



ENGINEERING STANDARDS LIBRARY / COMPARATIVE EVALUATIONS

Final Comparative Evaluation and Benchmark System



The canonical system for controlled Mythos comparative evaluations: mandatory gates, weighted capability, workload profiles, evidence confidence, explicit limitations, and repeatable proof.

PERMANENT DOCUMENT ID

MYTHOS-BMK-CMP-0001

PUBLICATION

Candidate Benchmark Method
Version 1.1.0-candidate

ASSURANCE PROFILE

Candidate Evaluation
Documentation-Grounded

PERMANENT PUBLICATION IDENTITY

Final Comparative Evaluation and Benchmark System

The canonical system for controlled Mythos comparative evaluations: mandatory gates, weighted capability, workload profiles, evidence confidence, explicit limitations, and repeatable proof.

FIELD	VALUE
DOCUMENT ID	MYTHOS-BMK-CMP-0001
VERSION	1.1.0-candidate
STATUS	Candidate Benchmark Method
PUBLISHER	Mythos Systems
PUBLICATION DATE	2026-07-24
SCHEDULED REVIEW	2026-10-24
CLASSIFICATION	Public
PUBLICATION DOCTRINE Permanent identity is independent of internal production sequence. Claims remain bounded by evidence, assumptions, profiles, mandatory gates, controlled exceptions, and formal revision.	

PUBLICATION ARCHITECTURE

METHOD, EVIDENCE, DECISION, AND REVISION

01 / SCOPE

Define the exact decision, target workloads, candidates, versions, exclusions, and stakeholders.

02 / GATES

Test non-compensable security, state, recovery, change, portability, and evidence conditions.

03 / SCORES

Measure atomic capability under explicit profile weights without hiding architectural differences.

04 / CONFIDENCE

Discount scores by source grade, currency, repeatability, and environment relevance.

05 / PROOF

Run controlled scenarios, failures, migration, recovery, economics, and operator-cognition tests.

06 / DECISION

Record conditional recommendation, residual risk, acceptance, owner, review date, and triggers.

ASSURANCE POSITION

This publication governs a decision process. It does not certify a candidate, guarantee future performance, or replace environment-specific engineering, legal, commercial, security, and operational review.

INTERACTIVE NAVIGATION

PUBLICATION CONTENTS

01 Purpose and governing principles

The final benchmark system prevents attractive scorecards from overstating what is known or hiding non-compensable...

02 Decision contract

Every evaluation begins with a written question, target outcome, stakeholder set, workload profiles, acceptance...

03 Candidate boundary

The unit of evaluation includes product, control plane, hosting, dependencies, identity, data, integrations, staffing,...

04 Mandatory gate model

Security and trust architecture.

05 Atomic criterion engineering

Each criterion has one measurable subject, weight, gate state, definition, minimum evidence, scoring rule, owner, a...

06 Capability score scale

0 - absent or contradicted.

07 Evidence confidence scale

E0 unsupported: modifier 0.00.

08 Weighted score calculation

For each criterion, multiply capability score by criterion weight and workload-profile weight.

09 Workload profiles and sensitivity

Profiles represent distinct operating models rather than cosmetic personas.

10 Architecture and trust assessment

Model authority, identity, data, state, execution, observation, recovery, and exit.

NAVIGATION

The native PDF outline links to each section and remains stable across portfolio assembly and archival packaging.

Purpose and governing principles

THE CANONICAL SYSTEM FOR CONTROLLED MYTHOS COMPARATIVE EVALUATIONS: MANDATORY GATES, WEIGHT

OPERATING POSITION

The final benchmark system prevents attractive scorecards from overstating what is known or hiding non-compensable failures.

01

No product is selected from one aggregate score.

02

Mandatory gates precede weighted ranking.

03

Scores are workload-specific and evidence-adjusted.

04

Architectural differences remain visible.

05

Conclusions state limitations and revalidation triggers.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Decision contract

OPERATING POSITION

Every evaluation begins with a written question, target outcome, stakeholder set, workload profiles, acceptance boundaries, exclusions, and decision date.

01

The contract distinguishes procurement interest from engineering fitness.

02

Candidate versions, services, deployment models, and assumptions are frozen for the evidence snapshot.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Candidate boundary

OPERATING POSITION

The unit of evaluation includes product, control plane, hosting, dependencies, identity, data, integrations, staffing, governance, recovery, migration, and exit.

01

Feature families or service tiers are not silently generalized.

02

Adjacent products may be included when their differing architectural layer is itself decision-relevant.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Mandatory gate model

OPERATING POSITION

Security and trust architecture.

01

Governance and approval.

02

State integrity and recoverability.

03

Observability and evidence.

04

Resilience and continuity.

05

Change safety and rollback.

06

Portability and exit.

07

Assurance confidence.

08

A failed gate cannot be offset by optional features.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Atomic criterion engineering

OPERATING POSITION

Each criterion has one measurable subject, weight, gate state, definition, minimum evidence, scoring rule, owner, and review date.

01

Compound marketing categories are decomposed until evidence can be attached unambiguously.

02

Criterion changes create a new benchmark revision and preserved lineage.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Capability score scale

OPERATING POSITION

0 - absent or contradicted.

01

1 - materially deficient.

02

2 - partial with major limitations.

03

3 - adequate for a bounded profile.

04

4 - strong and operationally credible.

05

5 - leading for the defined workload.

06

Scores describe the frozen candidate surface, not universal product quality.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Evidence confidence scale

OPERATING POSITION

E0 unsupported: modifier 0.00.

01

E1 vendor assertion: modifier 0.38.

02

E2 official documentation or source: modifier 0.64.

03

E3 controlled Mythos lab: modifier 0.78.

04

E4 production evidence: modifier 0.91.

05

E5 independent repeatable validation: modifier 1.00.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Weighted score calculation

OPERATING POSITION

For each criterion, multiply capability score by criterion weight and workload-profile weight.

01

For evidence-adjusted scoring, also multiply by the evidence modifier.

02

Normalize against the maximum possible weighted score to obtain a 0-100 index.

03

Report raw capability and adjusted confidence separately.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Workload profiles and sensitivity

OPERATING POSITION

Profiles represent distinct operating models rather than cosmetic personas.

01

Weight changes are approved before scoring.

02

Rank movement across profiles is expected and documented.

03

Near-ties trigger deeper proof rather than false precision.

04

A candidate remains ineligible where a mandatory profile gate fails.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Architecture and trust assessment

OPERATING POSITION

Model authority, identity, data, state, execution, observation, recovery, and exit.

01

Identify managed versus sovereign boundaries.

02

Record external services, plugins, providers, connectors, agents, and control loops.

03

Test confused-deputy, excess-privilege, dependency, and recovery paths.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Controlled proof lifecycle

OPERATING POSITION

Prepare representative architecture, data, configuration, versions, scenarios, expected results, instrumentation, owners, and acceptance thresholds.

01

Run normal, degraded, adversarial, recovery, migration, and retirement scenarios.

02

Preserve raw artifacts and deviations.

03

Repeat material failures after correction.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Performance and scale protocol

OPERATING POSITION

Use representative data, topology, concurrency, latency, retention, and failure conditions.

01

Measure saturation, backpressure, partial success, queue growth, state growth, and recovery.

02

Vendor or synthetic benchmarks remain contextual inputs until reproduced in the adopting environment.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Security and resilience protocol

OPERATING POSITION

Validate identity, RBAC, privileged actions, secrets, cryptography, signing, isolation, audit, incident response, backup, restore, HA, dependency loss, and degraded mode.

01

Critical failures are mandatory-gate outcomes, not weighted deductions.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Economics and lifecycle protocol

OPERATING POSITION

Model licenses, metering, infrastructure, data, storage, support, staff, integration, training, migration, downtime, and exit over multiple growth scenarios.

01

Validate upgrade, compatibility, support, content, firmware, plugin, provider, and decommission paths.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Decision and recommendation model

OPERATING POSITION

Produce profile-specific conditional recommendations.

01

Name required compensating controls and missing evidence.

02

Record rejected alternatives and why they were rejected.

03

Identify the next irreversible commitment and the evidence required before it.

04

Never publish a universal winner when workload fit differs.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Benchmark record schema

OPERATING POSITION

Evaluation ID, title, version, status, publisher, dates, route, and classification.

01

Decision contract, candidate snapshot, criteria, profiles, gates, raw and adjusted scores.

02

Sources, assumptions, limitations, risks, PoC scenarios, artifacts, findings, recommendation, owner, and revalidation triggers.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Publication quality gate

OPERATING POSITION

Permanent public identity with no production-sequence branding.

01

Interactive outline, consistent metadata, hashes, source register, route, and tags.

02

No unsupported release, certification, performance, or enterprise-readiness claim.

03

Visual and programmatic QA pass before packaging.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

Revalidation and supersession

OPERATING POSITION

Scheduled review is no later than 90 days for fast-moving product comparisons.

01

Immediate review follows major release, acquisition, license change, critical vulnerability, material outage, failed PoC, or changed workload.

02

Superseded editions remain archived with explicit status and lineage.

DECISION DISCIPLINE

Every conclusion names evidence, uncertainty, owner, residual risk, acceptance state, and revalidation trigger.

SOURCE AUTHORITY AND REVISION

CURRENT EVIDENCE SNAPSHOT

SOURCE ID	PUBLISHER	TITLE	CHECKED
NIST-CSF-20	NIST	Cybersecurity Framework 2.0	2026-07-24
NIST-SP800-53	NIST	Security and Privacy Controls for Information Systems and Organizations	2026-07-24
NIST-AI-RMF	NIST	AI Risk Management Framework	2026-07-24
CIS-CONTROLS	Center for Internet Security	CIS Critical Security Controls v8.1	2026-07-24

REVISION HISTORY

1.1.0-candidate / 2026-07-24 - permanent-publication candidate edition. Scheduled review: 2026-10-24. Earlier review is required after a major release, evidence failure, critical vulnerability, commercial change, or architecture transition.