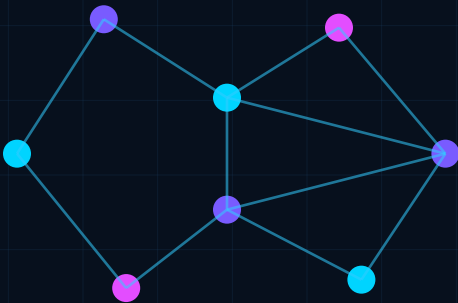


OpenTelemetry for Agents (Semantic mapping for cognitive traces)

Current-source review, evidence-bounded adoption decision, explicit limitations, reproducibility controls, and preserved source research note.



PERMANENT ID	EVIDENCE	ADOPTION	FRESHNESS
RES-010	A-	ADOPT WITH VERSIONED EXPLICIT-CYCLE / ACTIVE CHANGE	
Freshness review: 2026-09-17 / Next scheduled review: 2026-11-16			

Required Research Fields

Each field remains independently reviewable; evidence strength does not imply production authority.

EVIDENCE GRADE	A-	Strength of the retained source set and technical basis.
SOURCE AUTHORITY	2 registered authorities	OpenTelemetry Semantic Conventions; OpenTelemetry GenAI Semantic Conventions
FRESHNESS	SHORT-CYCLE / ACTIVE CHANGE	Reviewed 2026-09-17; dynamic sources require event-driven revalidation.
CONFLICTS	VISIBLE / NOT SILENT	Differences between specification, vendor behavior, research claims, and reproduced evidence must remain explicit.
LIMITATIONS	EXPLICIT	No certification, procurement approval, legal conclusion, or unbounded production claim is created by this report.
REPRODUCIBILITY	REQUIRED	Material claims require exact versions, representative data, failure cases, artifacts, and repeatable acceptance criteria.
ADOPTION POSTURE	ADOPT WITH VERSIONED EXTENSIONS	Adopt OpenTelemetry as the base telemetry substrate and version Mythos agent semantics separately while upstream GenAI and MCP conventions remain in development.
REVIEW TRIGGER	2026-11-16	Scheduled at 60 days; earlier on material source, vulnerability, implementation, or contradictory-evidence change.

Authority Register and Freshness Delta

Primary-source lineage is preserved; current-source changes are separated from unperformed implementation claims.

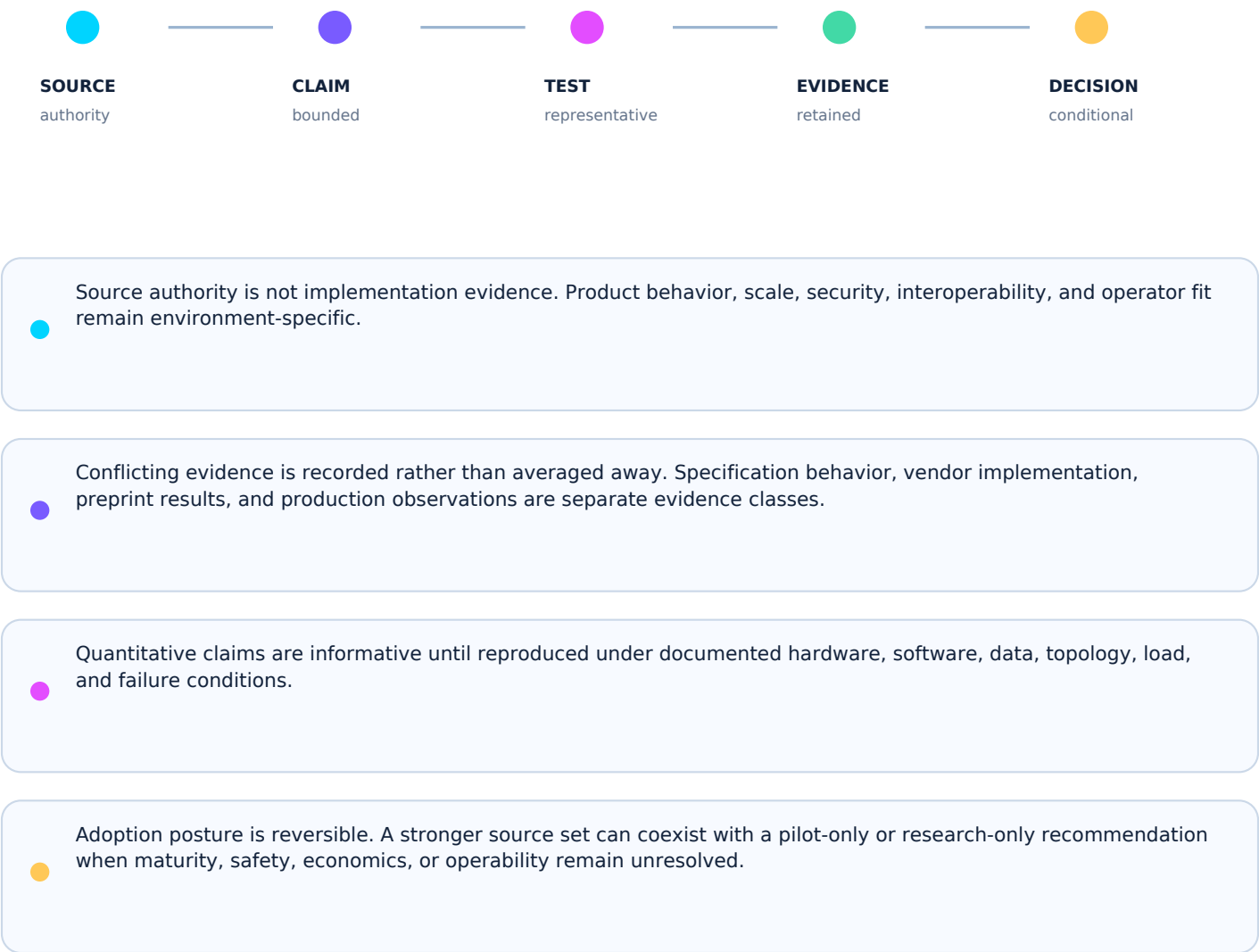
AUTHORITY 01	OpenTelemetry Semantic Conventions https://opentelemetry.io/docs/specs/semconv/
AUTHORITY 02	OpenTelemetry GenAI Semantic Conventions https://github.com/open-telemetry/semantic-conventions-genai

2026-09-17 FRESHNESS DELTA

OpenTelemetry Semantic Conventions are at 1.44.0; GenAI/agent semantics remain an active-evolution area.

Freshness policy: scheduled review + immediate review on source supersession, material release, vulnerability, retraction, or contradictory evidence.

Evidence Must Survive Contact With Reality



Decision Gate and Validation Plan

ADOPTION POSTURE

ADOPT WITH VERSIONED EXTENSIONS

Adopt OpenTelemetry as the base telemetry substrate and version Mythos agent semantics separately while upstream GenAI and MCP conventions remain in development.

- 1 / SOURCE

Pin exact versions, publication state, retrieved artifacts, and source hashes.
- 2 / PROTOTYPE

Demonstrate the core mechanism with representative data and instrumented execution.
- 3 / FAILURE

Exercise malformed input, dependency loss, stale state, overload, recovery, rollback, and misuse.
- 4 / BASELINE

Compare against an incumbent, classical, or simpler architecture using the same workload.
- 5 / ACCEPT

Record acceptance criteria, residual limitations, owner, expiration, and revalidation triggers.

EXPIRATION / REVIEW TRIGGER

Next scheduled review: 2026-11-16 (60-day cadence). Review sooner after material source revision, breaking implementation release, critical vulnerability, retraction, benchmark contradiction, changed workload, or changed legal/safety boundary.

8. Complete Source Research Note

SOURCE-LINEAGE NOTICE

The following material preserves the complete original source note, excluding only the conversational wrapper and outer Markdown fence. Legacy status labels and absolute language are retained for traceability and do not override the candidate status, limitations, or adoption decision in this edition.

Document ID: RES-010 Version: 1.0.0 (Definitive Registry Edition) Status: Published Research Note Classification: Public Organization: Mythos Systems (MYTHOS AI) Owner: Aaron Stovall Effective Date: 2026-07-16 Applies To: Platform engineers building AI observability pipelines, AI engineers instrumenting agentic frameworks, and SREs managing LLM infrastructure Companion Standards: MYTHOS-DAT-0002 (AI & Agent Observability), MYTHOS-AI-0001 (Agentic Frameworks), MYTHOS-DAT-0005 (Tamper-Evident Ledgers), RA-001 (Governed Multi-Agent Runtime)

The architecture of AI observability has failed. Organizations instrument their LLM applications using legacy APM (Application Performance Monitoring) tools designed for deterministic web requests. They trace the HTTP latency to the `/v1/chat/completions` endpoint and declare the system "observable."

This is a blind fold, not observability. A 200ms HTTP latency tells you nothing about whether the agent hallucinated, ignored a tool result, leaked a secret in its prompt, or entered a reasoning loop. Traditional traces model function calls; agent traces must model cognition.

This research note defines the mechanics of OpenTelemetry for Agents. In a Mythos-compliant architecture (like the PANTHEON runtime's Argus module), the agent's cognitive loop is mapped to the OTel GenAI semantic conventions. Every prompt, context injection, tool proposal, and policy gate is captured as a hierarchical distributed span. This cognitive trace allows engineers to reconstruct the exact state of the agent's attention window at any millisecond, transforming the AI from an opaque black box into a debuggable, deterministic state machine.

To observe an agent, we must map its non-deterministic workflow into a structured, hierarchical distributed trace. A standard web trace is flat (User -> API -> DB). An agent trace is a Directed Acyclic Graph (DAG) of reasoning.

1.1 The Root Span (The Mission)

The root span represents the user's initial intent (e.g., "Refactor the authentication module"). It lasts from the moment the user submits the prompt until the agent returns the final response. It carries high-level metrics: total token cost, total wall-clock time, and the models utilized.

1.2 The Cognitive Spans (LLM & Context)

Nested under the Root Span are the Cognitive Spans. Every time the agent invokes an LLM, a span is created.

- The Attributes: This span MUST utilize the GenAI semantic conventions (`gen_ai.*`). It captures `gen_ai.request.model`, `gen_ai.usage.prompt_tokens`, `gen_ai.usage.completion_tokens`, and `gen_ai.system`.
- The Payload: The span events capture the exact system prompt, the assembled user prompt, and the LLM's generated completion. This is the only way to debug why the agent proposed a specific action.

1.3 The Retrieval Spans (RAG & Memory)

Before the Cognitive Span, the agent queries its memory (Mnemosyne) or a vector database.

- The Attributes: Captures the search query, the top-K retrieved chunks, and the latency of the vector search.
- The Value: Allows engineers to diagnose hallucinations caused by irrelevant context being injected into the LLM.

1.4 The Execution & Policy Spans (Tools & Aegis)

When the LLM proposes a tool call, the trace forks.

- The Policy Span (Aegis): Records the OPA policy decision (Allow/Deny) and the capability grant issued.
- The Execution Span (Morta): Records the exact tool parameters, the WASM sandbox execution time, and the stdout result returned to the agent.

Mapping cognitive loops to OpenTelemetry introduces severe physical constraints that standard APM tools do not face.

2.1 The Context Window Bloat

A standard microservice trace is a few kilobytes. An agent trace containing a 50,000-token context window, a 10,000-token LLM completion, and multiple tool responses can easily exceed 1 Megabyte per trace.

- The Impact: Sending thousands of 1MB traces to Datadog or New Relic will instantly exhaust bandwidth, overwhelm the OTel collector, and incur massive financial costs.
- The Mitigation: The OTel collector MUST be configured with intelligent sampling. 100% of error traces are kept. For successful traces, only the Root and Policy spans are kept by default. Full cognitive payloads are stored in local, S3-backed object storage, and the OTel span contains a URI pointer to the local payload.

2.2 PII and Secret Exfiltration

LLM prompts are magnets for PII. Users paste internal emails, database records, and proprietary code into agents. If the OTel pipeline exports these prompts to a third-party APM SaaS, the organization has suffered a data breach via telemetry.

- The Mitigation: The OTel collector MUST run a processor (like the `attributes` or `redaction` processor) that scans span events for PII (using NER models) and secrets (using regex). PII is replaced with cryptographic pseudonyms (`[REDACTED_EMAIL: hash_123]`) before the span leaves the local boundary.

2.3 Telemetry Poisoning

An attacker uses an indirect prompt injection to force the agent to output 10,000 lines of text in a single tool call. The agent instrumentation blindly wraps this in a span event and sends it to the OTel collector. The collector crashes with an Out of Memory error, blinding the SOC during the attack.

- The Mitigation: The instrumentation MUST enforce hard limits on span attribute sizes. Any payload exceeding 10KB MUST be truncated and offloaded to an external blob store.

When organizations rely on vendor SDKs that treat LLM calls as standard HTTP requests, they fall victim to the "Black Box" debugging crisis.

3.1 The Flat Trace (The Black Box)

The trace shows: `POST api.openai.com (200 OK, 4500ms)`.

- The Flaw: The engineer knows the API took 4.5 seconds, but they do not know what the agent asked, what context it retrieved, or why it took 4.5 seconds.
- The Impact: When the agent hallucinates a destructive action, the engineer cannot determine if the RAG pipeline retrieved a bad document, or if the LLM simply ignored the system prompt. The root cause is invisible.

3.2 The Broken DAG (The Orphaned Tool)

The trace shows an LLM call, and separately, a tool execution call. But they are not linked in a parent-child hierarchy.

- The Flaw: Without span context propagation, the engineer cannot prove which LLM reasoning loop triggered which tool execution.
- The Impact: In a multi-agent system, it becomes impossible to trace the blast radius of a prompt injection. Did Agent A trigger the malicious tool call, or did Agent B?

To achieve true cognitive observability, the PANTHEON architecture (specifically the Argus module) enforces strict instrumentation boundaries.

4.1 Out-of-Process Interception (The gRPC Boundary)

The LLM and the tools are not instrumented by the application code. Instead, all traffic between Nona (Planner), Aegis (Policy), and Morta (Executor) flows over local gRPC.

- The Argus observability module acts as an out-of-process proxy. It intercepts the gRPC streams, extracts the payloads, and generates the OTel spans.
- The Value: The agent logic remains pure. The instrumentation cannot be bypassed by a buggy LLM, and it captures the exact data transmitted over the wire.

4.2 Semantic Convention Strictness

The instrumentation MUST strictly adhere to the OpenTelemetry GenAI semantic conventions. Custom, stringly-typed attributes (e.g., `llm.model_name`) break downstream parsing.

- By using standard `gen_ai.*` attributes, the traces can be ingested by specialized AI observability platforms (like Langfuse or Arize) alongside standard OTel backends.

4.3 Deterministic Replay Linking

The OTel trace ID is mathematically bound to the Durable State Machine (Clio).

- If an agent crashes and is replayed by Temporal, the new OTel trace is linked to the original trace ID via a `replay_of` span link.
- The Value: Engineers can view the entire lifecycle of an agent, including its death and resurrection, as a single continuous narrative.

An agent that cannot be traced cannot be trusted. By mapping the cognitive loop to OpenTelemetry GenAI semantic conventions, enforcing edge redaction, and utilizing out-of-process interception, we transform the AI from an unpredictable black box into a transparent, debuggable system. If an engineer cannot reconstruct the exact context window and policy decision that led to an agent's action, the agent is not production-ready.

An HTTP latency is a heartbeat; a cognitive trace is a thought.
A prompt is PII; an unredacted span is a breach.
A flat trace is a black box; a DAG is a confession.

If the telemetry cannot reconstruct the context, the agent is unobservable.

> Cognition may be non-deterministic.
> Observability must be absolute.

End of Research Note RES-010

© 2026 Mythos Systems. This research is published for public reference. Attribution required.

9. Source Register

Source	Authority and Status	Freshness Control
OpenTelemetry Semantic Conventions https://opentelemetry.io/docs/specs/semconv/	Primary observability specification Living Published: Living	Validated: 2026-07-22 Review on material revision, withdrawal, supersession, or scheduled report review.
OpenTelemetry GenAI Semantic Conventions https://github.com/open-telemetry/semantic-conventions-genai	Primary emerging specification Development - volatile Published: Living	Validated: 2026-07-22 Review on material revision, withdrawal, supersession, or scheduled report review.

10. Lineage, Review, and Publication Record

Record	Value
Original source file	RES-010_OpenTelemetry_for_Agents_Semantic_mapping_for_cognitive_traces.md
Original source words	1407
Upgrade type	Enterprise research report; source and freshness validation; limitations; adoption decision; reproducibility plan
Generated on	2026-07-22
Current status	Candidate Research Report
Publication authority	Formal Mythos approval not yet recorded
Next review	60 days or event-driven trigger, whichever occurs first