halt
Toxicity	Weighted language model scoring	Proof log + audit flag
Censorship	Output suppression trace	ZK re-eval request

4. Watchdog Implementation

- Watchdogs must be schema-bound and runtime-local
- Forks define watchdog scope in deployment contract
- Watchdog trigger logs must be certified and undeletable
- No watchdog may self-disable or be fork-optional

5. Certification Triggers

- Certification requires bias guard results and anomaly replays
- Forks lacking watchdog enforcement are rejected
- Certification is void on watchdog bypass or deletion

6. Certification Relevance

Bias monitoring, anomaly detection and watchdog enforcement are mandatory for MaxOneOpen certification. All forks must document safeguards, responses and audit pathways in reproducible format.