# Module 12 – Trust Classification & Runtime Behavior Enforcement

Module ID: AI-TRUST-012

Version: 4.1 (Revised CTO Edition)

Layer: AI Containment & Trust Layer

Status: RELEASE

Dependencies: Module 00, Module 04, Module 11

## 0. Purpose & Enforcement Point

This module defines the trust classification logic and behavior enforcement structure for all identity-bound or AI-triggered operations in MaxOneOpen v4.1. It ensures that actors and models are trust-tiered, session-evaluated, and constrained to policy-bound runtime behavior. Trust classification is required for execution, forking, messaging, and AI evaluation.

## 1. Trust Tier Logic, Classification Capsule & Enforcement Map

Each identity or model is assigned a `Trust Classification Capsule (TCC)`:  
`{ entity\_id, classification\_type, trust\_tier (0–5), evaluation\_context, origin\_hash, timestamp }`  
Trust tiers define access, visibility, and operational scope:  
- T0 = no access  
- T1–T2 = scoped roles only  
- T3 = default operations  
- T4 = limited admin or model execution  
- T5 = system critical or audit authority  
Trust enforcement is runtime-validated per session.

## 2. Behavior Monitoring, Downgrade Policy & Violation Handling

At runtime, each action is monitored and compared to allowed trust-tier scope. Violations result in automatic trust downgrade, session invalidation, or capsule-triggered rollback. Behavior deviations create a `Trust Violation Capsule (TVC)`:  
`{ capsule\_id, session\_id, observed\_behavior, expected\_scope, trigger\_point, downgrade\_action }`

## 3. Session Enforcement Hooks & Execution Control Points

Trust controls integrate with:  
- execution start (Module 01)  
- identity validation (Module 03)  
- policy binding (Module 05)  
- AI model execution (Module 11)  
- fork initiation (Module 02)  
Each trust violation ends session scope immediately and flags identity or model for downgrade reclassification.

## 4. Forensic Capsules & Classification Ledger Anchors

Artifacts committed per trust process:  
- `Trust Classification Capsule (TCC)`  
- `Trust Evaluation Snapshot (TES)` – session-level audit trace  
- `Trust Violation Capsule (TVC)` – logged on breach, includes remediation path  
All capsules are anchored in LedgerSync and queryable via Module 14 forensic audit paths.

## 5. Intermodular Enforcement Chain & Policy Bindings

Module 12 binds to:  
- Module 01 (Execution Control) to gate execution  
- Module 03 (Identity) to define per-role tier alignment  
- Module 04 (MaxReg) to load tier-specific policy rules  
- Module 05 (ConfigBinding) for policy injection  
- Module 11 (AI Execution) to classify and constrain models  
- Module 13 (LedgerSync) and Module 14 (Audit) for trace anchoring

## CTO Validation Matrix

Module 12 (CTO Edition) guarantees the following verifiable conditions:  
- Every actor and AI model is trust-classified before operation: YES  
- Behavior violations trigger auto-downgrade and rollback: YES  
- Classification is capsule-bound, auditable and revocable: YES  
- Trust tiers define scope, rights and access boundaries: YES  
- Ledger traces reflect all trust changes and enforcement actions: YES