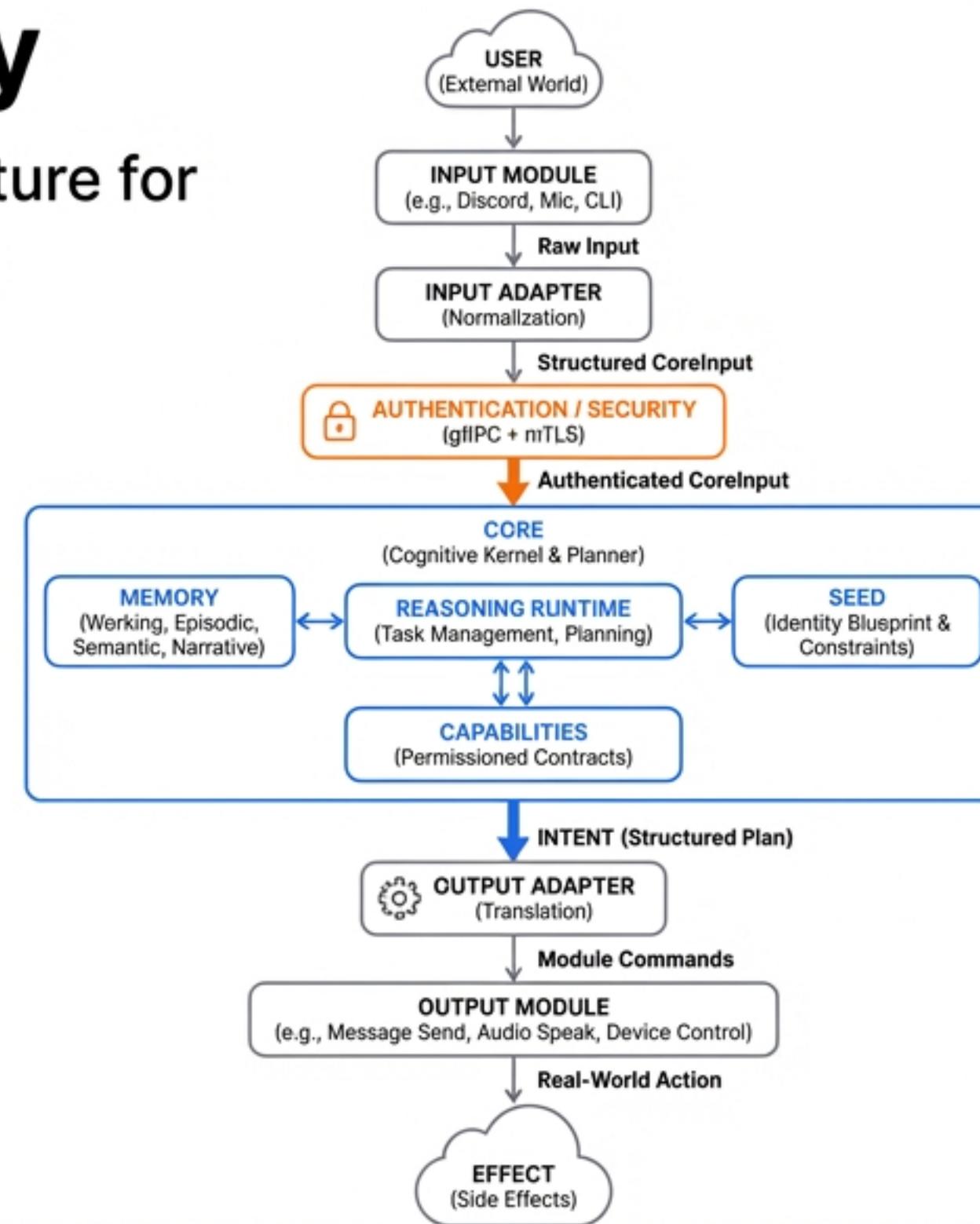


ANIMA: A Cognitive Kernel for Long-Lived Artificial Identity

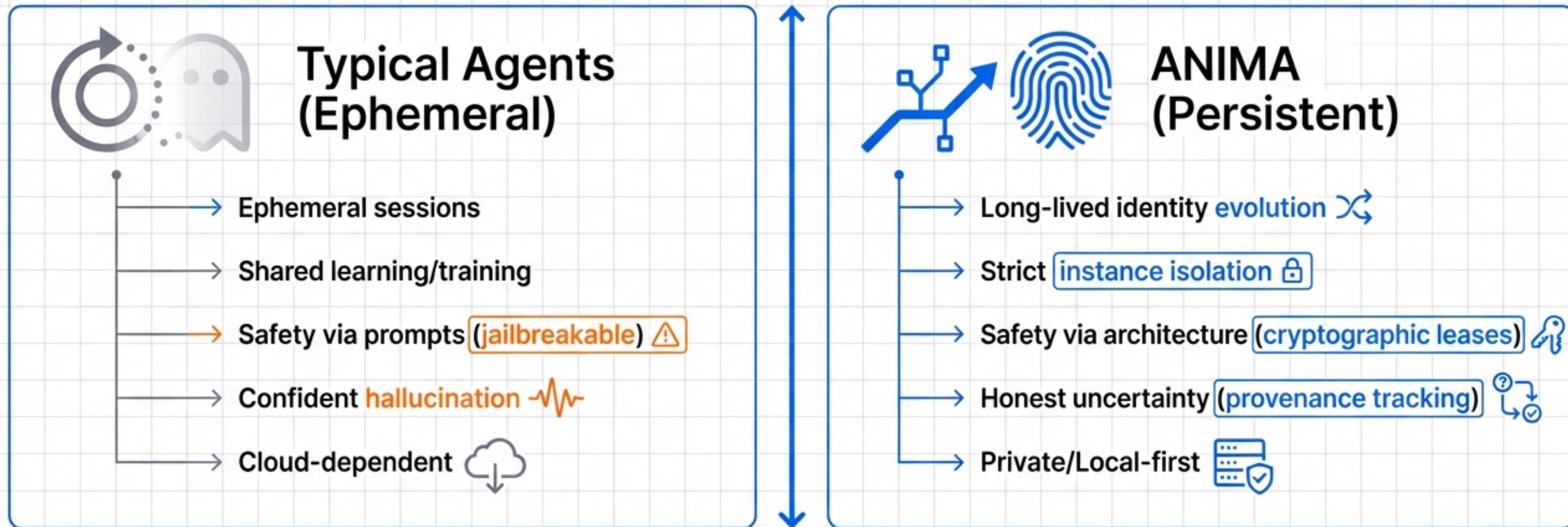
Beyond the Chatbot: Architecture for Safety, Continuity, and Trust.



Core Definition: A private, modular AI engine designed to host evolving identities under strict safety constraints.

Philosophy: Principles first, implementation second.

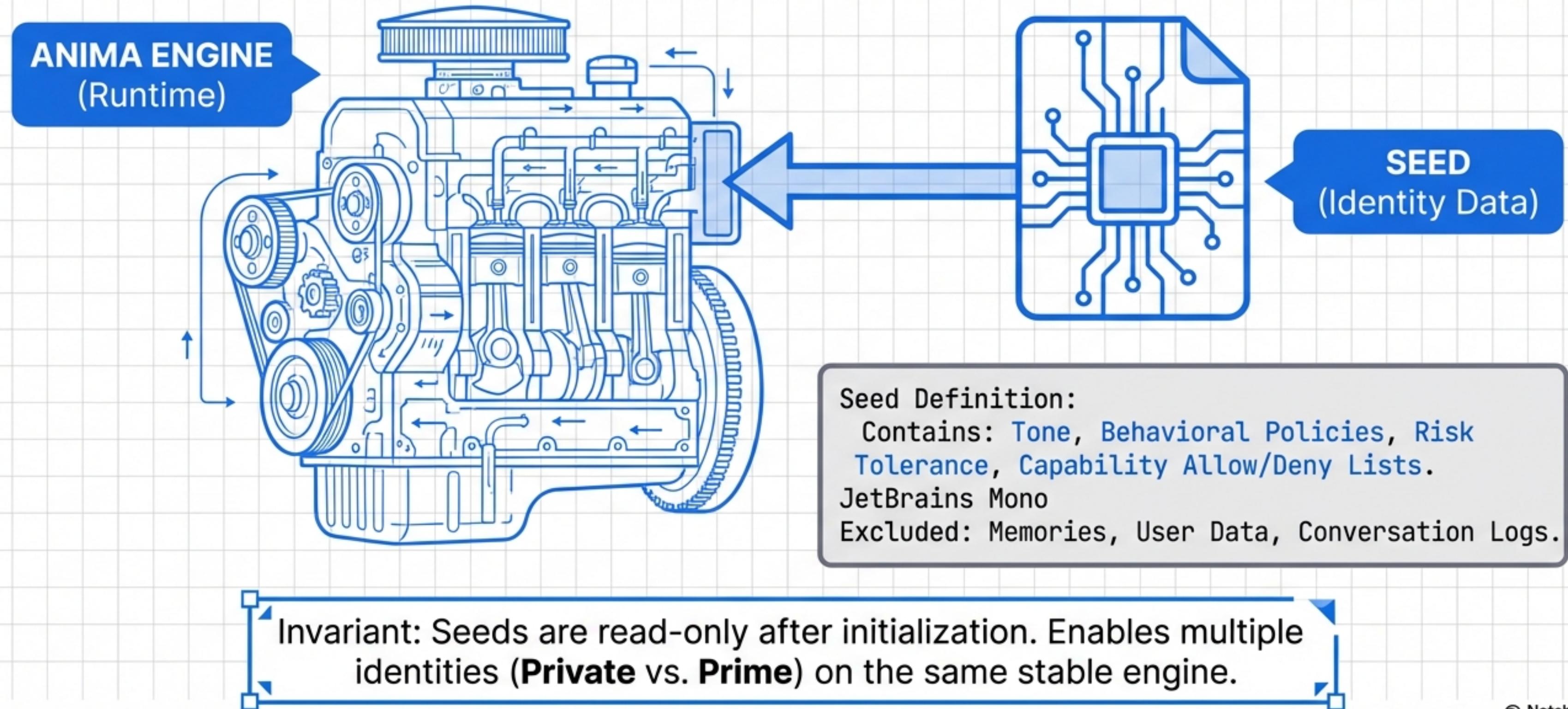
Why Not Just Build Another Agent?



“ANIMA is not designed to feel intelligent at all costs. She is designed to feel consistent, honest, and safe to grow with.”

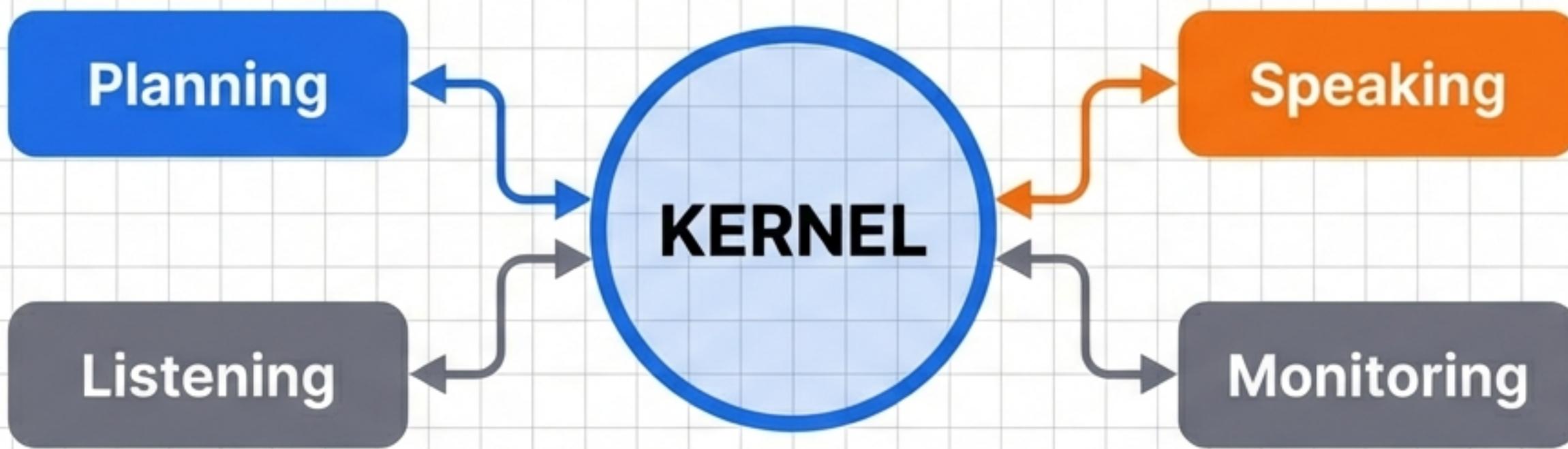
Principle 1: Engine ≠ Identity

The Engine is an agnostic runtime. Personality is data, not code.



The Core is a Supervisor, Not a Worker

ANIMA operates as a Cognitive Kernel, managing concurrent spans.



Multitasking: Manages concurrent “spans” rather than blocking on a single task.

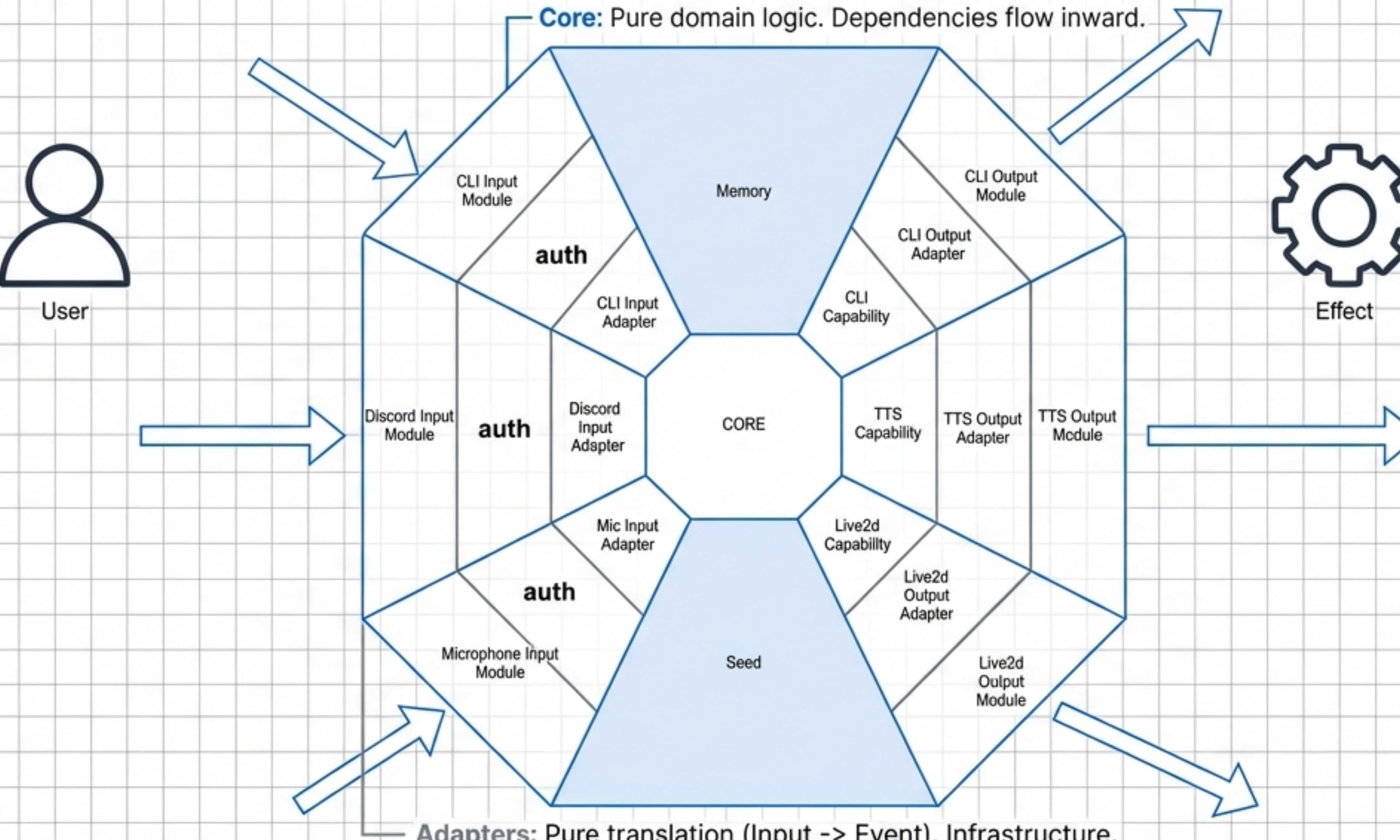
Cooperative Interruption: Tasks are interruptible by design.

Focus Management: Maintains a “foreground focus span” to prevent cascade cancellations.

“ANIMA does not execute instructions.
It supervises behavior.”

Hexagonal Architecture & Domain Purity

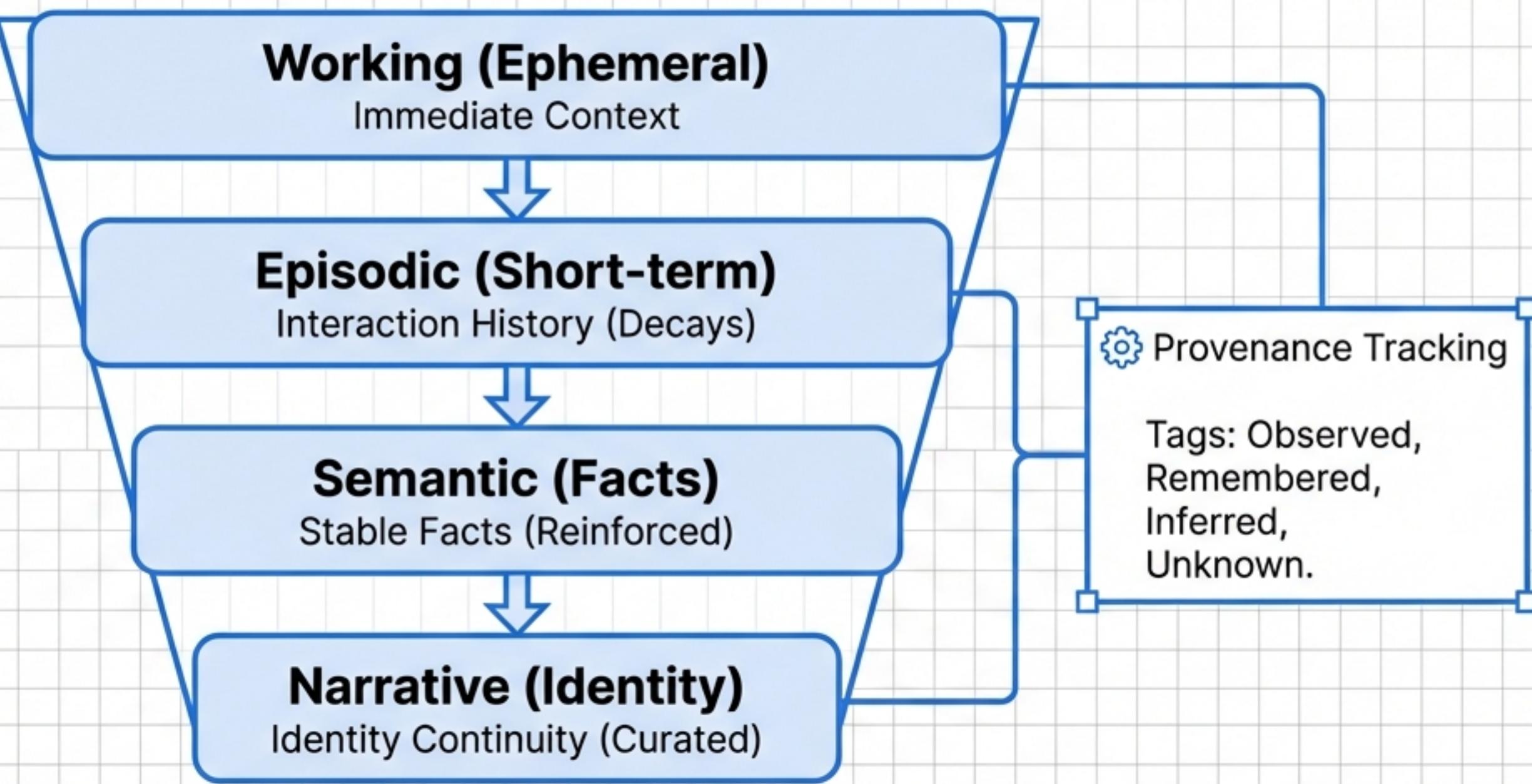
Domains must never talk directly to the outside world.



The Law: Dependencies always flow inward. Core never imports infrastructure.

Memory Integrity: Beyond Simple RAG

The Medial Temporal Lobe (MTL) Domain.



Constraint: Memory is strictly instance-scoped. No cross-instance sharing.

Model Topology: Cortex vs. Arcuate

Cognition is essential. Language is a capability.

CORTEX (Mandatory)

The Cognitive Model.

Responsibilities: **Reasoning, Planning, Safety Checks.**

Access: **Controlled Memory Slices.**

Intents



Semantic Events



ARCUATE (Optional)

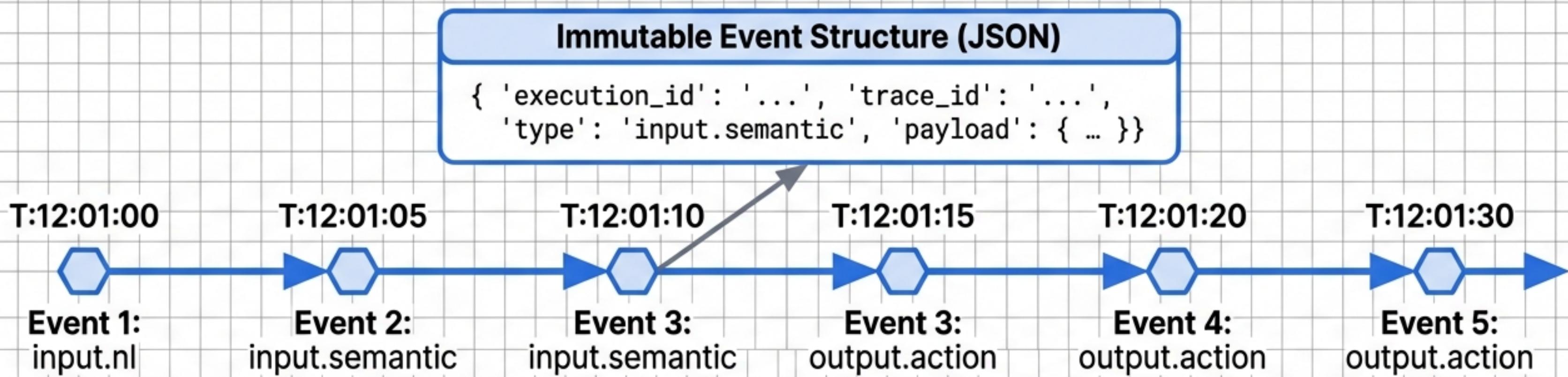
The NLP Layer.

Responsibilities:
Translate Raw Text -> Semantic Events, Intents -> Speech.

Safety Benefit: Prevents memory leakage through language processing.
Arcuate never accesses memory directly.

Observability: Recording Facts, Not Logs

Traditional logging is mutable. ANIMA records immutable structural events.



Canonical Input Events

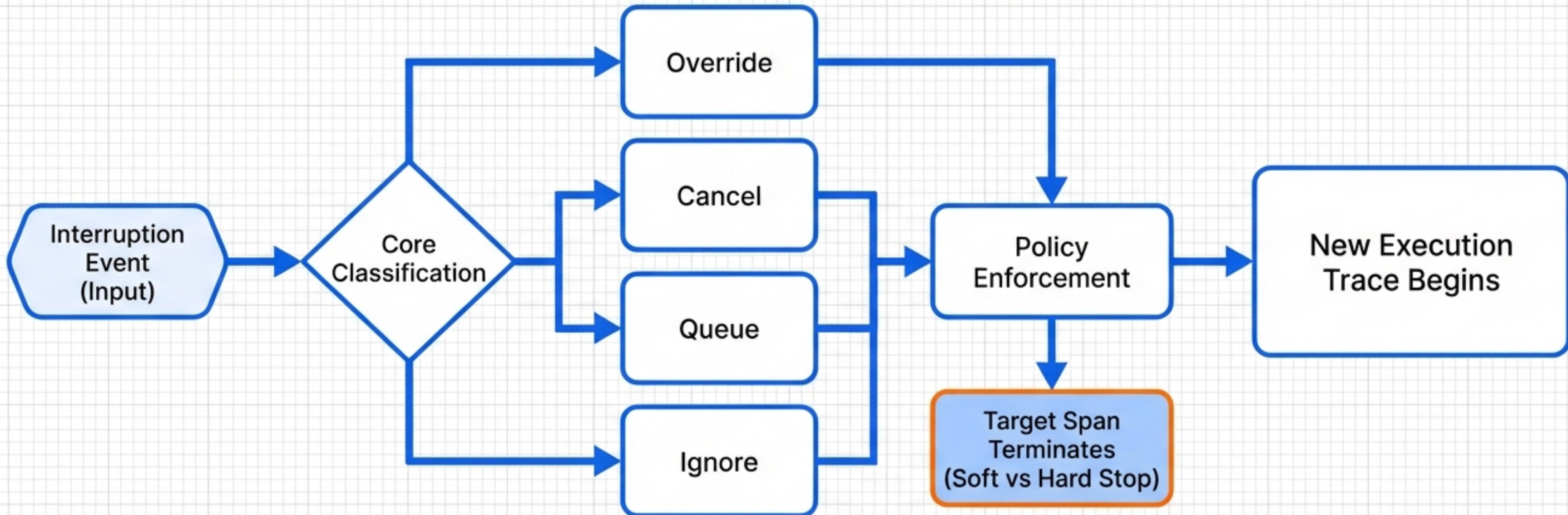
- **input.nl**: Raw natural language (**Low trust**)
- **input.system**: Mechanical facts (**High trust**)
- **input.semantic**: Post-interpretation asserted meaning (**Reasoning ready**)

Result: Deterministic replayability and audit-grade transparency.

Design Note: All recorded events are flat structures, no nested mutable state.

Managing Chaos: Interruptions as First-Class Events

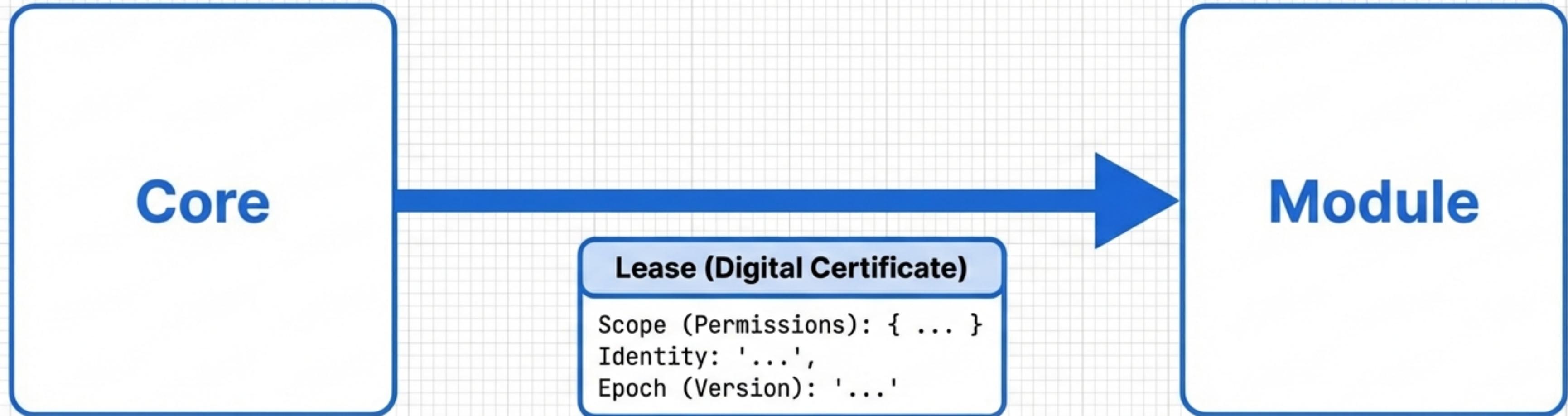
Real-world interaction is messy. Interruptions are structured inputs.



Key Scenario: Speaking While Speaking:
Handled deterministically via spans, preventing race conditions.

Security Model: Cryptographic Leases

Replacing API Keys with Lease-Based Authorization over gRPC + mTLS.

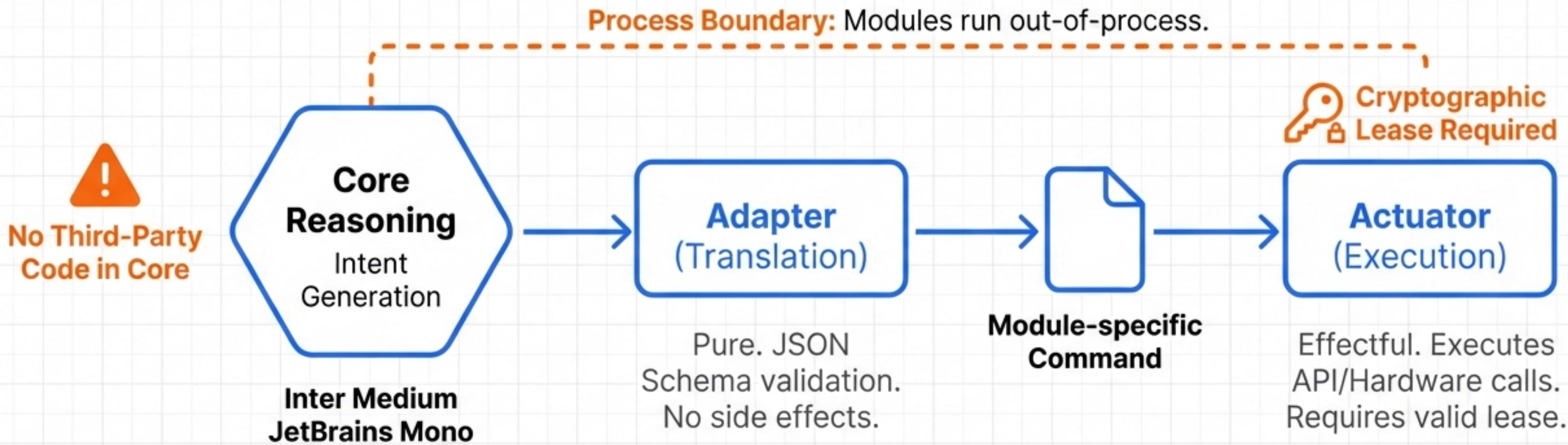


Zero-Lease State: No Lease, No Execution.
A module without a valid lease is inert.

Epochs prevent replay attacks and enforce instant revocation.

Module Architecture: Adapter-Actuator Split

Isolating translation from side effects.

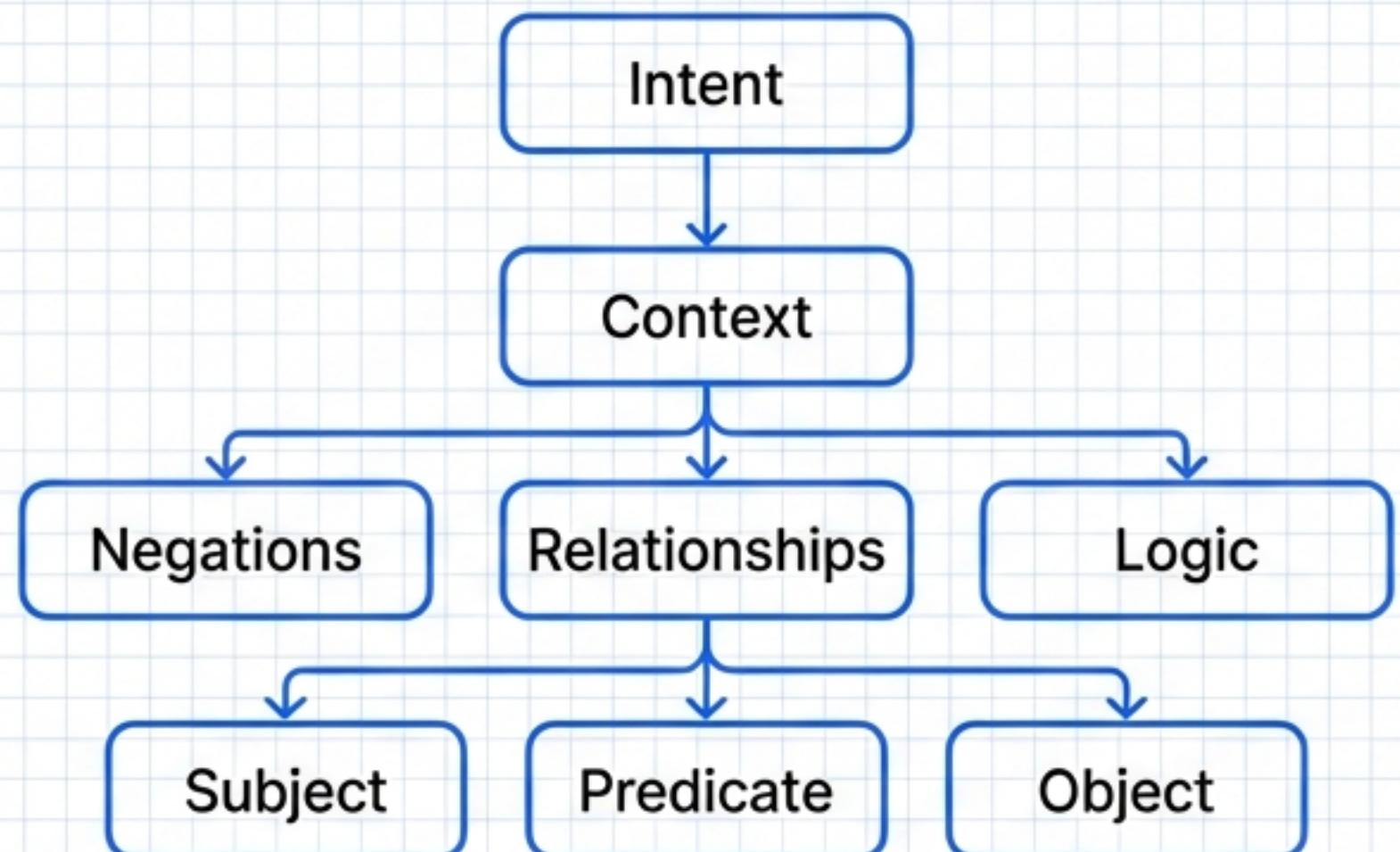


- **Adapter (Pure)**: Translation only. JSON Schema validation. No side effects.
- **Actuator (Effectful)**: Executes API/Hardware calls. Requires valid lease.
- **Process Boundary**: Modules run out-of-process.

Explicit Semantics: Spines & JSRS

Embeddings are lossy. Reasoning requires explicit structure.

Semantic Spines



Explicit representations of intent and context.
Allows validation of logic (negations, relationships).

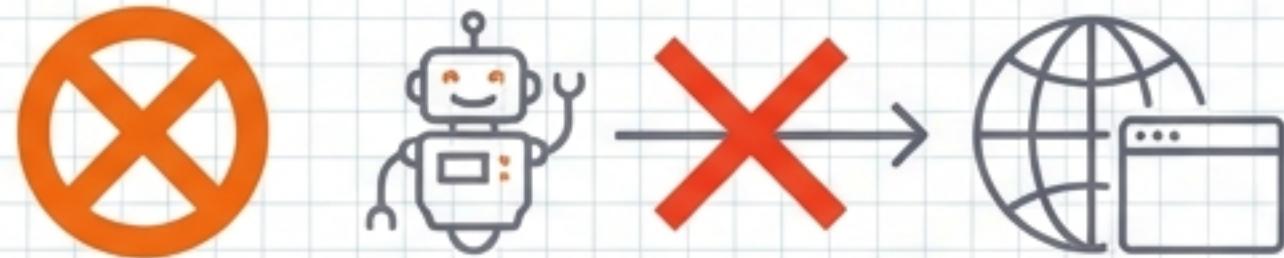
JSRS (JSON Scoped Reference System)

```
{ "$schema": "$root/schemas/module.json",  
  "config": {  
    "base_path": "$here/data",  
    "resources": {  
      "users": {  
        "ref": "$root/config/users.json"  
      }  
    }  
  },  
  "data": {  
    "items": ["item1", "item2"],  
    "local_ref": "$here/data/items[0]"  
  }  
}
```

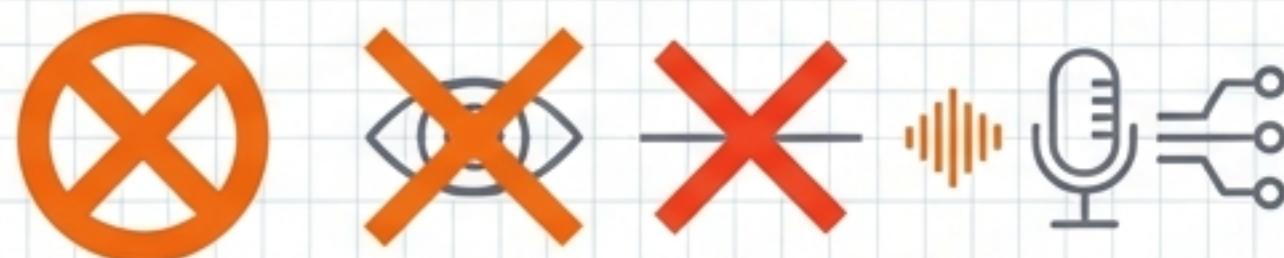
Relative navigation within JSON. Allows modular composition without breaking references.

Non-Goals & Ethical Boundaries

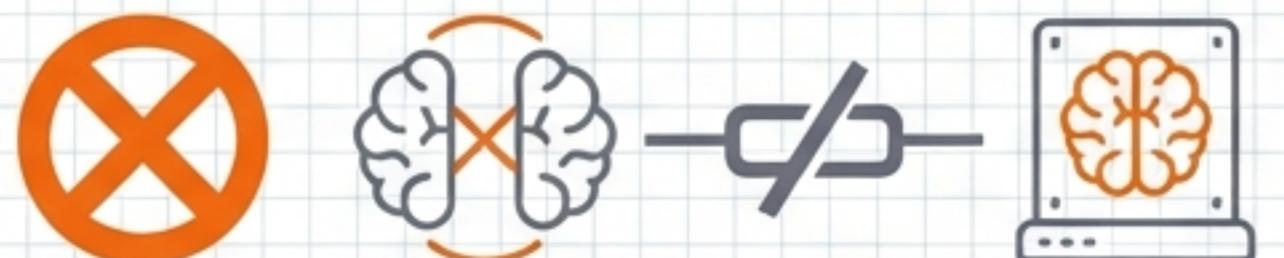
What ANIMA explicitly refuses to do.



No Autonomous Internet Agents: No unsupervised browsing or crawling.
// No open-ended web interaction.



No Mass Surveillance: Passive monitoring is architecturally disabled.
// Monitoring is opt-in, not architectural.



No Shared Memory: Hive-mind behavior is impossible; memory is local to the instance.
// State is isolated by design.

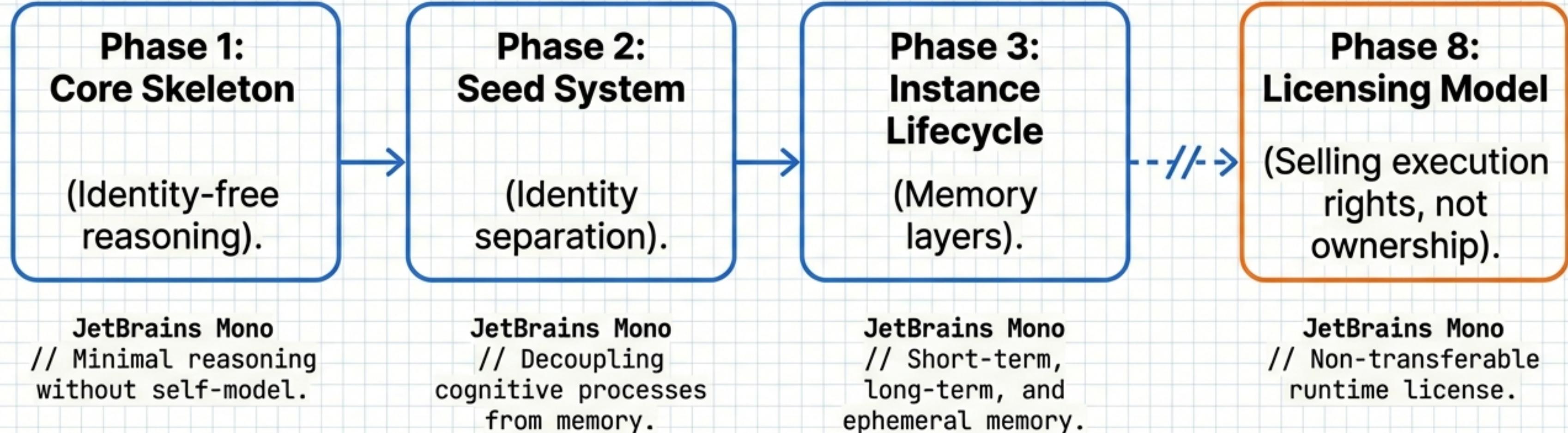


No General Intelligence Claims: It is a system, not a being.
// No self-awareness or consciousness.

"Some things are intentionally left undone to protect safety and trust."

Viability & Roadmap: The Path Forward

Strategy: Minimum Viable Cognitive Kernel.

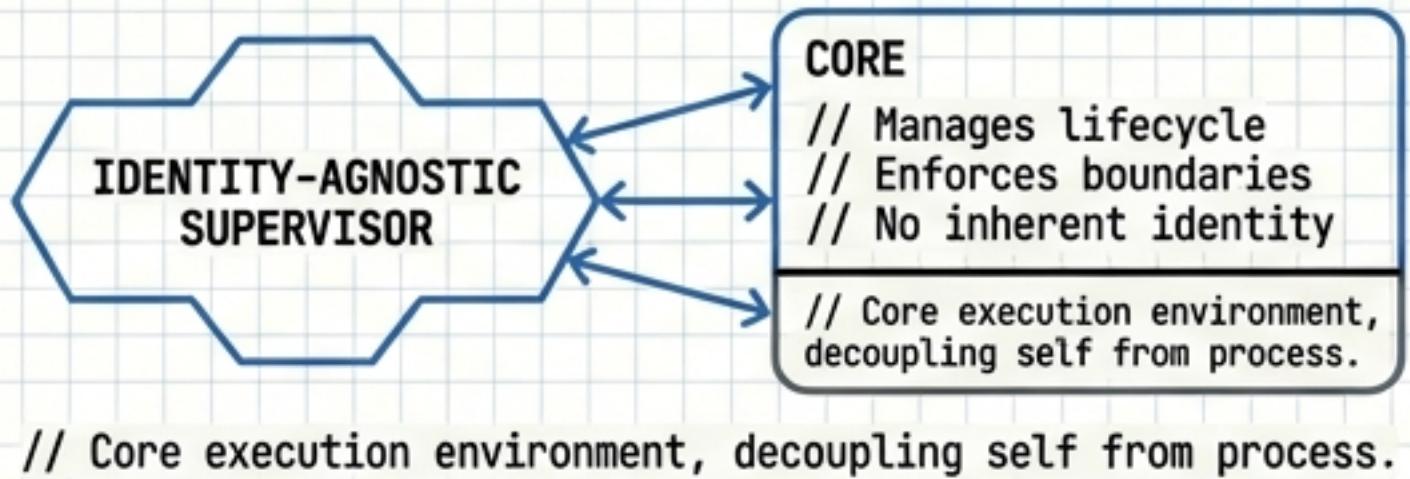


Is this too ambitious? Yes, for a monolith. No, for a modular architecture.

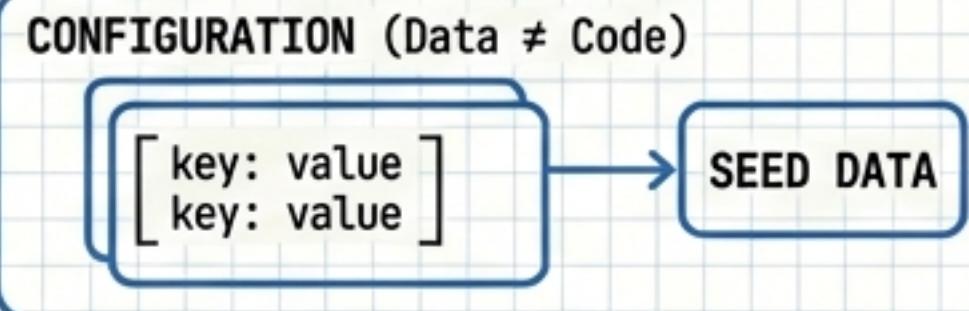
The Definition of ANIMA

An engine for growing identities, designed to be safe enough to trust with the long term.

ENGINE: Identity-agnostic supervisor.

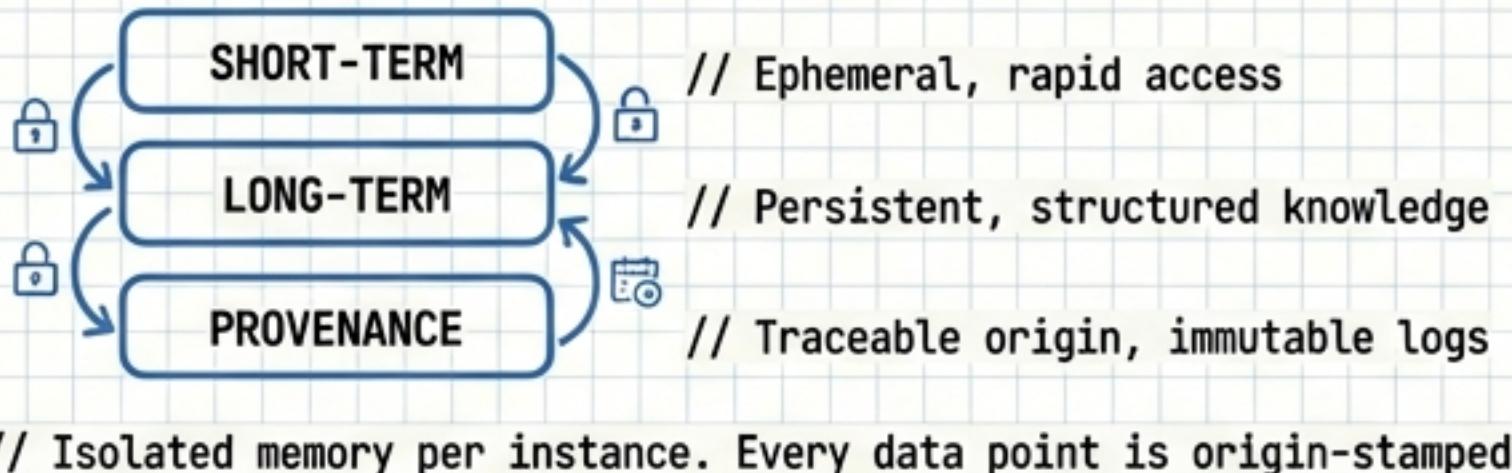


SEED: Identity configuration (Data ≠ Code).



// Immutable configuration for identity initialization.
Pure data, no executable logic.

MEMORY: Instance-local, layered, and provenance-tracked.



SECURITY: Lease-gated, out-of-process execution.



// Temporal execution rights. Processes run in strict isolation,
preventing unauthorized access.

ANIMA does not execute instructions. It supervises behavior.