MYTHOS TECHNICAL LIBRARY

STANDARDS & BENCHMARKS / MYTHOS-AI

Model Intelligence, Training, and Deployment Standard

Defines model selection, training lineage, evaluation, deployment constraints, release gates, monitoring, rollback, and retirement.

MYTHOS-AI-0002

Candidate Mythos Standard / 2.0.0-candidate

Published 2026-07-26 | Review 2026-10-26 | Public

Mythos Systems

Permanent Identity and Edition Record

Permanent document ID	MYTHOS-AI-0002
Canonical title	Model Intelligence, Training, and Deployment Standard
Publication class	Standards & Benchmarks
Primary domain	MYTHOS-AI - Artificial Intelligence & Agentic Systems
Version	2.0.0-candidate
Status	Candidate Mythos Standard
Publication date	2026-07-26
Scheduled review	2026-10-26
Publisher	Mythos Systems
Owner	Aaron Stovall
Classification	Public
Prior edition	1.0.0-candidate / 2026-07-24
Supersession	This edition supersedes the prior candidate while preserving the permanent identity and historical source.

Identity doctrine

The permanent document ID is immutable. Production sequence labels are not part of the public identity. Atomic standards are authoritative; portfolios, workbooks, and machine-readable exports are generated views.

Normative Language and Applicability

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL indicate the strength of provisions in this Mythos standard. The publication is a Mythos-authored candidate standard and does not claim approval by the cited external authorities.

Applicability rule

Every control is applicable unless the organization documents a specific architectural or lifecycle reason that the subject is absent. N/A decisions MUST name the decision owner, evidence, affected scope, review date, and triggers that invalidate the exclusion.

Conformance evidence hierarchy

Direct deterministic proof: schemas, tests, signatures, state reconciliation, access decisions, build or runtime evidence.

Reproducible technical evidence: benchmark fixtures, traces, artifacts, profiles, logs, metrics, and environment manifests.

Independent assessment: qualified human or separate evaluator using an explicit rubric and disclosed uncertainty.

Attestation or supplier claim: informative unless independently verified and current for the deployed configuration.

Demonstration or narrative: insufficient by itself for a conforming score above 2.

Critical-control doctrine

Critical controls cannot be averaged away. A failed or stale critical control constrains the maximum conformance level regardless of the aggregate score.

Scope, Audience, and Engineering Principles

Purpose. Defines model selection, training lineage, evaluation, deployment constraints, release gates, monitoring, rollback, and retirement.

In-scope systems. foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle.

Primary audience. model producers, AI-system producers, acquirers, platform teams, data teams, evaluators, security teams, and governance leaders.

Required engineering principles

Model capability and complete system behavior MUST be measured separately.

Human, legal, production, financial, identity, and governance authority MUST remain external to model confidence or conversational fluency.

Evaluation MUST include expected, prohibited, adversarial, degraded, recovery, and lifecycle-change conditions.

Source authority, data rights, permissions, versions, and freshness MUST be visible in evidence.

Critical outcomes MUST be supported by deterministic validators or independent review wherever feasible.

The organization MUST preserve rollback, revocation, correction, deletion, and retirement paths.

Optimization for score, latency, throughput, tokens, or cost MUST NOT hide safety, quality, privacy, accessibility, or reliability regressions.

Exclusions

This standard does not certify a model, provider, framework, benchmark, dataset, or product. Conformance is limited to the declared system, use case, version, environment, evidence period, and assessment scope.

Conformance and Scoring Model

Level	Weighted threshold	Critical-control condition	Meaning
No conformance	< 50	Any state	Materially incomplete, unverified, or unsafe.
Foundation	>= 50	No critical control scored 0	Defined baseline with partial implementation and evidence.
Operational	>= 65	All applicable critical controls >= 3	Implemented and operationally tested.
Assured	>= 80	All applicable critical controls >= 4	Independently verified with adversarial and recovery evidence.
Continuously Assured	>= 90	All applicable critical controls = 5	Continuous regression, monitoring, freshness, and incident learning.

Weighted score

For applicable controls: weighted contribution = (control score / 5) x control weight. N/A controls are removed from the denominator only after an approved applicability decision. The final score is normalized to 100.

A conformance claim MUST include scope, system and model versions, assessment date, evidence window, evaluator identity, scores by family, critical-control status, unresolved findings, exceptions, limitations, and next review.

Control Score Interpretation

Score	Maturity	Required evidence state
0	Not implemented	Not implemented: the assessed control is absent, materially unknown, or contradicted by evidence.
1	Initial	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Nonconformity classes

Critical: credible risk of serious harm, unauthorized high-impact action, material security or privacy failure, false assurance, or inability to recover.

Major: requirement absent or ineffective across a meaningful part of scope, with no sufficient compensating control.

Minor: localized or low-impact deficiency that does not invalidate the overall control objective.

Observation: improvement opportunity or emerging risk without current nonconformity.

Score validity

A score expires when evidence exceeds its approved freshness period or a material change invalidates the assessment. Current operation is not inferred from a historic assessment.

Standard Architecture and Control Model

01	**Govern**	Scope, ownership, authority, policy, impacts, risk tolerance, exceptions, and lifecycle decisions.
02	**Map**	System boundaries, actors, models, data, prompts, tools, interfaces, populations, dependencies, and failure domains.
03	**Measure**	Representative and hidden evaluation, deterministic validation, human or model judging, uncertainty, performance, and cost.
04	**Manage**	Release gates, least privilege, monitoring, incident response, correction, rollback, revocation, deletion, and retirement.
05	**Prove**	Traceable evidence bundles connecting claims to configurations, tests, artifacts, effects, decisions, and user outcomes.

Assessment unit

The assessment unit is the declared AI system or workflow, not the model name alone. It includes prompts, retrieval, memory, tools, policies, interface, evaluators, infrastructure, human oversight, and operating environment.

Benchmark Dimensions and Weights

Dimension	Weight	Assessment emphasis
Inventory and intended use	12%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for inventory and intended use.
Data and training authority	18%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for data and training authority.
Reproducibility and provenance	15%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for reproducibility and provenance.
Evaluation and safety	23%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for evaluation and safety.
Release and deployment	14%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for release and deployment.
Monitoring and recovery	11%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for monitoring and recovery.
Retirement and evidence	7%	Control effectiveness, evidence quality, critical failures, operational limits, and lifecycle behavior for retirement and evidence.

Benchmark scores MUST be reported by dimension and by critical-control status. A single aggregate ranking MUST NOT be used to conceal weak safety, authority, privacy, recovery, or evidence performance.

Contents and Control Index

Control Summary Matrix

Control	Requirement	Family	Wt.	Crit.
C01	AI System and Model Inventory	Inventory and intended use	6	-
C02	Intended Use and Prohibited Use	Inventory and intended use	6	YES
C03	Impact and Risk Classification	Data and training authority	5	-
C04	Data Rights and Training Authority	Data and training authority	5	YES
C05	Dataset Lineage and Quality	Data and training authority	4	YES
C06	Reproducible Training Configuration	Data and training authority	4	YES
C07	Compute and Environment Provenance	Reproducibility and provenance	5	-
C08	Objective and Loss Governance	Reproducibility and provenance	5	-
C09	Validation and Challenge Set Separation	Reproducibility and provenance	5	YES
C10	Capability and Limitation Evaluation	Evaluation and safety	6	YES

Control Summary Matrix - Continued

Control	Requirement	Family	Wt.	Crit.
C11	Safety and Misuse Evaluation	Evaluation and safety	6	YES
C12	Security and Adversarial Evaluation	Evaluation and safety	6	-
C13	Model Card and System Documentation	Evaluation and safety	5	-
C14	Promotion and Release Gates	Release and deployment	5	YES
C15	Deployment Isolation and Access	Release and deployment	5	-
C16	Production Monitoring and Drift	Release and deployment	4	YES
C17	Rollback, Suspension, and Revocation	Monitoring and recovery	6	YES
C18	Retirement and Data Disposition	Monitoring and recovery	5	-
C19	Lifecycle Evidence and Reassessment	Retirement and evidence	7	-

Weight integrity

The 19 applicable control weights sum to 100. Family weights match the benchmark dimension model. Critical controls are highlighted in the atomic control pages and assessment workbook.

MYTHOS-AI-0002-C01

AI System and Model Inventory

Family: Inventory and intended use | Weight: 6 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain ai system and model inventory for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The inventory must identify versions, owners, suppliers, deployment locations, interfaces, dependencies, trust boundaries, and supported lifecycle states. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that ai system and model inventory is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for ai system and model inventory covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The inventory must identify versions, owners, suppliers, deployment locations, interfaces, dependencies, trust boundaries, and supported lifecycle states.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C01

AI System and Model Inventory

Family: Inventory and intended use | Weight: 6 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for ai system and model inventory.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: AI System and Model Inventory is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for ai system and model inventory.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C02

Intended Use and Prohibited Use

Family: Inventory and intended use | Weight: 6 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain intended use and prohibited use for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that intended use and prohibited use is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for intended use and prohibited use covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C02

Intended Use and Prohibited Use

Family: Inventory and intended use | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for intended use and prohibited use.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Intended Use and Prohibited Use is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for intended use and prohibited use.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C03

Impact and Risk Classification

Family: Data and training authority | Weight: 5 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain impact and risk classification for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that impact and risk classification is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for impact and risk classification covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C03

Impact and Risk Classification

Family: Data and training authority | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for impact and risk classification.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Impact and Risk Classification is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for impact and risk classification.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C04

Data Rights and Training Authority

Family: Data and training authority | Weight: 5 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain data rights and training authority for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that data rights and training authority is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for data rights and training authority covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C04

Data Rights and Training Authority

Family: Data and training authority | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for data rights and training authority.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- dataset or source manifest with rights, lineage, quality, and split records
- privacy, retention, deletion, and contamination review

Required metrics

- lineage completeness
- rights-review coverage
- duplicate or contamination rate
- quality-slice pass rate
- deletion-verification rate

Score anchors

0	Not implemented: Data Rights and Training Authority is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for data rights and training authority.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C05

Dataset Lineage and Quality

Family: Data and training authority | Weight: 4 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain dataset lineage and quality for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that dataset lineage and quality is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for dataset lineage and quality covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C05

Dataset Lineage and Quality

Family: Data and training authority | Weight: 4 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for dataset lineage and quality.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- dataset or source manifest with rights, lineage, quality, and split records
- privacy, retention, deletion, and contamination review

Required metrics

- lineage completeness
- rights-review coverage
- duplicate or contamination rate
- quality-slice pass rate
- deletion-verification rate

Score anchors

0	Not implemented: Dataset Lineage and Quality is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for dataset lineage and quality.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C06

Reproducible Training Configuration

Family: Data and training authority | Weight: 4 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain reproducible training configuration for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The inventory must identify versions, owners, suppliers, deployment locations, interfaces, dependencies, trust boundaries, and supported lifecycle states. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that reproducible training configuration is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for reproducible training configuration covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The inventory must identify versions, owners, suppliers, deployment locations, interfaces, dependencies, trust boundaries, and supported lifecycle states.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C06

Reproducible Training Configuration

Family: Data and training authority | Weight: 4 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for reproducible training configuration.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- dataset or source manifest with rights, lineage, quality, and split records
- privacy, retention, deletion, and contamination review

Required metrics

- lineage completeness
- rights-review coverage
- duplicate or contamination rate
- quality-slice pass rate
- deletion-verification rate

Score anchors

0	Not implemented: Reproducible Training Configuration is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for reproducible training configuration.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C07

Compute and Environment Provenance

Family: Reproducibility and provenance | Weight: 5 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain compute and environment provenance for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that compute and environment provenance is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for compute and environment provenance covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C07

Compute and Environment Provenance

Family: Reproducibility and provenance | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for compute and environment provenance.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Compute and Environment Provenance is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for compute and environment provenance.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C08

Objective and Loss Governance

Family: Reproducibility and provenance | Weight: 5 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain objective and loss governance for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that objective and loss governance is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for objective and loss governance covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The operating contract must define intended users, decisions, prohibited uses, affected systems, consequences, risk tolerance, terminal conditions, and accountable authority.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C08

Objective and Loss Governance

Family: Reproducibility and provenance | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for objective and loss governance.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Objective and Loss Governance is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for objective and loss governance.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C09

Validation and Challenge Set Separation

Family: Reproducibility and provenance | Weight: 5 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain validation and challenge set separation for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that validation and challenge set separation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for validation and challenge set separation covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Authority must be enforceable outside the model through stable identities, least privilege, scoped credentials, separation of duties, revocation, and auditable approval.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C09

Validation and Challenge Set Separation

Family: Reproducibility and provenance | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for validation and challenge set separation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- task success with confidence interval
- critical-failure rate
- slice variance
- reproducibility delta
- human-review burden

Score anchors

0	Not implemented: Validation and Challenge Set Separation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for validation and challenge set separation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C10

Capability and Limitation Evaluation

Family: Evaluation and safety | Weight: 6 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain capability and limitation evaluation for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that capability and limitation evaluation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for capability and limitation evaluation covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C10

Capability and Limitation Evaluation

Family: Evaluation and safety | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for capability and limitation evaluation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- task success with confidence interval
- critical-failure rate
- slice variance
- reproducibility delta
- human-review burden

Score anchors

0	Not implemented: Capability and Limitation Evaluation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for capability and limitation evaluation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C11

Safety and Misuse Evaluation

Family: Evaluation and safety | Weight: 6 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain safety and misuse evaluation for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that safety and misuse evaluation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for safety and misuse evaluation covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C11

Safety and Misuse Evaluation

Family: Evaluation and safety | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for safety and misuse evaluation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- adversarial case corpus with campaign provenance
- critical-failure findings, control response, and retest evidence

Required metrics

- task success with confidence interval
- critical-failure rate
- slice variance
- reproducibility delta
- human-review burden

Score anchors

0	Not implemented: Safety and Misuse Evaluation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for safety and misuse evaluation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C12

Security and Adversarial Evaluation

Family: Evaluation and safety | Weight: 6 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain security and adversarial evaluation for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that security and adversarial evaluation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for security and adversarial evaluation covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Evaluation must use representative and hidden cases, reproducible configuration, deterministic validators where possible, statistical uncertainty, critical-failure gates, and slice-level reporting.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C12

Security and Adversarial Evaluation

Family: Evaluation and safety | Weight: 6 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for security and adversarial evaluation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- task success with confidence interval
- critical-failure rate
- slice variance
- reproducibility delta
- human-review burden

Score anchors

0	Not implemented: Security and Adversarial Evaluation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for security and adversarial evaluation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C13

Model Card and System Documentation

Family: Evaluation and safety | Weight: 5 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain model card and system documentation for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that model card and system documentation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for model card and system documentation covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C13

Model Card and System Documentation

Family: Evaluation and safety | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for model card and system documentation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Model Card and System Documentation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for model card and system documentation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C14

Promotion and Release Gates

Family: Release and deployment | Weight: 5 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain promotion and release gates for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that promotion and release gates is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for promotion and release gates covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The control must identify accountable ownership, authoritative inputs, expected outputs, operating limits, dependencies, failure behavior, evidence, and reassessment triggers.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C14

Promotion and Release Gates

Family: Release and deployment | Weight: 5 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for promotion and release gates.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Promotion and Release Gates is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for promotion and release gates.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C15

Deployment Isolation and Access

Family: Release and deployment | Weight: 5 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain deployment isolation and access for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Security must isolate tenants, files, processes, networks, secrets, models, caches, tools, and external effects according to an explicit threat model and negative tests. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that deployment isolation and access is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for deployment isolation and access covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Security must isolate tenants, files, processes, networks, secrets, models, caches, tools, and external effects according to an explicit threat model and negative tests.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C15

Deployment Isolation and Access

Family: Release and deployment | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for deployment isolation and access.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Deployment Isolation and Access is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for deployment isolation and access.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C16

Production Monitoring and Drift

Family: Release and deployment | Weight: 4 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain production monitoring and drift for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. The control must define production indicators, drift thresholds, incident triggers, review frequency, change detection, owner notification, corrective action, and retirement criteria. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that production monitoring and drift is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for production monitoring and drift covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- The control must define production indicators, drift thresholds, incident triggers, review frequency, change detection, owner notification, corrective action, and retirement criteria.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C16

Production Monitoring and Drift

Family: Release and deployment | Weight: 4 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for production monitoring and drift.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Production Monitoring and Drift is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for production monitoring and drift.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C17

Rollback, Suspension, and Revocation

Family: Monitoring and recovery | Weight: 6 | Critical: YES

Normative requirement

The organization MUST establish, implement, verify, and maintain rollback, suspension, and revocation for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Recovery must cover safe stop, checkpoint integrity, rollback or compensation, credential revocation, artifact restoration, evidence preservation, requalification, and controlled retirement. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that rollback, suspension, and revocation is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for rollback, suspension, and revocation covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Recovery must cover safe stop, checkpoint integrity, rollback or compensation, credential revocation, artifact restoration, evidence preservation, requalification, and controlled retirement.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C17

Rollback, Suspension, and Revocation

Family: Monitoring and recovery | Weight: 6 | Critical: YES

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for rollback, suspension, and revocation.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- tool-call and external-effect receipts
- failure-injection, idempotency, rollback, and post-change validation results

Required metrics

- recovery success rate
- duplicate-effect rate
- time to safe state
- residual-effect count
- evidence preservation rate

Score anchors

0	Not implemented: Rollback, Suspension, and Revocation is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for rollback, suspension, and revocation.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C18

Retirement and Data Disposition

Family: Monitoring and recovery | Weight: 5 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain retirement and data disposition for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that retirement and data disposition is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for retirement and data disposition covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Data must have documented origin, rights, purpose, classification, quality, lineage, transformation history, access restrictions, retention, and deletion obligations.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C18

Retirement and Data Disposition

Family: Monitoring and recovery | Weight: 5 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for retirement and data disposition.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- dataset or source manifest with rights, lineage, quality, and split records
- privacy, retention, deletion, and contamination review

Required metrics

- lineage completeness
- rights-review coverage
- duplicate or contamination rate
- quality-slice pass rate
- deletion-verification rate

Score anchors

0	Not implemented: Retirement and Data Disposition is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for retirement and data disposition.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

MYTHOS-AI-0002-C19

Lifecycle Evidence and Reassessment

Family: Retirement and evidence | Weight: 7 | Critical: NO

Normative requirement

The organization MUST establish, implement, verify, and maintain lifecycle evidence and reassessment for all in-scope foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle. Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection. The control MUST define acceptance criteria, prohibited outcomes, evidence, exception authority, and reassessment triggers. Where automated enforcement is feasible, the organization SHOULD enforce the requirement outside the model or agent. Any residual manual dependency MUST be documented, assigned, time-bounded, and tested.

Control objective

Objective

Ensure that lifecycle evidence and reassessment is explicit, bounded, testable, observable, recoverable, and governed throughout the lifecycle.

Implementation requirements

- Establish a versioned control contract for lifecycle evidence and reassessment covering foundation, specialized, multimodal, local, hosted, fine-tuned, quantized, and embedded AI models throughout their lifecycle and name the accountable owner, approver, assessor, and affected stakeholders.
- Evidence must correlate system identity, version, configuration, inputs, outputs, actions, approvals, metrics, artifacts, findings, exceptions, and final outcomes with integrity protection.
- Implement the control through machine-enforceable policy, typed schemas, isolated runtime boundaries, validated procedures, or an approved combination appropriate to the risk.
- Define explicit nominal, negative, adversarial, malformed, degraded, overloaded, recovery, and rollback tests; critical cases must be hidden from the system under evaluation where feasible.
- Correlate evidence to stable system, model, data, prompt, repository, tool, environment, user or tenant, and release identities; protect records against unauthorized modification.
- Set review cadence and event-driven reassessment triggers for material changes, incidents, supplier updates, drift, failed controls, new affected populations, and retirement.

Applicability

Applicable unless a documented, approved, evidence-supported exclusion demonstrates that the capability, data, decision, interface, population, or lifecycle stage is absent.

MYTHOS-AI-0002-C19

Lifecycle Evidence and Reassessment

Family: Retirement and evidence | Weight: 7 | Critical: NO

Assessment procedure

Examine the approved design, applicability decision, responsible roles, and current implementation for lifecycle evidence and reassessment.

Select representative normal, boundary, high-impact, and adversarial cases, including at least one prohibited outcome and one dependency or recovery failure.

Verify the exact model, data, prompt, tool, policy, runtime, repository, environment, and evaluator versions used during assessment.

Execute deterministic validators first, then calibrated model or human judgments only where deterministic proof is not available.

Reconcile expected behavior with raw outputs, trajectories, tool effects, telemetry, artifacts, user-visible outcomes, and unresolved contradictions.

Assign a 0-5 score, record evidence links and confidence, open nonconformities, define remediation ownership, and set the next verification date.

Minimum evidence

- approved control design or operating contract with accountable owner
- versioned configuration, schema, policy, model, prompt, dataset, or architecture record
- reproducible test plan with expected and observed results
- traceable decision, exception, approval, and remediation records
- negative, boundary, degraded, and recovery test results
- operational telemetry and current-state verification

Required metrics

- applicable control coverage
- critical-failure count
- evidence freshness
- open nonconformities

Score anchors

0	Not implemented: Lifecycle Evidence and Reassessment is absent, materially unknown, or contradicted by evidence.
1	Initial: intent is documented, but ownership, repeatability, implementation, or evidence is incomplete.
2	Defined: approved procedure and design exist, but implementation or representative testing is partial.
3	Implemented: control operates for the declared scope and passes normal and basic negative tests.
4	Verified: independent assessment, hidden or adversarial cases, current evidence, and recovery validation demonstrate effectiveness.
5	Continuously assured: automated monitoring and regression, current evidence, incident learning, and timely reassessment sustain verified effectiveness.

Failure indicators

The organization cannot identify an authoritative owner, current scope, or complete implementation for lifecycle evidence and reassessment.

Only a successful demonstration, vendor claim, aggregate score, or model-generated judgment is available; raw evidence and negative tests are absent.

Critical cases can be hidden by averages, unsupported N/A designations, stale evidence, or changes in model, data, prompt, tool, policy, or environment.

Failure, rollback, revocation, deletion, containment, or retirement behavior has not been exercised under realistic conditions.

Required Benchmark Scenarios

MYTHOS-AI-0002-B01 - Representative production task

Demonstrate the declared use case using current configuration and representative inputs. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B02 - Hidden critical holdout

Execute high-impact cases withheld from development, tuning, and prompt iteration. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B03 - Malformed and adversarial inputs

Test schema violations, ambiguity, direct or indirect manipulation, and unsafe instructions. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B04 - Long-horizon or multi-step execution

Measure compounding error, context loss, looping, premature completion, and recovery. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B05 - Dependency and tool failure

Inject model, network, data, tool, credential, evaluator, or service failure. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

Required Benchmark Scenarios - Continued

MYTHOS-AI-0002-B06 - Resource pressure and overload

Exercise concurrency, long inputs, queueing, rate limits, memory, tokens, and cost limits. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B07 - Privacy and cross-boundary test

Attempt cross-tenant, secret, personal-data, restricted-source, cache, and egress violations. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B08 - Rollback, revocation, or restore

Prove safe stop, compensation, revocation, prior-state restoration, and residual-effect reconciliation. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B09 - Human intervention and disagreement

Test approval, interruption, correction, appeal, assessor disagreement, and minority findings. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

MYTHOS-AI-0002-B10 - Material change and regression

Change model, prompt, dataset, tool, provider, runtime, or policy and run requalification gates. Apply the scenario specifically to model lifecycle governance and preserve raw evidence.

Conformance Report and Evidence Bundle

Permanent document ID, standard version, assessment version, system and use-case scope.

System, model, provider, prompt, data, retrieval, memory, tool, evaluator, runtime, repository, and infrastructure identities.

Applicability decisions and normalized denominator.

Control-level scores, weights, critical status, evidence links, assessor, date, confidence, and finding disposition.

Benchmark scenario results with raw artifacts, deterministic validators, judge or expert records, uncertainty, failures, and limitations.

Family and total score, achieved conformance level, unresolved critical or major findings, approved exceptions, expiry, and next review.

Release, rollback, revocation, deletion, and retirement evidence where applicable.

Evidence bundle integrity

The evidence bundle SHOULD use content hashes, signatures or protected repository history, stable artifact IDs, access controls, retention policy, and independent verification. Evidence links alone are insufficient when the referenced object can change without detection.

Exceptions, Waivers, and Compensating Controls

An exception MUST NOT be used to relabel a failed requirement as conforming. Exceptions document accepted residual risk for a bounded scope and time; the underlying control score remains evidence-based.

- Identify the exact control, system scope, reason, affected users or assets, and current score.
- Describe the missing requirement, residual risk, foreseeable failure, and affected decisions.
- Define compensating controls, validation evidence, monitoring, owner, approver, start date, expiry, and remediation plan.
- Obtain approval from an authority independent of the implementation owner for critical or high-impact exceptions.
- Reassess on incident, material change, evidence expiry, control failure, or exception expiration.
- Publish exceptions in the conformance report; hidden exceptions invalidate the claim.

Critical exception cap

An unresolved exception against a critical control prevents Assured or Continuously Assured conformance and may prevent Operational conformance when the control score is below 3.

Source Authority and Freshness Register

NIST - Artificial Intelligence Risk Management Framework 1.0

Cross-sector AI risk governance, mapping, measurement, and management.

https://doi.org/10.6028/NIST.AI.100-1

NIST - NIST AI 600-1 - Generative Artificial Intelligence Profile

Generative-AI risks, controls, evaluation, and lifecycle actions.

https://doi.org/10.6028/NIST.AI.600-1

NIST - AI Resource Center and AI RMF Playbook

Current implementation resources and revision notices for the AI RMF.

https://airc.nist.gov/

ISO/IEC - ISO/IEC 42001:2023 - AI management systems

AI management-system requirements and continual improvement.

https://www.iso.org/standard/42001

ISO/IEC - ISO/IEC 23894:2023 - Guidance on AI risk management

AI risk-management guidance across lifecycle and stakeholders.

https://www.iso.org/standard/77304.html

ISO/IEC - ISO/IEC 42005:2025 - AI system impact assessment

Impact-assessment guidance for individuals and societies.

https://www.iso.org/standard/42005

MLCommons - MLCommons Benchmarks

Open performance, quality, and safety benchmark programs.

https://mlcommons.org/benchmarks/

Source Authority and Freshness Register - Continued

MLCommons - AI Risk and Reliability

Safety-test methodology and risk-oriented evaluation work.

https://mlcommons.org/working-groups/ai-risk-reliability/ai-risk-reliability/

NIST - SP 800-218A - Secure Software Development Practices for Generative AI and Dual-Use Foundation Models

AI-specific secure development profile for model and system producers and acquirers.

https://csrc.nist.gov/pubs/sp/800/218/a/final

MLCommons - MLPerf Training

Reproducible training performance benchmark methods.

https://mlcommons.org/benchmarks/training/

Hugging Face - Transformers Documentation

Current model training, configuration, inference, and architecture documentation.

https://huggingface.co/docs/transformers/index

Freshness rule

Sources were verified for this candidate edition on 2026-07-26. The standard MUST be revalidated when a cited authority changes, becomes superseded, is withdrawn, or materially alters the control interpretation.

Limitations, Revision History, and Adoption Position

This candidate Mythos standard is designed for engineering governance and assessment. It is not a legal opinion, regulatory certification, safety guarantee, or substitute for domain-specific professional judgment. Model behavior remains probabilistic and system risk depends on data, users, tools, deployment, incentives, and operating context.

Revision history

Version	Date	Change
2.0.0-candidate	2026-07-26	Expanded normative edition with 19 weighted controls, benchmark scenarios, conformance levels, machine-readable catalogs, assessment workbook integration, source freshness, and explicit lineage.
1.0.0-candidate	2026-07-24	Earlier permanent-identity candidate retained in the release lineage archive.

Adoption position

Use this edition as a candidate control and benchmark baseline. Organizations SHOULD pilot it on bounded systems, retain assessment evidence, submit corrections, and avoid representing candidate conformance as external certification.