

Leio: Sub-Millisecond Decision Infrastructure

High-Throughput Ontology Lattices for Autonomous Software & AI Agents

Seed Financing Memorandum
Q4 2026

Mycelia

2026-10-02

AVERAGE LLM LATENCY

The Agent Inference Bottleneck

1,450 ms

PER 1M OUTPUT TOKENS

\$15.00+

Per 1M Output Tokens

Autoregressive generation burns massive GPU compute pretending to write text for simple routing choices.

HALLUCINATION INVARIANT

Stochastic Schema Failures

Non-Zero

Unbounded generative models fail hard on schema invariants and hierarchical regulatory constraints.

AUTONOMIC REFLEX ENGINE

The Missing Layer

System 1

Autonomous agents lack a sub-conscious reflex layer to handle deterministic routing without invoking the full model.

DETERMINISTIC FAST PATH

Kahneman Cognitive Architecture

92%

IN-MEMORY REFLEX PATH

< 1 μ s

In-Memory Reflex Path

Evaluates routine observations, schema triage, tool selection, and invariant guardrails in microseconds.

DELIBERATIVE SYNTHESIS

System 2: Deep LLMs

8%

Reserved exclusively for open-ended synthesis, customer-facing creative copy, and complex multi-turn negotiation.

OVERALL LATENCY CUT

Compound Efficiency

85%

By pre-filtering 92% of events through Leio, agent pipelines drop latency by 85% and context token burn by 90%.

ZERO-DEPENDENCY PAYLOAD

Pure Rust Core & WebAssembly

189 KB

WASM NATIVE EXECUTION

800 ns

WASM Native Execution

Compiles directly into standalone WebAssembly running client-side in browsers, Cloudflare Workers, and edge devices.

FULL 730 DROID CLASSES

Memory Footprint

< 8 MB

Ultra-compact memory overhead enables synchronous execution in V8 isolates without cold starts.

ZERO DATA RETENTION

Air-Gapped Privacy

100%

Zero external telemetry or outbound calls. Certified for air-gapped private VPCs, SOC 2 Type II, and HIPAA BAA.

MONOTONICITY INVARIANT

Algebraic Order-Theoretic Proofs

$$P(A) \leq P(B)$$

EXCLUSION KERNELS

$$A \sqcap B = \perp$$

Exclusion Kernels

Mutually exclusive concept branches are algebraically zeroed, preventing contradictory state collisions.

AXIOMATICALLY BOUNDED

Hierarchical Consistency

Zero Drift

If a diagnosis is 95% Cancer, it is mathematically impossible for the system to say it is not a Disease.

BOUNDED DECISION SPACE

Down-Set Confinement

100%

Decisions are strictly confined to reachable transitive closures within the formal ontology directed graph.

PER 1M INPUT TOKENS

Radical Cloud Unit Economics

\$0.042

OUTPUTS ALWAYS FREE

\$0.00

Outputs Always Free

Because output decisions are typed machine representations, there is zero autoregressive generation cost.

SAAS INFRASTRUCTURE MOAT

Gross Margin Profile

92%+

O(K) candidate generation +
O(1) transitive closures yield
negligible marginal compute
costs.

NET COST REDUCTION

Customer Annual Savings

250x

High-volume agent workflows save
over \$50,000 to \$200,000 annually
compared to general LLM API bills.

TIME TO FIRST DECISION

The Agent Ecosystem Wedge

30 Seconds

CLAUDE & CURSOR READY

Native MCP

Claude & Cursor Ready

Single-line execution via `uvx olaya-mcp` or `npx @mycelia/olaya-mcp` connects Leio directly into agent frameworks.

PYTHON, TS, RUST, GO

Universal Client SDKs

4 Runtimes

Native client packages provide typed interfaces and drop-in LangChain / CrewAI / Vercel AI SDK wrappers.

IN-BROWSER SANDBOX

Self-Serve POC Velocity

Zero Login

Developers test custom category schemas and copy production-ready code directly from the web workbench.

PRE-COMPILED LATTICES

Standardized Domain Ontologies

6 Standard

HIGH-SPEED LOGISTICS

EPCIS 2.0

High-Speed Logistics

Powers real-time conveyor sorting, cold-chain compliance, and hazmat routing at 4 m/s in automated cross-docks.

ICD-10 HARMONIZATION

DOID Clinical Medicine

730 Classes

Categorizes clinical records, pathology observations, and FHIR interoperability without transmitting patient PII.

PROPRIETARY FINE-TUNING

Custom Enterprise Trees

< 15 Mins

Enterprise customers import proprietary product taxonomies, ERP material groups, and internal support trees.

ACTIVE EDGE INFRASTRUCTURE

Production Deployment Proof

100% Live

SERVERLESS HTTP API

< 15 ms

Serverless HTTP API

Tested and deployed globally on Vercel Edge with sub-microsecond in-browser WebAssembly fallbacks.

AUTONOMOUS AGENT
RUNTIME

Interactive
Evidence

3 Scenarios

Real-time side-by-side agent execution traces for Document OCR, Retail Normalization, and System Interoperability.

MONTHLY LATENCY SAVED

Customer Value Proof

400+ Hrs

At 1M decisions/month, customers eliminate over 400 hours of cumulative agent blocking time.

DEDICATED VPC APPLIANCE

Enterprise Monetization Engine

\$42K+ ACV

HIGH-THROUGHPUT EDGE API

\$49 / mo

High-Throughput Edge API

Usage-based on-ramp capturing hundreds of thousands of independent developers and AI startups.

TEAM ORCHESTRATION

**Tier 2: Production
Growth**

\$499 / mo

Dedicated rate limits, priority SLA, WebMCP agent monitoring, and custom fine-tuning pipelines.

DEDICATED PRIVATE CLUSTER

**Tier 3: Enterprise Air-
Gapped**

\$3,500 / mo

Single-tenant air-gapped VPCs, Net-30 procurement, custom MSA, 99.999% SLA, and custom ontology curation.

18-MONTH RUNWAY TARGET

Seed Financing Objectives

\$2.5M Seed

MASSIVE GRAPH CAPACITY

1M+ Nodes

Massive Graph Capacity

Scale contrastive embedding kernels to support million-concept enterprise retail and biomedical taxonomies.

FORTUNE 500 DEPLOYMENTS

Enterprise Pilot Pipeline

10 Pilots

Close dedicated air-gapped VPC agreements across global logistics networks, healthtech, and fintech leaders.

W3C & AGENT STANDARDS

Standardization Leadership

WebMCP

Establish Leio as the canonical System 1 decision protocol across the autonomous agent ecosystem.