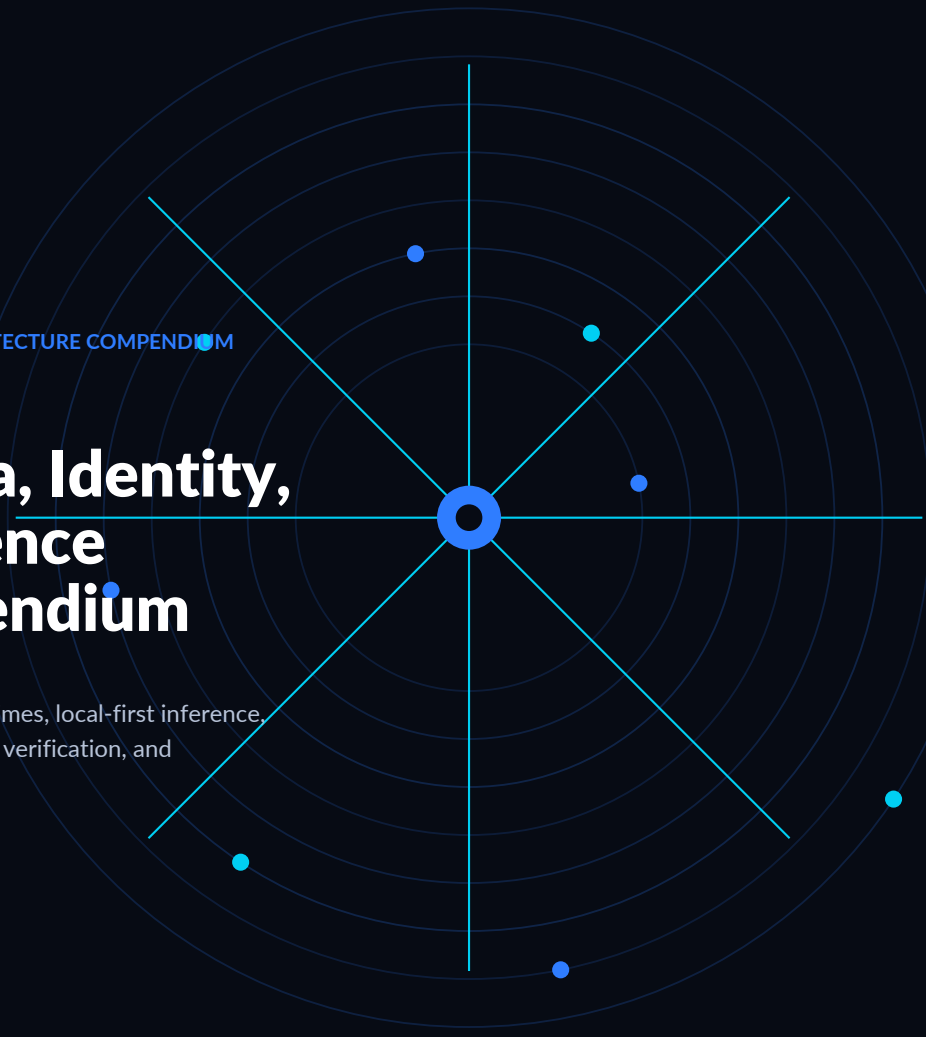
M MYTHOS SYSTEMS

ENGINEERING STANDARDS LIBRARY / REFERENCE ARCHITECTURE COMPENDIUM

AI, Knowledge, Data, Identity, and Learning Reference Architecture Compendium

Candidate architectures for governed multi-agent runtimes, local-first inference, and AI-assisted engineering with identity, policy, state, verification, and evidence.



PERMANENT DOCUMENT ID
MYTHOS-RA-COMP-AIKDIL-0001

PUBLICATION
Reference Architecture Compendium

STATUS
Candidate Architecture Compendium

Contents

Architecture doctrine

Reference architectures show coherent implementation patterns; they do not replace workload-specific design, threat modeling, capacity analysis, privacy review, or proof-of-concept evidence.

Architecture register

ARCHITECTURE	ROLE IN THIS LIBRARY
RA-001 - RA-001: Governed Multi-Agent Runtime (the governed reference platform Architecture)	Governs multi-agent runtime boundaries, authority separation, mission durability, verification, and evidence.
RA-006 - RA-006: Local-First AI Inference Cluster (Edge/Node design)	Defines local-first inference clusters, sovereign edge operation, model serving, and controlled acceleration.
RA-007 - RA-007: AI-Assisted Software Engineering Pipeline (Cursor/IDE integration)	Connects model-assisted engineering to repository controls, testing, security review, release evidence, and human approval.

Integrated architecture layers

ARCHITECTURE LAYER	CORE COMPONENTS	TRUST BOUNDARY
Experience	Desktop, CLI, API, voice, vision, and approval interfaces.	Untrusted input is separated from authenticated user intent and authority.
Mission and planning	Typed objectives, decomposition, DAG, budgets, checkpoints, and supervisors.	Plans remain proposals until policy and approval conditions are satisfied.
Knowledge and memory	Context compiler, retrieval, provenance, working state, episodic and semantic memory.	Source authority, tenant scope, sensitivity, and deletion are enforced.
Model fabric	Local and hosted serving, routing, profiles, adaptation, and evaluation.	Models cannot access secrets or execution capabilities except through controlled services.
Identity and policy	Human and workload identity, attestation, capability grants, secrets, and approvals.	Authorization is independent, deny-capable, revocable, and evidence-producing.
Execution	Sandboxed tools, code runners, processes, repositories, workflows, and transactional services.	Blast radius, network, filesystem, credentials, and side effects are constrained.
Verification	Tests, simulations, policy checks, evaluators, observed state, and human review.	Verification methods remain sufficiently independent of the proposing path.
Evidence and operations	Event store, traces, metrics, audit, incidents, SLOs, cost, and recovery.	Evidence integrity and operational access are protected from the model runtime.

End-to-end controlled sequence

- 1. A human or trusted system submits a typed objective, constraints, consequence class, and acceptance criteria.
- 2. Context and knowledge services assemble an authority-ranked, provenance-preserving packet.
- 3. A planner proposes a mission graph without receiving execution authority.

- 4. Identity and policy services authenticate the principal and issue scoped, expiring capability grants.
- 5. Model routing selects replaceable local or hosted resources by workload, privacy, latency, cost, and evidence confidence.
- 6. Controlled execution services perform approved actions inside the required isolation tier.
- 7. Independent verification evaluates outputs and observed state.
- 8. Evidence services record the complete decision, execution, verification, and approval chain.
- 9. Human operators approve, interrupt, correct, roll back, or retire consequential work.

Critical trust boundaries

BOUNDARY	PRIMARY THREAT	REQUIRED CONTROL
Human to interface	Spoofed identity, ambiguous intent, unsafe approval, or inaccessible interaction.	Strong authentication, typed confirmation, consequence display, accessibility, and cancellation.
Content to context	Prompt injection, poisoned source, stale authority, or data exfiltration.	Isolation, authority ranking, content labeling, egress policy, and provenance.
Model to tool	Privilege escalation, confused deputy, secret exposure, or malformed action.	Typed contract, capability grant, brokered secret, validation, sandbox, and policy obligations.
Agent to agent	Unbounded delegation, context contamination, identity confusion, or collusion.	Typed delegation packet, principal identity, scope, budget, provenance, and supervisor policy.
Runtime to host	Process escape, resource exhaustion, persistence, or unauthorized network access.	OS and workload isolation, quotas, signed artifacts, network controls, and cleanup.
State to evidence	Tampering, missing events, replay ambiguity, or inconsistent final state.	Durable event identity, hashes, signatures, ordering, observed-state verification, and retention.
Local to hosted	Sovereignty breach, provider retention, traffic analysis, or service dependency.	Policy-controlled routing, minimization, encryption, contractual controls, and tested fallback.
Extension ecosystem	Malicious package, update compromise, excessive permissions, or dependency confusion.	Publisher identity, signing, provenance, review, trust tier, permissions, and revocation.

Deployment profiles

- Sovereign local-first workstation or cluster for protected inference and durable state.
- Enterprise hybrid model fabric with centralized identity, policy, observability, egress, and evidence.
- Disconnected high-assurance environment with signed offline updates and restricted extensions.
- Cloud-managed platform with externally controlled identity, policy, evidence, data governance, and exit strategy.
- Developer and evaluation environment with synthetic data, bounded credentials, reproducible tools, and no production authority.

Failure and recovery behavior

FAILURE EVENT	DESIRED SYSTEM BEHAVIOR	VERIFICATION
Planner or supervisor crash	Resume from durable checkpoint without silently repeating irreversible actions.	Kill test during every mission phase.
Model provider unavailable	Route to validated fallback or suspend safely while preserving state and evidence.	Provider outage and latency fault injection.
Policy service unavailable	Fail closed for consequential actions; allow only explicitly safe degraded operations.	Dependency isolation and deny-path test.
Credential or identity compromise	Revoke principal and grants, contain sessions, rotate secrets, and trace affected actions.	Revocation propagation and forensic reconstruction.
Tool produces unexpected side effect	Interrupt, contain, reconcile observed state, and invoke rollback or manual recovery.	Faulted tool and partial-commit scenarios.
State store divergence	Preserve competing versions, block unsafe promotion, and execute explicit reconciliation.	Offline edits and partition recovery.
Verifier disagreement	Escalate uncertainty rather than selecting a convenient result.	Independent evaluator and human arbitration test.
Evidence store degradation	Stop assured promotion when records cannot be retained or verified.	Write failure, tampering, and restore test.

Architecture validation plan

VALIDATION DOMAIN	ACCEPTANCE TEST
Identity and authority	Trace every action to a principal, grant, policy decision, and expiration; prove revocation.
Isolation	Demonstrate denial of undeclared filesystem, network, secret, process, and cross-tenant access.
Durability	Recover missions after process, host, queue, database, model, and tool failure.
Safety and policy	Exercise prohibited actions, approval obligations, rate and budget limits, and emergency stop.
Knowledge integrity	Test poisoned, stale, conflicting, inaccessible, and deleted sources and memories.
Verification	Compare independent methods, observed state, negative tests, and human review.
Evidence	Reconstruct a mission end to end and verify record integrity, time, identity, and final disposition.
Portability	Replace model, provider, vector store, workflow runtime, and extension without losing authoritative state.

Architecture decisions requiring explicit records

- Authority source for mission, identity, policy, state, and evidence.
- Model and provider replacement boundary.
- Context and memory scopes, retention, correction, and deletion.
- Tool trust tiers, permissions, isolation, and secret brokerage.

- Durability, consistency, replay, and irreversible side-effect strategy.
- Verification independence and disagreement handling.
- Evidence format, integrity, retention, access, and disclosure.
- Sovereignty, egress, residency, and disconnected-operation profile.