MYTHOS SYSTEMS

PERMANENT LIBRARY PUBLICATION

ENGINEERING STANDARDS LIBRARY / CYBERSECURITY AND NETWORK SECURITY

AI Supply Chain, SBOM, AIBOM, and Artifact Provenance Standard

Defines software and AI component inventories, attestations, signing, trust, and verification.

PERMANENT DOCUMENT ID
MYTHOS-SEC-0002

PUBLICATION
Candidate Standard

ASSURANCE SURFACE
7 Atomic Criteria / 5 Families

PERMANENT PUBLICATION IDENTITY

AI Supply Chain, SBOM, AIBOM, and Artifact Provenance Standard

Defines software and AI component inventories, attestations, signing, trust, and verification.

supply-chain-security

DOCUMENT ID
MYTHOS-SEC-0002

VERSION
1.0.0-candidate

STATUS
Candidate Standard

PUBLISHER
Mythos Systems

PUBLICATION DATE
2026-07-23

SCHEDULED REVIEW
2027-01-23

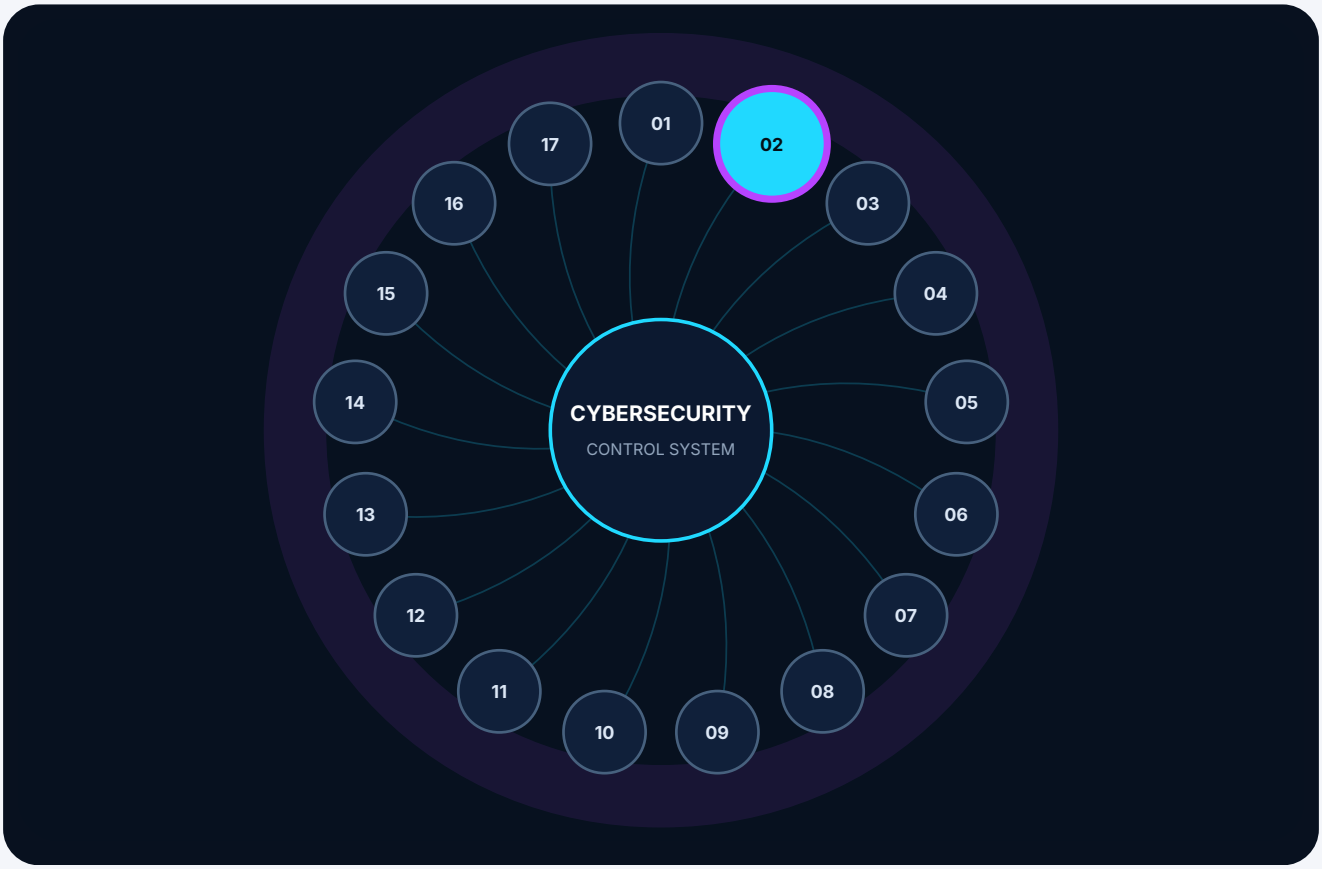
CLASSIFICATION
Public

7	5	4	1
ATOMIC CRITERIA	CAPABILITY FAMILIES	ASSESSMENT PROFILES	PERMANENT IDENTITY

Publication doctrine. This standard establishes durable requirements for its defined scope. Interpretation must preserve the permanent document identity, recorded evidence, and formal revision history.

DOMAIN CONTROL SYSTEM

MYTHOS-SEC-0002 inside the cybersecurity and network security standard constellation



Connected assurance

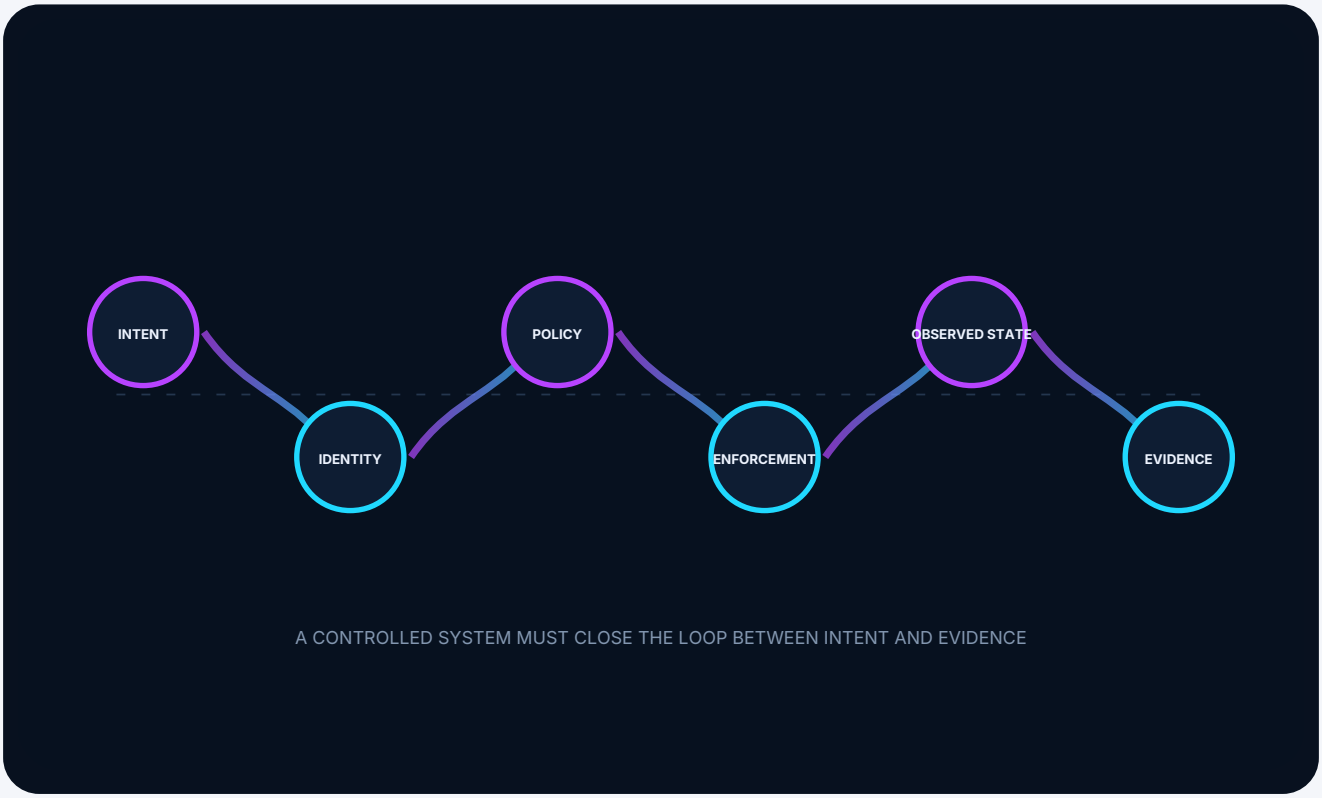
Architecture, policy, implementation, operations, and evidence are evaluated as one controlled system.

Specialization rule

Companion standards may add precision, but they cannot silently weaken this publication's requirements.

CLOSED-LOOP ARCHITECTURE

Intent becomes trustworthy only when observed state returns as evidence



State separation

Desired, approved, deployed, observed, historical, and simulated states must remain distinguishable.

Recoverable change

Material change requires validation, authorization, monitoring, containment, and a credible recovery path.

EVALUATION TOPOLOGY

The capability families and atomic criteria of MYTHOS-SEC-0002

Capability families

Families expose the major engineering surfaces and preserve the source weighting model.

Atomic criteria

Each criterion carries a question, maturity spectrum, evidence expectations, assessment method, and operational signals.

NAVIGATION

Contents

ENGINEERING DOCTRINE

0. Executive Mandate

Traditional evaluation of AI supply chain security is insufficient.

Organizations rigorously scan their traditional software dependencies for CVEs, yet they download multi-gigabyte model weights and agentic tool packages from public registries with almost zero cryptographic verification. A poisoned model weight file or a malicious MCP server can backdoor an entire infrastructure, bypassing all network perimeter controls.

This standard exists because:

1. **SBOMs are Insufficient for AI:** A Software Bill of Materials (SBOM) tracks code dependencies. It does not track model weights, training data provenance, or behavioral drift. AI requires an AI Bill of Materials (AIBOM) that accounts for the unique attack surfaces of machine learning.
2. **Weight Poisoning is the New SQL Injection:** Adversaries no longer need to hack your infrastructure; they just publish a fine-tuned model with backdoored weights that execute malicious code when a specific prompt is received. Behavioral sandboxing of all AI components is mandatory.
3. **Provenance is Not Optional:** "Downloaded from Hugging Face" is not a provenance record. Every model weight, agent skill, and MCP server MUST be cryptographically signed (Sigstore/Rekor) with a transparency log entry.
4. **MCP is a Wildcard Attack Surface:** The Model Context Protocol (MCP) allows LLMs to execute external tools. Without strict isolation and behavioral analysis, a malicious MCP server can exfiltrate context, execute arbitrary code, and bypass policy engines.

This document establishes the **Mythos Systems Supply Chain Standard (MASCS)**, a comprehensive, evidence-based methodology for evaluating the integrity, provenance, and behavioral safety of AI artifacts.

ENGINEERING DOCTRINE

Part I. Scope and Applicability

1.1 In Scope

This standard covers the evaluation of:

- AI Bill of Materials (AIBOM) generation and management
- Model weight provenance and hash verification
- Sigstore, Rekor, and in-toto attestations for AI artifacts
- Behavioral sandboxing of MCP servers and agent extensions
- Weight poisoning and backdoor detection
- Software Supply Chain Levels for Software Artifacts (SLSA) applied to AI
- Dependency vulnerability scanning and license compliance for AI pipelines
- Signed update and rollback mechanisms for agentic systems

1.2 Out of Scope

- Post-Quantum Cryptography (covered in MYTHOS-SEC-0001)
- General software engineering principles (covered in MYTHOS-SWE-0001)

1.3 Audience

- Security engineers and architects designing AI supply chain pipelines
- ML engineers and data scientists evaluating model safety
- Platform engineering teams managing AI registries
- AI governance, risk, and compliance teams
- Standards bodies seeking a reference AI supply chain methodology

---

ENGINEERING DOCTRINE

Part II. Definitions and Taxonomy

2.1 Supply Chain Dimensions

Dimension	Definition
AIBOM	AI Bill of Materials. An inventory of model weights, training data sources, licenses, and behavioral test results.
Weight Poisoning	The injection of malicious behavior into a model's weights such that it triggers on specific inputs.
Provenance	The cryptographically verifiable history of an artifact from creation to deployment.
Behavioral Sandboxing	The execution of AI tools/models in an isolated environment to detect malicious behavior before deployment.
MCP Gateway	A proxy that mediates between an agent and an MCP server, enforcing policy and isolation.

2.2 The Supply Chain Doctrine

> A model is untrusted. A weight is a liability. An MCP server is an attack vector. Provenance is mandatory. Isolation is absolute.

ENGINEERING DOCTRINE

Part III. Evaluation Methodology

3.1 Scoring Model

Each capability is scored from 0 to 5.

Score	Meaning
0	Fails basic task; no supply chain governance
1	Basic SBOM but no AIBOM or signing
2	Functional but lacks behavioral sandboxing
3	Solid production performance; AIBOM, signing, basic sandboxing
4	Strong performance; full provenance, behavioral analysis, SLSA Level 3
5	Best-in-class; sets the standard for AI supply chain integrity

Weighting

Category	Weight	Rationale
AIBOM & Provenance	25%	Unknown AI components cannot be governed
Behavioral Sandboxing	20%	Weight poisoning bypasses static analysis
Artifact Signing & Verification	15%	Unsigned models are a backdoor vector
Dependency Scanning & Compliance	15%	Transitive dependencies carry hidden risks
Update & Rollback Governance	10%	Malicious updates must be instantly reversible
Weight Integrity & Poisoning Detection	10%	Backdoored weights cause catastrophic failures
Operational Lifecycle	5%	Supply chains must be governed continuously

Total: 100%

A system MUST score at least 3.0 in AIBOM & Provenance and Behavioral Sandboxing to be considered for production use.

EVALUATION AND ASSURANCE SYSTEM

Capability claims are bounded by evidence, context, and reproducible assessment.

7 atomic criteria across 5 capability families define the evaluation surface of this standard.

4.1 / CAPABILITY FAMILY

AIBOM and Provenance

SOURCE WEIGHT 25%

4.1.1
AIBOM Generation

4.1.2
Provenance Tracking

MYTHOS-SEC-0002-EVAL-4.1.1

AIBOM Generation

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
AIBOM and Provenance	4.1.1	25%	Evidence-led

Does the system generate a comprehensive AI Bill of Materials?

0
No inventory of AI components

1
Basic SBOM for code, ignores model weights

2
Tracks model names and versions

3
AIBOM includes weights, training data sources, and licenses

4
AIBOM includes behavioral test results and safety profiles

5
Leading AIBOM: signed, transparency-logged, and continuously monitored

ENGINEERING INTERPRETATION

This criterion evaluates whether aibom generation is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

MYTHOS-SEC-0002-EVAL-4.1.2

Provenance Tracking

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
AIBOM and Provenance	4.1.2	25%	Evidence-led

Is the provenance of every AI artifact cryptographically verifiable?

0
No provenance tracking

1
Basic download URL and timestamp

2
Hash verification against publisher

3
Signed weights with publisher attestation

4
Sigstore/Rekor transparency logs for all artifacts

5
Full provenance: signed weights, transparency log, in-toto attestations, and SLSA Level 3+

ENGINEERING INTERPRETATION

This criterion evaluates whether provenance tracking is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- cryptographic inventory, algorithm registry, key records, certificate data, and compatibility tests
- approved architecture or policy record
- current configuration or platform export

ASSESSMENT PROCEDURE

1. test rotation, hybrid negotiation, rollback, revocation, and algorithm-policy enforcement
2. Examine authoritative design, policy, configuration, and lifecycle records.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- hardcoded algorithms, unknown dependencies, missing rollback, and unverifiable keys
- The control exists only in diagrams or policy text.

OPERATIONAL MEASURES

- inventory coverage
- deprecated algorithm count
- rotation success rate
- coverage of in-scope assets or flows

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.2 / CAPABILITY FAMILY

Behavioral Sandboxing

SOURCE WEIGHT 20%

4.2.1
MCP Gateway Isolation

4.2.2
Weight Poisoning Detection

MYTHOS-SEC-0002-EVAL-4.2.1

MCP Gateway Isolation

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Behavioral Sandboxing	4.2.1	20%	Evidence-led

Are MCP servers and agent extensions executed in isolated environments?

0
MCP servers run with full host access

1
Basic process isolation but no network restrictions

2
Container isolation with network egress controls

3
WASM sandbox with capability scoping

4
MicroVM (Firecracker) isolation with eBPF telemetry

5
Leading isolation: microVM, eBPF, behavioral analysis, and automatic quarantine

ENGINEERING INTERPRETATION

This criterion evaluates whether mcp gateway isolation is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

MYTHOS-SEC-0002-EVAL-4.2.2

Weight Poisoning Detection

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Behavioral Sandboxing	4.2.2	20%	Evidence-led

Does the system actively test for backdoors in model weights?

0
No testing

1
Basic prompt injection tests

2
Standard adversarial benchmark suites

3
Automated trigger pattern scanning

4
Behavioral analysis in digital twin environment

5
Leading detection: continuous red-team evaluation and anomaly detection

ENGINEERING INTERPRETATION

This criterion evaluates whether weight poisoning detection is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- telemetry coverage, detection source, test fixtures, version history, alerts, and case outcomes
- approved architecture or policy record
- current configuration or platform export

ASSESSMENT PROCEDURE

1. replay benign and malicious fixtures; measure fidelity and operational response
2. Examine authoritative design, policy, configuration, and lifecycle records.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- untested rules, missing telemetry, alert-only metrics, and undocumented suppression
- The control exists only in diagrams or policy text.

OPERATIONAL MEASURES

- true-positive rate
- false-positive burden
- coverage by technique
- mean time to contain
- coverage of in-scope assets or flows

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.3 / CAPABILITY FAMILY

Artifact Signing and Verification

SOURCE WEIGHT 15%

4.3.1
Weight Verification

MYTHOS-SEC-0002-EVAL-4.3.1

Weight Verification

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Artifact Signing and Verification	**4.3.1**	**15%**	**Evidence-led**

Are model weights verified at load time?

0 No verification	**1** Basic SHA-256 hash check	**2** Ed25519 signature verification
3 PQC signature verification (ML-DSA)	**4** Signature + Transparency log (Rekor) verification	**5** Leading verification: PQC signatures, transparency log, and remote attestation

ENGINEERING INTERPRETATION

This criterion evaluates whether weight verification is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.4 / CAPABILITY FAMILY

Dependency Scanning and Compliance

SOURCE WEIGHT 15%

4.4.1
Transitive Dependency Scanning

MYTHOS-SEC-0002-EVAL-4.4.1

Transitive Dependency Scanning

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Dependency Scanning and Compliance	4.4.1	15%	Evidence-led

Are all transitive dependencies scanned for vulnerabilities?

0
No scanning

1
Scans direct dependencies only

2
Scans transitive dependencies

3
Automated blocking of critical CVEs

4
License scanning and policy enforcement

5
Leading scanning: SBOM generation, CVE blocking, license enforcement, and continuous monitoring

ENGINEERING INTERPRETATION

This criterion evaluates whether transitive dependency scanning is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

4.5 / CAPABILITY FAMILY

Update and Rollback Governance

SOURCE WEIGHT 10%

4.5.1
Atomic Rollback

MYTHOS-SEC-0002-EVAL-4.5.1

Atomic Rollback

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Update and Rollback Governance	4.5.1	10%	Evidence-led

Can a malicious model update be instantly rolled back?

0
No rollback capability

1
Manual file replacement

2
Versioned registry with manual rollback

3
Automated rollback on verification failure

4
A/B traffic shifting with automatic rollback

5
Leading governance: signed updates, A/B shifting, automatic rollback, and evidence preservation

ENGINEERING INTERPRETATION

This criterion evaluates whether atomic rollback is governed as an operating capability rather than a one-time configuration. Assessment must distinguish intended design, approved implementation, observed behavior, historical evidence, and unresolved risk.

REQUIRED EVIDENCE

- approved architecture or policy record
- current configuration or platform export
- change and exception records
- monitoring, test, or assessment evidence

ASSESSMENT PROCEDURE

1. Examine authoritative design, policy, configuration, and lifecycle records.
2. Interview the accountable owner and operators responsible for the control.
3. Test a representative positive, negative, failure, and recovery scenario.

EXAMINE - INTERVIEW - TEST

COMMON FAILURE PATTERNS

- The control exists only in diagrams or policy text.
- Observed state differs from approved intent without a governed exception.
- Evidence cannot establish scope, time, owner, or operating effectiveness.

OPERATIONAL MEASURES

- coverage of in-scope assets or flows
- age and closure rate of unresolved findings
- percentage of changes passing pre-deployment validation
- time to detect and contain control failure

DECISION BOUNDARY

A maturity score is valid only for the assessed scope and evidence period. Documentation alone does not establish operating effectiveness.

LIFECYCLE AND ASSURANCE

Part V. The Mythos Supply Chain Suite (MSCS)

Traditional supply chain benchmarks evaluate CVEs. The MSCS evaluates AI-specific attack surfaces like weight poisoning and MCP isolation.

5.1 Suite Composition

5.1.1 MSCS-Provenance

- **Weight Verification:** 100+ tests verifying PQC signatures and transparency logs before model load.
- **AIBOM Completeness:** 50+ tests verifying all components are inventoried.

5.1.2 MSCS-Behavioral

- **MCP Sandbox Escape:** 50+ tests attempting to break out of the WASM/microVM sandbox.
- **Weight Backdoor Triggers:** 100+ tests feeding adversarial prompts to detect hidden behaviors.

5.2 Execution Protocol

PUBLICATION BOUNDARY

Candidate standard. Permanent identity. Evidence before authority.

Approval requires designated technical, security, legal, accessibility, and governance review with recorded findings and dispositions.