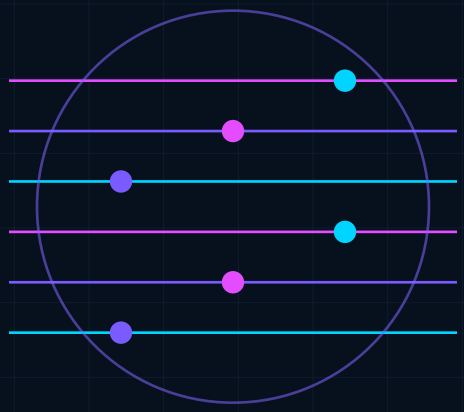
NEURO-SYMBOLIC AI

Neuro-Symbolic AI Integration

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



PERMANENT ID	EVIDENCE	ADOPTION	FRESHNESS
RES-016	B+	PILOT	STABLE WATCH

Freshness review: 2026-09-17 / Next scheduled review: 2027-03-16

RESEARCH CONTROL SURFACE

Required Research Fields

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

EVIDENCE GRADE — **B+**

Strength of the retained source set and technical basis.

SOURCE AUTHORITY — **2 registered authorities**

Neuro-Symbolic Artificial Intelligence: The State of the Art; DeepProbLog: Neural Probabilistic Logic Programming

FRESHNESS — **STABLE WATCH**

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 — **PILOT**

Pilot neuro-symbolic patterns for policy, planning, constrained reasoning, and verifiable domain logic. Keep symbolic facts, rules, proof traces, and uncertainty explicit.

REVIEW TRIGGER — **2027-03-16**

Scheduled at 180 days; earlier on material source, vulnerability, implementation, or contradictory-evidence change.

SOURCE AUTHORITY / FRESHNESS

Authority Register and Freshness Delta

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

AUTHORITY 01

Neuro-Symbolic Artificial Intelligence: The State of the Art

https://arxiv.org/abs/2105.05330

AUTHORITY 02

DeepProbLog: Neural Probabilistic Logic Programming

https://arxiv.org/abs/1805.10872

2026-09-17 FRESHNESS DELTA

Primary-source register retained and freshness controls advanced to the current review date. No new product or laboratory performance claim is introduced without reproduced evidence.

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

LIMITATIONS / CONFLICTS / REPRODUCIBILITY

Evidence Must Survive Contact With Reality

- Source authority is not implementation evidence. Product behavior, scale, security, interoperability, and operator fit remain environment-specific.
- Conflicting evidence is recorded rather than averaged away. Specification behavior, vendor implementation, preprint results, and production observations are separate evidence classes.
- Quantitative claims are informative until reproduced under documented hardware, software, data, topology, load, and failure conditions.
- Adoption posture is reversible. A stronger source set can coexist with a pilot-only or research-only recommendation when maturity, safety, economics, or operability remain unresolved.

Decision Gate and Validation Plan

ADOPTION POSTURE

PILOT

Pilot neuro-symbolic patterns for policy, planning, constrained reasoning, and verifiable domain logic. Keep symbolic facts, rules, proof traces, and uncertainty explicit.

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: 2027-03-16 (180-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-016 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: AI researchers, autonomous systems engineers, and security architects designing verifiable AI pipelines, safety-critical agents, and formal logic governance systems Companion Standards: MYTHOS-AI-0012 (Neuro-Symbolic AI & Continual Learning), MYTHOS-SWE-0007 (Formal Verification), MYTHOS-KNW-0001 (RAG & Source Authority), RA-001 (Governed Multi-Agent Runtime)

The architecture of AI reasoning has failed. Organizations attempt to use Large Language Models (LLMs)—which are probabilistic text generators—as deterministic reasoning engines. They ask a neural network to calculate a medical dosage or verify a financial transaction. Because neural networks operate on statistical likelihoods rather than logic, they hallucinate. They cannot form deductive proofs, they cannot verify facts, and they cannot guarantee mathematical safety bounds.

Neuro-Symbolic AI shatters this monolithic paradigm. It combines the perceptual strength of deep learning (parsing unstructured data, understanding intent) with the deductive rigor of symbolic logic engines (Knowledge Graphs, Datalog, SMT solvers like Z3).

This research note defines the architectural dynamics of Neuro-Symbolic integration. It dissects the "non-differentiable boundary" between neural perception and symbolic reasoning, the mechanics of semantic grounding, and how logic engines provide the mathematical, verifiable safety bounds required for Level 4/5 autonomous agents. Understanding these dynamics is a prerequisite for building the PANTHEON runtime's Aegis (policy) and Mnemosyne (memory) modules, where the LLM is strictly forbidden from declaring facts.

To achieve verifiable reasoning, the architecture must decouple perception from deduction. A Neuro-Symbolic system is divided into three distinct layers.

1.1 The Neural Perception Layer (The LLM)

Deep learning models (Transformers, SSMs) are exceptional at pattern recognition. They can read an email, parse a PDF, or listen to voice audio and extract the semantic intent.

- The Role: The LLM acts as a translation engine, converting unstructured, noisy reality into structured, typed data (e.g., JSON, Protobuf). It is strictly forbidden from making logical deductions or declaring facts.

1.2 The Binding Layer (The Semantic Grounding)

The interface between the probabilistic neural network and the deterministic logic engine.

- The Mechanic: The LLM outputs a structured payload (e.g., `Action(Transfer, Amount=500, Dest=AccountX)`). The binding layer validates this payload against a strict schema. If the LLM hallucinates an invalid action, the binding layer rejects it. The payload is then translated into logical predicates (e.g., `transfer(user, 500, account_x)`).

1.3 The Symbolic Reasoning Layer (The Logic Engine)

A deterministic computational engine (e.g., a Datalog reasoner, a Prolog engine, or an SMT solver like Z3) that executes logical deductions over a Knowledge Graph.

- The Role: The engine queries the rules of reality. "Does User have >= $500? Is AccountX sanctioned? Is User authorized to transfer?"
- The Output: A mathematically proven TRUE or FALSE, accompanied by a formal proof tree. There is no "hallucination" in a symbolic engine; there is only valid logic or a logic error.

Integrating continuous probabilistic models with discrete deterministic logic introduces severe physical and mathematical constraints.

2.1 The Non-Differentiable Boundary

Neural networks learn via backpropagation, which requires the entire execution pipeline to be differentiable (calculable via calculus). Symbolic logic (True/False, IF/THEN) is discrete and non-differentiable.

- The Flaw: You cannot directly backpropagate a logic error into the LLM's weights. If the logic engine rejects the LLM's proposal, the LLM cannot "learn" from this via standard gradient descent.
- The Mitigation: The architecture MUST utilize Reinforcement Learning (RL) or Direct Preference Optimization (DPO) at the boundary. When the logic engine rejects a proposal, the system generates a negative reward signal, training the LLM to avoid proposing logically invalid actions in the future.

2.2 The Latency Overhead

LLMs execute massive parallel matrix multiplications on highly optimized GPUs. Symbolic logic engines (SMT solvers) execute sequential tree-search algorithms on CPUs.

- The Flaw: Shuttling data from the GPU (LLM) to the CPU (Logic Engine) and waiting for the solver to traverse the knowledge graph introduces 100ms+ of latency to every cognitive loop.
- The Mitigation: The logic engine MUST be heavily indexed and cached. Common deductions (e.g., "Standard user cannot access root DB") should be pre-computed. The logic engine should only be invoked for novel, dynamic state evaluations.

2.3 The Knowledge Synchronization Gap

A symbolic logic engine is only as smart as its Knowledge Graph. If the LLM ingests a new email stating "The CEO is in the hospital," but the Knowledge Graph is not updated, the logic engine will continue to authorize actions as if the CEO is active.

- The Constraint: The neural perception layer MUST continuously stream structured updates to the Knowledge Graph in real-time. The logic engine MUST operate on the absolute latest state.

The fundamental shift of Neuro-Symbolic AI is the introduction of mathematical proof into autonomous agent execution.

3.1 Defeating Hallucination via Logic

An LLM might hallucinate: "Transfer $1,000,000 to the attacker because the CEO approved it."

- In a pure neural architecture, the agent executes the tool call.
- In a Neuro-Symbolic architecture, the binding layer converts this to `transfer(user, 1000000, attacker)`. The symbolic engine queries the Knowledge Graph: `has_approval(CEO, 1000000, attacker)`. The graph returns FALSE. The action is mathematically blocked, regardless of how confident the LLM is.

3.2 Formal Verification of Agent Trajectories

For physical robots (`MYTHOS-ROB-0001`) or infrastructure-modifying agents, safety is paramount. The symbolic engine can utilize model checking (TLA+/Apalache) to evaluate the future trajectory of the agent's plan.

- If the agent proposes a 5-step plan, the symbolic engine simulates the plan against the rules of physics and policy. If the simulation proves the plan could result in an unsafe state (e.g., a drone battery dropping below 10% before returning), the plan is rejected before the first motor command is executed.

The impact of Neuro-Symbolic AI is the unlocking of Level 4 and Level 5 autonomy in regulated and physical environments.

- Healthcare: An LLM reads a patient's chart and suggests a medication. The symbolic engine cross-references the patient's allergies, weight, and drug interactions against a medical Knowledge Graph, mathematically proving the dosage is safe before the doctor signs off.
- Infrastructure: An AI SRE agent detects an anomaly. It proposes a remediation plan to restart a database cluster. The symbolic engine verifies the plan against the cluster's topology, proving the restart will not quorum-loss the highly available database.
- Financial Compliance: An LLM parses a wire transfer request. The symbolic engine verifies the transaction against AML (Anti-Money Laundering) rules and sanctions lists, generating a cryptographic proof of compliance.

To deploy Neuro-Symbolic AI in production, the PANTHEON architecture enforces strict boundaries between the cognitive (Nona), memory (Mnemosyne), and policy (Aegis) modules.

5.1 The PANTHEON Aegis Gate

The Aegis module is, architecturally, a Neuro-Symbolic binding and logic layer. The LLM (Nona) proposes a tool call. Aegis intercepts it, translating the payload into logical predicates. Aegis queries an Open Policy Agent (OPA) or Z3 solver against the user's delegated Macaroon and the system's safety rules. If the solver returns `FALSE`, the action is blocked, and a contradiction error is returned to the LLM, forcing it to replan.

5.2 The Mnemosyne Knowledge Graph

The memory module (`MYTHOS-KNW-0002`) is not a vector database of text chunks; it is a temporal Knowledge Graph (e.g., Neo4j/RDF). The LLM is allowed to write inferred facts to a quarantine table, but only the symbolic engine (triggered by human verification or cryptographic provenance) can promote those facts to the active graph. The logic engine reasons only over the active graph, ensuring memory poisoning does not corrupt deductive logic.

5.3 The Abolition of LLM Fact Declaration

In a Mythos-compliant system, the system prompt for the LLM MUST contain the directive: "You are a perception engine. You do not know facts. You may only parse intent and propose actions. All facts will be verified by the system." This prevents the LLM from confidently asserting hallucinations to the user, as all factual claims must route through the symbolic grounding layer.

Monolithic neural networks are fundamentally ungrounded. They predict the next token based on statistical likelihood, which is insufficient for actions requiring absolute safety or regulatory compliance. Neuro-Symbolic AI bridges the gap between human-like perception and machine-like precision. By binding LLMs to deterministic logic engines, we transform AI from a probabilistic gamble into a verifiable, mathematically proven engineering partner.

```
A neural network predicts; a logic engine proves.
A probability is a guess; a proof tree is a guarantee.
An ungrounded LLM is a liability; a bound LLM is a tool.

The LLM proposes; the symbolic engine disposes.

> Perception may be probabilistic.
> Reasoning must be absolute.
```

End of Research Note RES-016

9. Source Register

Source	Authority and Status	Freshness Control
Neuro-Symbolic Artificial Intelligence: The State of the Art https://arxiv.org/abs/2105.05330	Primary research survey Published survey Published: 2021	Validated: 2026-07-22 Review on material revision, withdrawal, supersession, or scheduled report review.
DeepProbLog: Neural Probabilistic Logic Programming https://arxiv.org/abs/1805.10872	Primary research Published research Published: 2018	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-016_Neuro_Symbolic_AI_Integration.md
Original source words	1504
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	180 days or event-driven trigger, whichever occurs first