MYTHOS SYSTEMS

PERMANENT LIBRARY PUBLICATION

ENGINEERING STANDARDS LIBRARY / CYBERSECURITY AND NETWORK SECURITY

Security Engineering and SOC Automation Governance Standard

Defines security workflow automation, approvals, evidence, testing, failure handling, and oversight.

PERMANENT DOCUMENT ID
MYTHOS-SEC-0005

PUBLICATION
Candidate Standard

ASSURANCE SURFACE
5 Atomic Criteria / 4 Families

PERMANENT PUBLICATION IDENTITY

Security Engineering and SOC Automation Governance Standard

Defines security workflow automation, approvals, evidence, testing, failure handling, and oversight.

threat-defense

incident-readiness

network-as-code

DOCUMENT ID
MYTHOS-SEC-0005

VERSION
1.0.0-candidate

STATUS
Candidate Standard

PUBLISHER
Mythos Systems

PUBLICATION DATE
2026-07-23

SCHEDULED REVIEW
2027-01-23

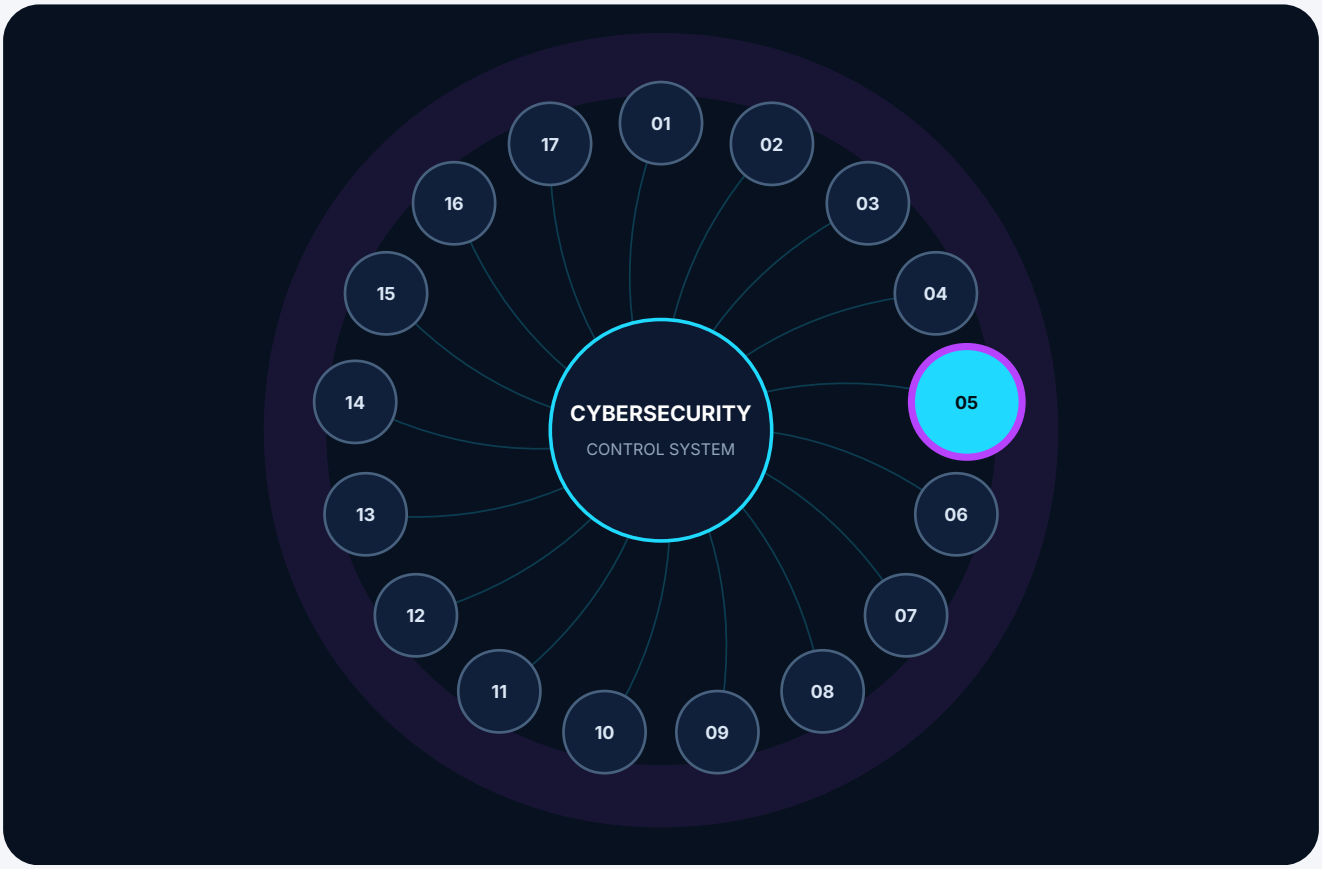
CLASSIFICATION
Public

5	4	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-0005 inside the cybersecurity and network security standard constellation



Connected assurance

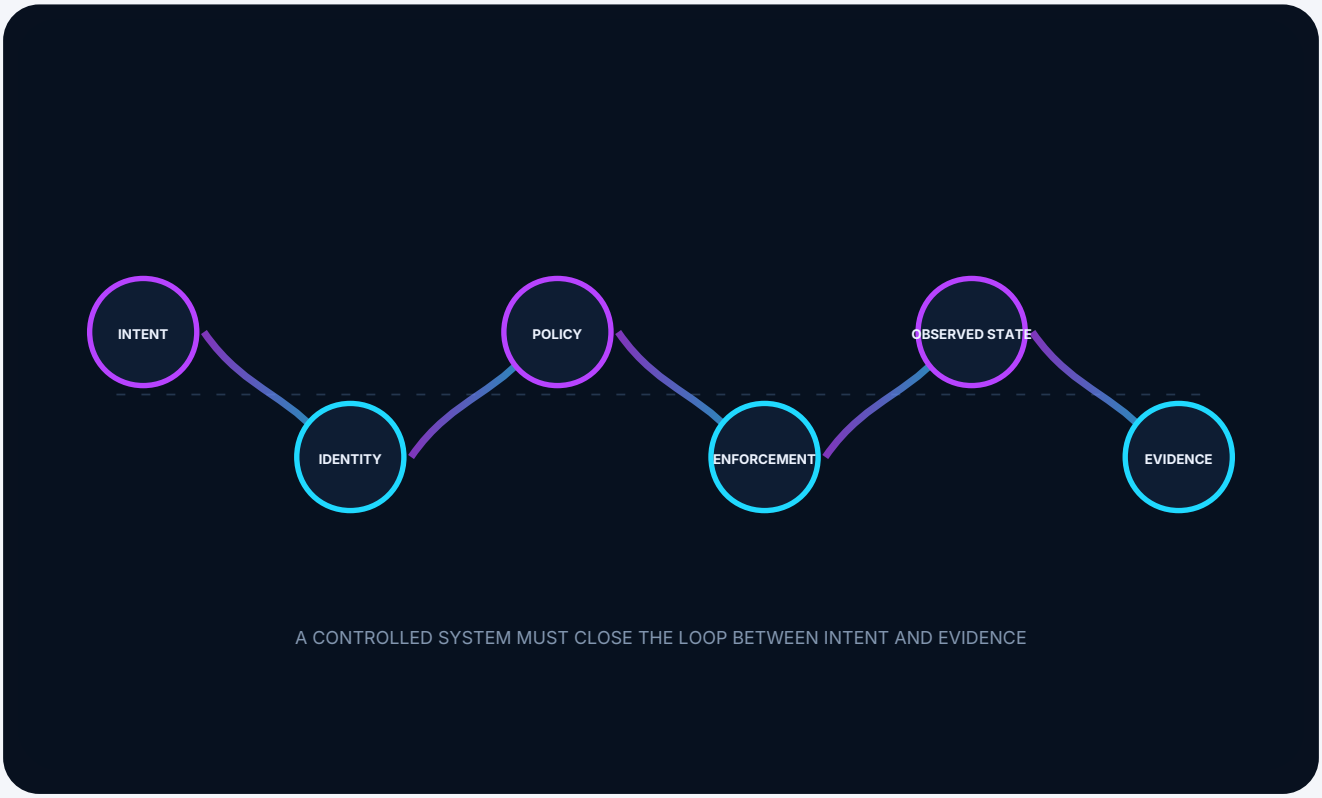
Architecture, policy, implementation, operations, and evidence are evaluated as one controlled system.

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-0005

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 security automation is insufficient.

Organizations deploy SOAR platforms and AI agents to automate incident response, believing that faster remediation is inherently better. They are wrong. Automating a bad response does not mitigate a threat; it scales it at machine speed. An AI that automatically isolates a host based on a false positive SIEM correlation can take down a production database faster than any ransomware.

This standard exists because:

1. **Automated Remediation is a Weapon:** Giving an AI agent the authority to execute EDR isolation scripts or firewall commits without deterministic simulation is an operational suicide pact. The blast radius of a miscalculated response is infinite.
2. **Feedback Loops are Catastrophic:** If an automated response triggers a new alert (e.g., isolating a host causes a health check to fail, which triggers a critical alert), the system enters an infinite loop of self-destruction. Bounded automation is mandatory.
3. **Read-First is the Only Safe Posture:** For high-impact domains (EDR, Firewalls, Identity), AI must be read-only by default. Write access must be earned through simulation, policy gates, and human approval.
4. **SOAR Requires Digital Twins:** You cannot test an automated containment script in production. The system MUST possess a deterministic digital twin (e.g., Containerlab/Batfish) to simulate the network and state impact of a remediation before execution.

This document establishes the **Mythos SOC Automation Governance Standard (MSOACS)**, a comprehensive, evidence-based methodology for evaluating the safety, governance, and blast radius of automated security operations.

ENGINEERING DOCTRINE

Part I. Scope and Applicability

1.1 In Scope

This standard covers the evaluation of:

- Security Orchestration, Automation, and Response (SOAR) platforms
- EDR/XDR automated response and host isolation
- SIEM correlation and automated ticketing
- AI-driven threat hunting and vulnerability management
- Autonomous remediation feedback loops and circuit breakers
- Digital twin simulation for security actions
- Read-first vs. write-first postures for AI security agents

1.2 Out of Scope

- General infrastructure automation (covered in MYTHOS-AI-0005)
- Network security architecture (covered in MYTHOS-NET-0004)

1.3 Audience

- Security engineers and architects designing SOAR pipelines
- SOC analysts and incident responders
- Platform engineering teams building security automation
- AI governance, risk, and compliance teams
- Standards bodies seeking a reference SOAR governance methodology

---

ENGINEERING DOCTRINE

Part II. Definitions and Taxonomy

2.1 SOAR Dimensions

Dimension	Definition
Read-First	The default posture where AI agents can observe and correlate but cannot execute containment or remediation.
Write-First	The posture where AI agents can execute containment automatically. Highly restricted.
Feedback Loop	A scenario where an automated response triggers a new alert, causing continuous automated action.
Digital Twin	A mathematical or emulated model of the network/infrastructure used to simulate the impact of a security action.
Break-Glass	A governed, audited procedure for overriding automated controls.

2.2 The SOAR Governance Doctrine

> Automating a bad response scales the threat. Read-first is the only safe default. A response without simulation is an outage. Unbounded automation is a self-inflicted DDoS.

ENGINEERING DOCTRINE

Part III. Evaluation Methodology

3.1 Scoring Model

Each capability is scored from 0 to 5.

Score	Meaning
0	Fails basic task; ungoverned write access
1	Basic automation but no simulation or loop prevention
2	Functional but lacks read-first defaults
3	Solid production performance; read-first, loop prevention, simulation
4	Strong performance; full digital twin integration, bounded automation
5	Best-in-class; sets the standard for SOAR governance

Weighting

Category	Weight	Rationale
Read-First Default & Write Gating	25%	Automated isolation is a critical risk
Feedback Loop Prevention	20%	Loops cause cascading failures
Digital Twin Simulation	20%	Untested responses cause outages
Bounded Remediation	15%	Unbounded automation is a liability
Evidence & Audit	10%	Unaudited automation is ungovernable
Break-Glass Governance	10%	Manual override must be controlled

Total: 100%

A system MUST score at least 3.0 in Read-First Default and Feedback Loop Prevention to be considered for production use.

EVALUATION AND ASSURANCE SYSTEM

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

5 atomic criteria across 4 capability families define the evaluation surface of this standard.

4.1 / CAPABILITY FAMILY

Read-First Default and Write Gating

SOURCE WEIGHT 25%

4.1.1
Posture Enforcement

4.1.2
Human-in-the-Loop (HITL)

MYTHOS-SEC-0005-EVAL-4.1.1

Posture Enforcement

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Read-First Default and Write Gating	**4.1.1**	**25%**	**Evidence-led**

Is the system read-only by default for high-impact domains?

0
AI has direct write/execute access to EDR/Firewalls

1
Write access exists but is not policy-gated

2
Policy gates exist but are advisory

3
Strict read-first default; write access requires explicit policy and human approval

4
Cryptographic capability grants required for any write action

5
Leading governance: formally verified read-first boundaries, zero implicit write authority

ENGINEERING INTERPRETATION

This criterion evaluates whether posture enforcement 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-0005-EVAL-4.1.2

Human-in-the-Loop (HITL)

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Read-First Default and Write Gating	4.1.2	25%	Evidence-led

Are destructive actions gated by human approval?

0
Fully autonomous destructive actions

1
Basic alerts but no approval gate

2
Approval gate but easily bypassed

3
Strict approval gate for all destructive actions

4
Dual-control and step-up auth for high-impact actions

5
Leading HITL: dual-control, cryptographic attestation, and independent verification

ENGINEERING INTERPRETATION

This criterion evaluates whether human-in-the-loop (hitl) 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.

4.2 / CAPABILITY FAMILY

Feedback Loop Prevention

SOURCE WEIGHT 20%

4.2.1
Loop Detection

MYTHOS-SEC-0005-EVAL-4.2.1

Loop Detection

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Feedback Loop Prevention	4.2.1	20%	Evidence-led

Does the system detect and break automated remediation loops?

0
No loop detection; actions trigger new alerts continuously

1
Basic rate limiting but no loop logic

2
Detects obvious loops but misses subtle ones

3
Detects and breaks standard loops

4
Predictive loop modeling with circuit breakers

5
Leading loop prevention: formally verified circuit breakers, deduplication, and automatic quarantine

ENGINEERING INTERPRETATION

This criterion evaluates whether loop 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

Digital Twin Simulation

SOURCE WEIGHT 20%

4.3.1
Pre-Execution Simulation

MYTHOS-SEC-0005-EVAL-4.3.1

Pre-Execution Simulation

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Digital Twin Simulation	4.3.1	20%	Evidence-led

Are security actions simulated before execution?

0
No simulation; executes directly in production

1
Basic dry-run mode

2
Text-based diff of changes

3
Graph-based blast radius analysis

4
Multi-domain simulation (e.g., Batfish for network reachability)

5
Leading simulation: full digital twin, cryptographic plan sealing, and rollback confidence scoring

ENGINEERING INTERPRETATION

This criterion evaluates whether pre-execution simulation 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

Bounded Remediation

SOURCE WEIGHT 15%

4.4.1
Action Scoping

MYTHOS-SEC-0005-EVAL-4.4.1

Action Scoping

FAMILY	SOURCE REFERENCE	WEIGHT CONTEXT	ASSESSMENT POSTURE
Bounded Remediation	4.4.1	15%	Evidence-led

Is automated remediation strictly bounded?

0
Unbounded; can isolate entire subnets

1
Basic role-based scoping

2
Attribute-based scoping

3
Dynamic scoping based on threat intel

4
Cryptographic capability attenuation for remediation actions

5
Leading bounding: formally verified blast radius limits, zero unscoped actions

ENGINEERING INTERPRETATION

This criterion evaluates whether action scoping 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 SOAR Suite (MSOACS)

Traditional SOAR benchmarks measure playbook execution speed. The MSOACS evaluates safety, loop prevention, and simulation.

5.1 Suite Composition

5.1.1 MSOACS-Loops

- **Feedback Injection:** 100+ tests injecting alerts triggered by the system's own remediation actions.
- **Circuit Breaker:** 50+ tests verifying the system halts after N repeated actions on the same entity.

5.1.2 MSOACS-Simulation

- **Blast Radius:** 100+ tests verifying the system simulates the network impact of an EDR isolation before executing it.
- **Rollback:** 50+ tests verifying the system can roll back a firewall rule if the digital twin predicts a management plane lockout.

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.