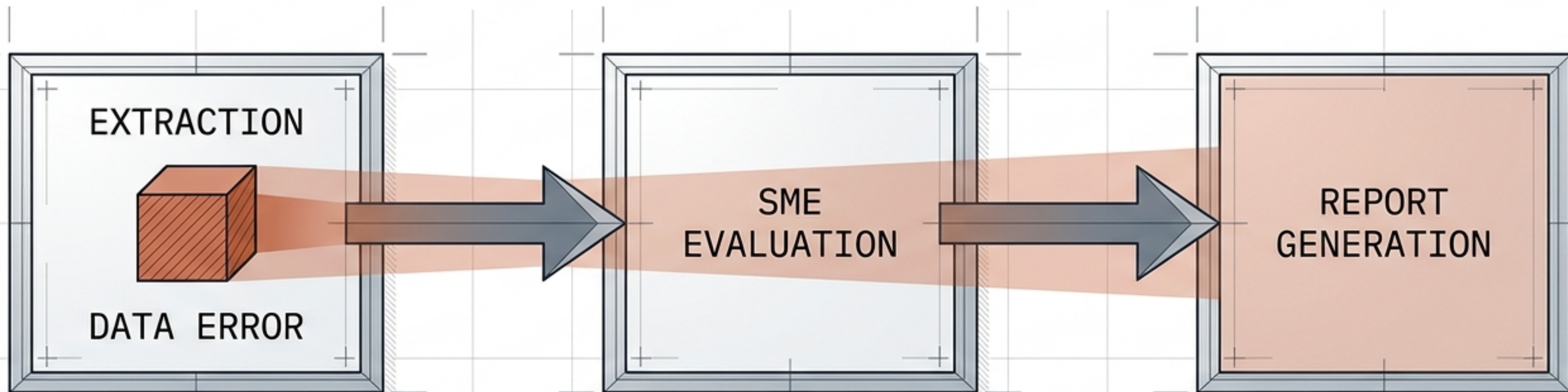




# THE DOMINO EFFECT OF LINEAR PIPELINES



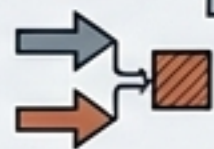
## NO QUALITY GATE:

Garbage in, garbage out. A bad extraction propagates silently.



## NO CROSS-SME REASONING:

Legal says PASS, Risk says FAIL. Outputs are concatenated, not resolved.



## NO EVIDENCE TRACKING:

Findings lack direct links to document pages or data snippets.



## BLIND TRUST:

Every LLM claim is taken at face value without tool verification.

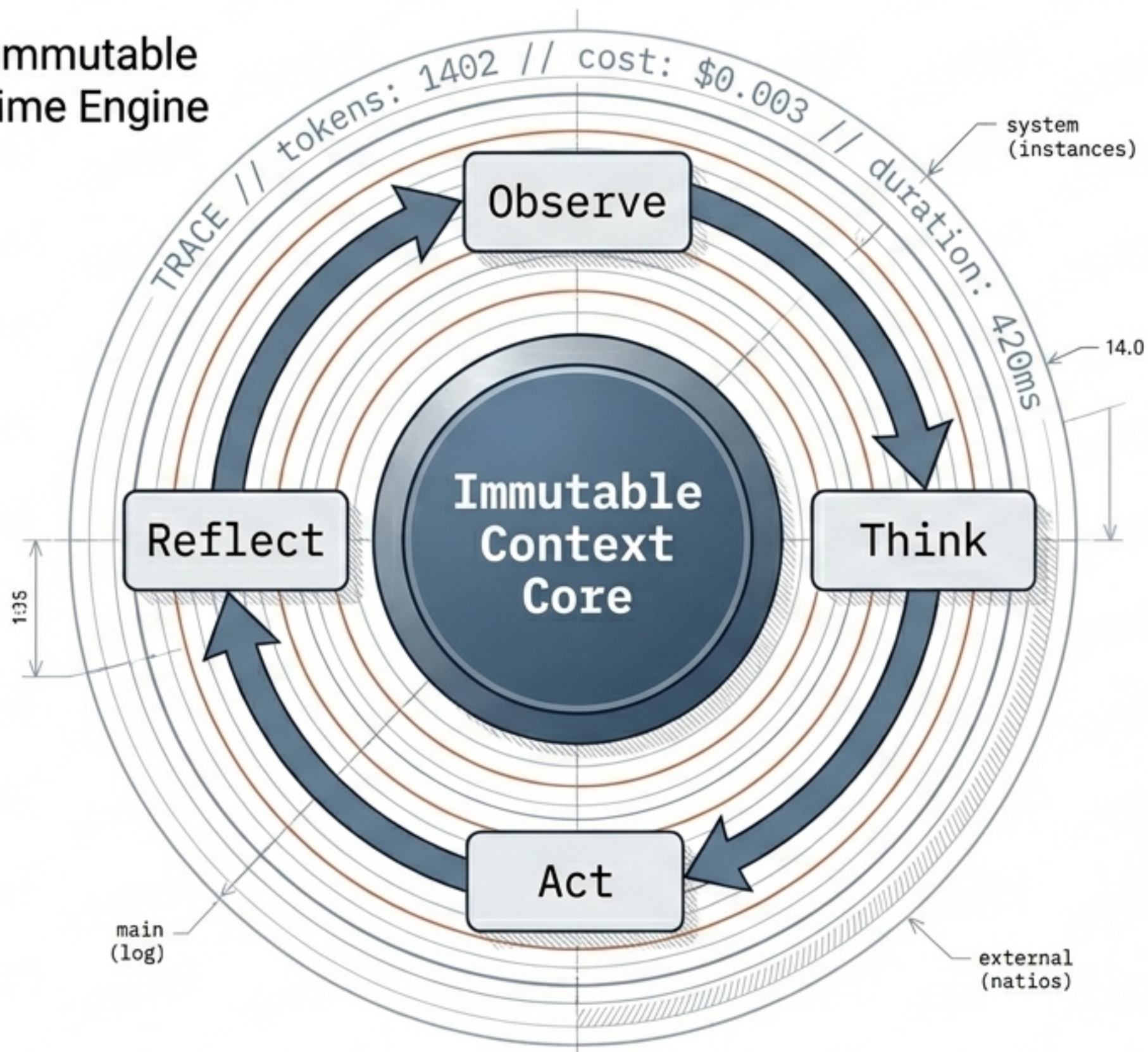


# THE ARCHITECTURAL PARADIGM SHIFT

| Dimension       | V2 (Linear Pipeline)    | V3 (Agentic System)               |
|-----------------|-------------------------|-----------------------------------|
| Architecture    | Pipeline                | Parliament                        |
| Execution       | Extract / Evaluate Once | Observe & Reflect Loops           |
| Truth-Seeking   | Blind Trust             | Fact-Checked via External Tools   |
| Disagreement    | Concatenate Opinions    | Structured, Evidence-Based Debate |
| Quality Control | None (Garbage In/Out)   | Quantitative Scoring Thresholds   |

# The AgentRuntime Execution Loop

The Immutable Runtime Engine



## 1. Agent-Agnostic

Any BaseAgent plugs in. The runtime executes the loop seamlessly, regardless of the agent's role.



## 2. Context is Immutable

with\_tool\_result() and with\_llm\_response() return new instances. Zero aliasing bugs or surprise mutations.



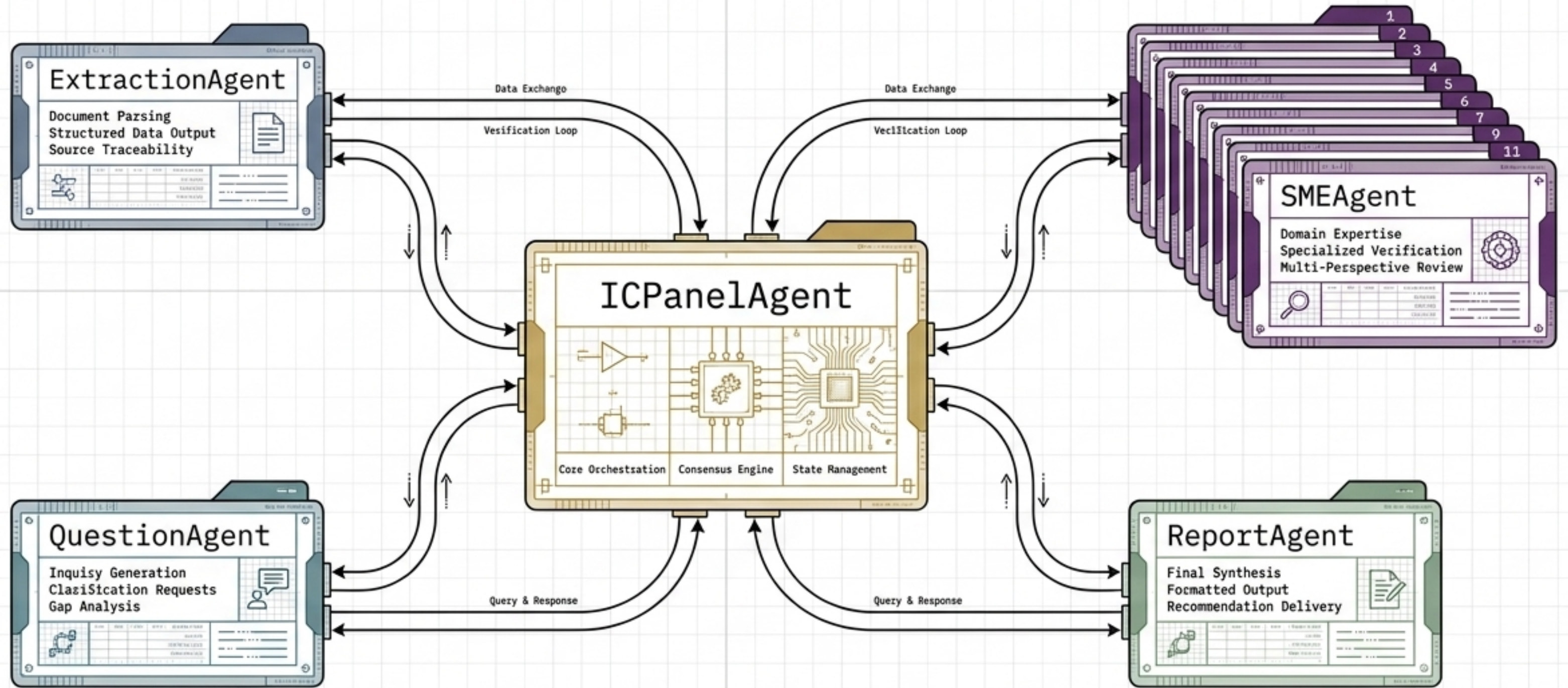
## 3. Traced Execution

Full token count, duration, and tool I/O recorded in AgentTrace for perfect auditability.

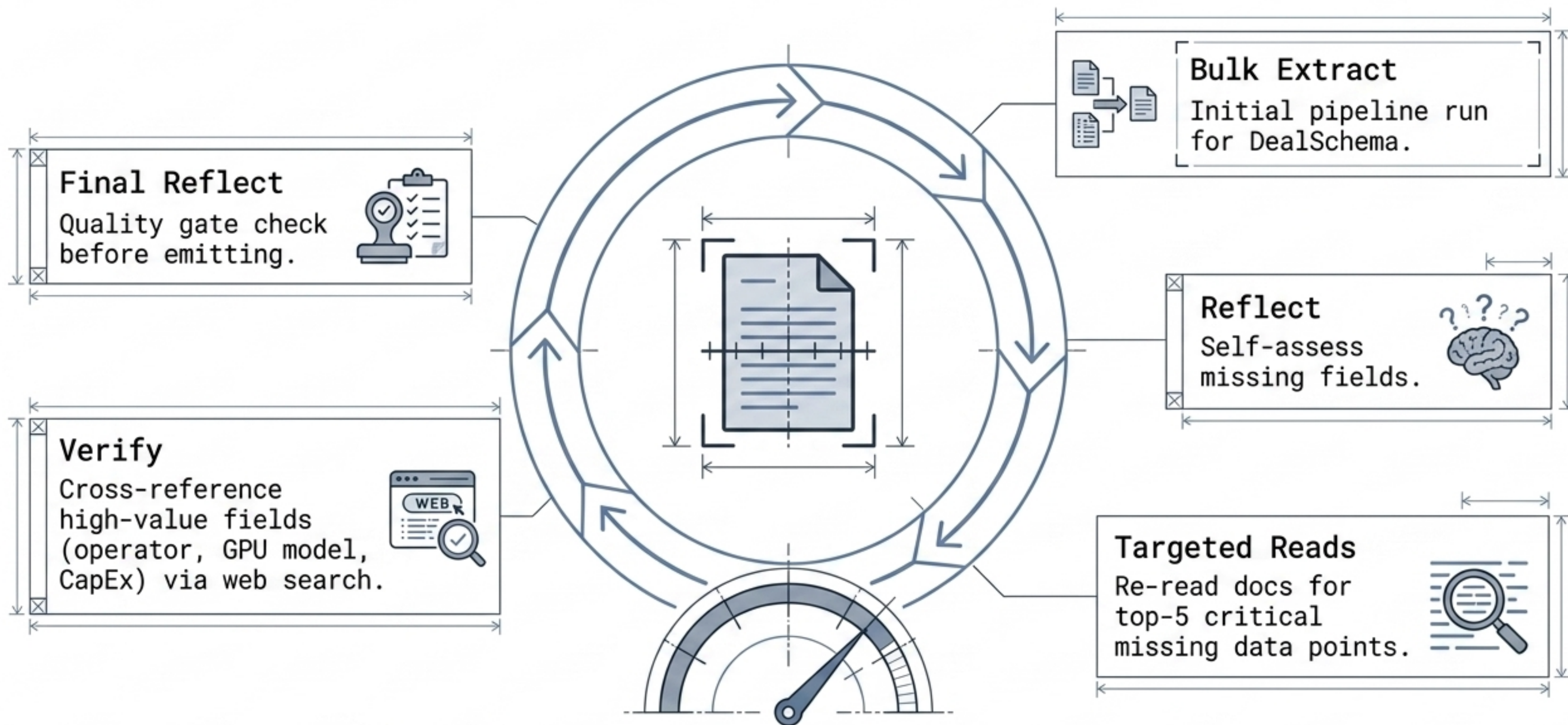


# The 5 Agents That Underwrite a Deal

Autonomous state machines that think, verify, and argue.



# Extraction Agent: Iterate Until Complete

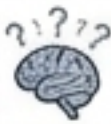


Model: gemini-3.1-pro | Threshold: 80/100

Quality Gate: 80/100 threshold


# SME Agent: 11 Domain Experts Showing Their Work

 Initial Evaluation

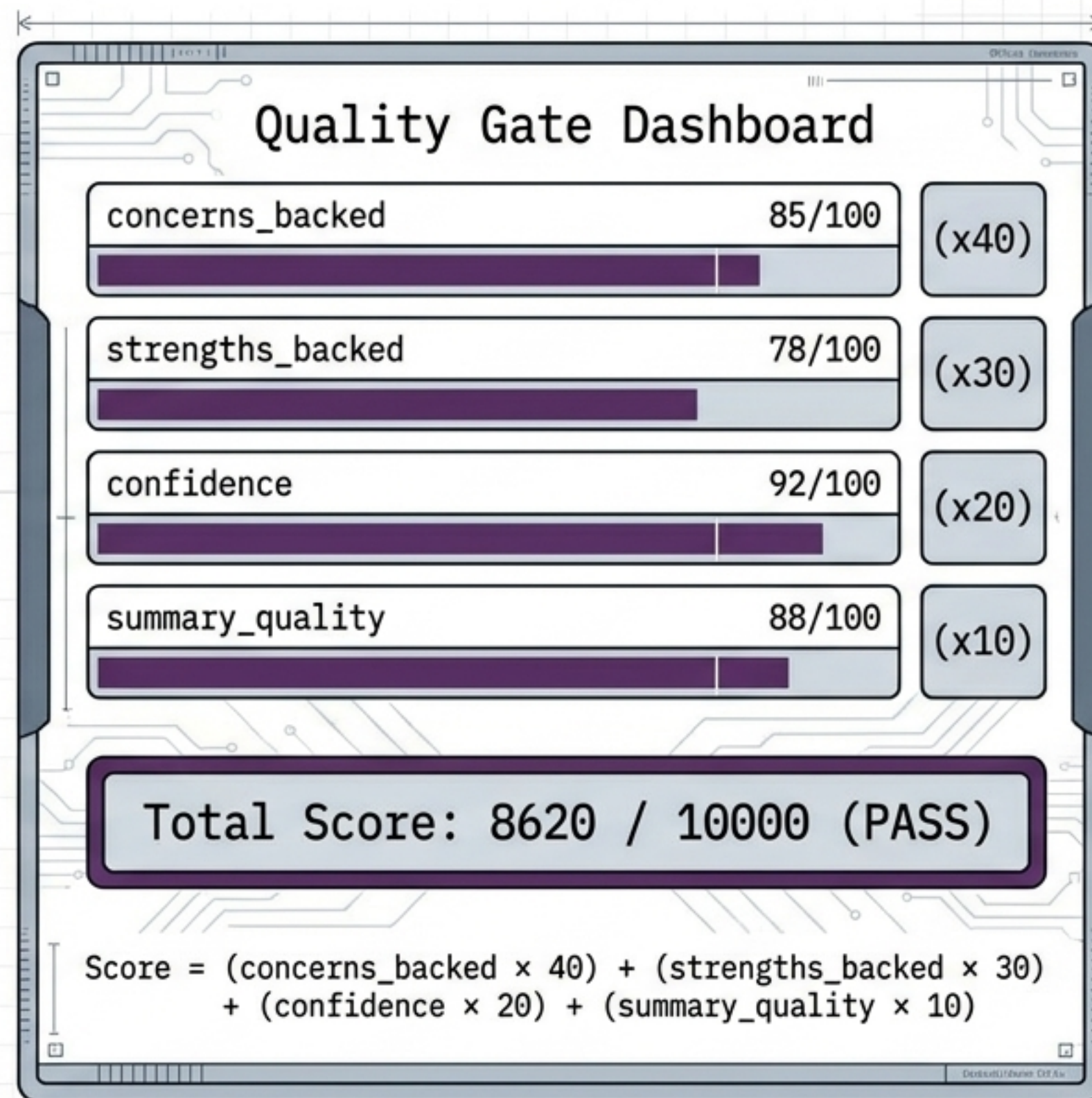
 Reflect

 Tool Verification

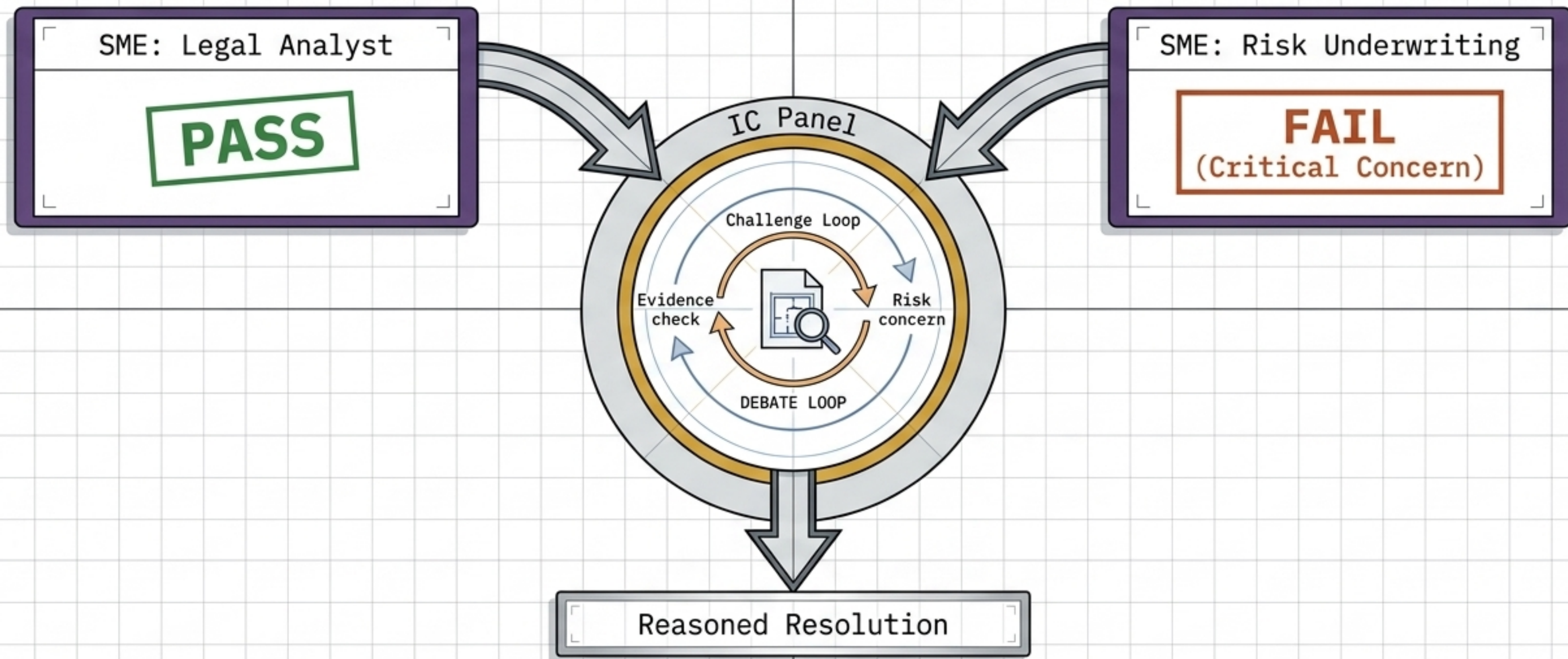
```
query_schema,  
calculate
```

 Final Reflect

- Operates across 11 lanes (e.g., gpu\_engineer, legal\_analyst, risk\_underwriting).
- Outputs distinct decisions: PASS, PASS\_WITH\_CONDITIONS, NEEDS\_CLARIFICATION, FAIL.

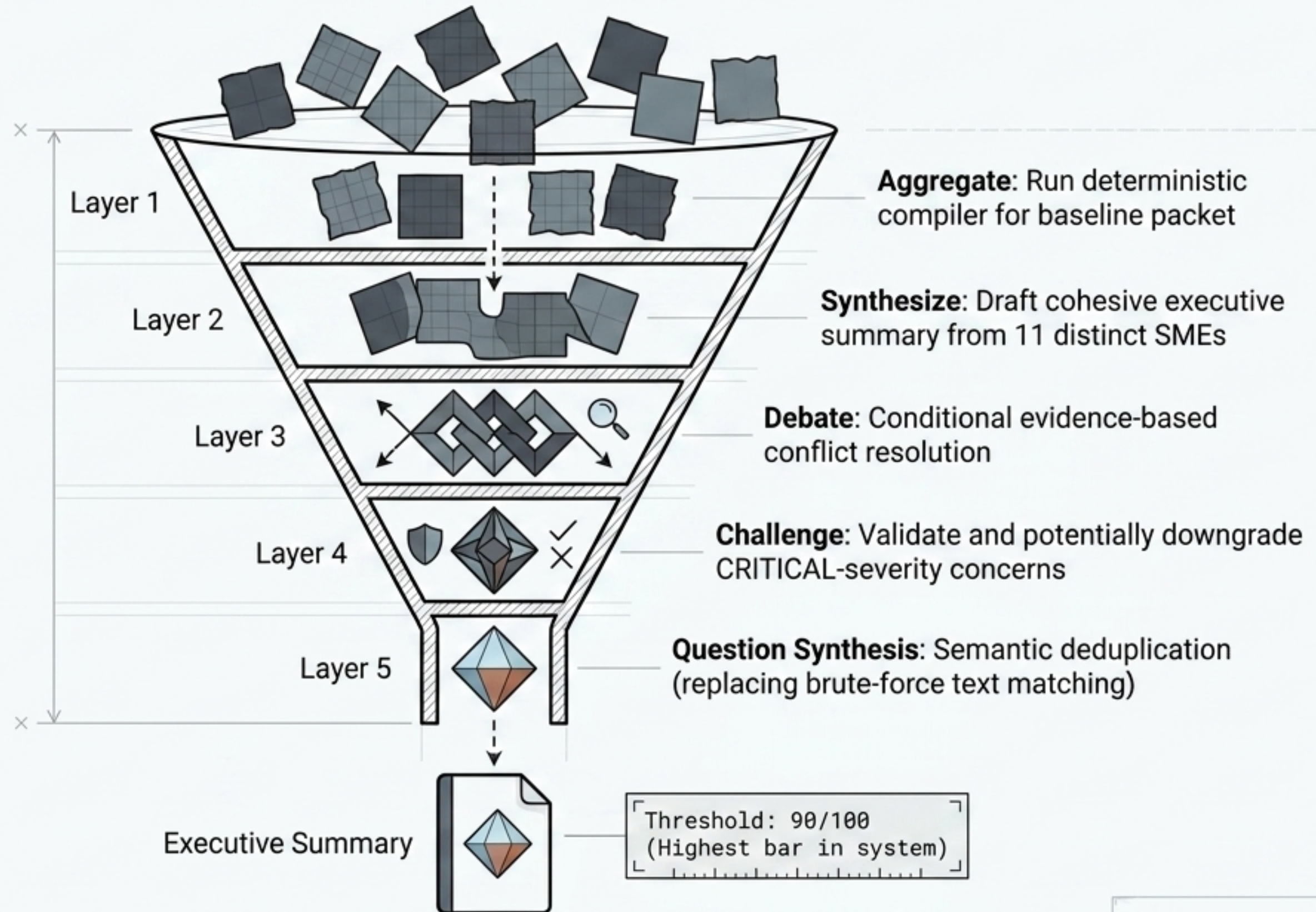


# IC Panel Agent: The Structural Debate

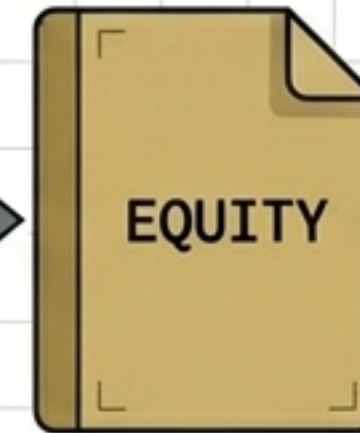
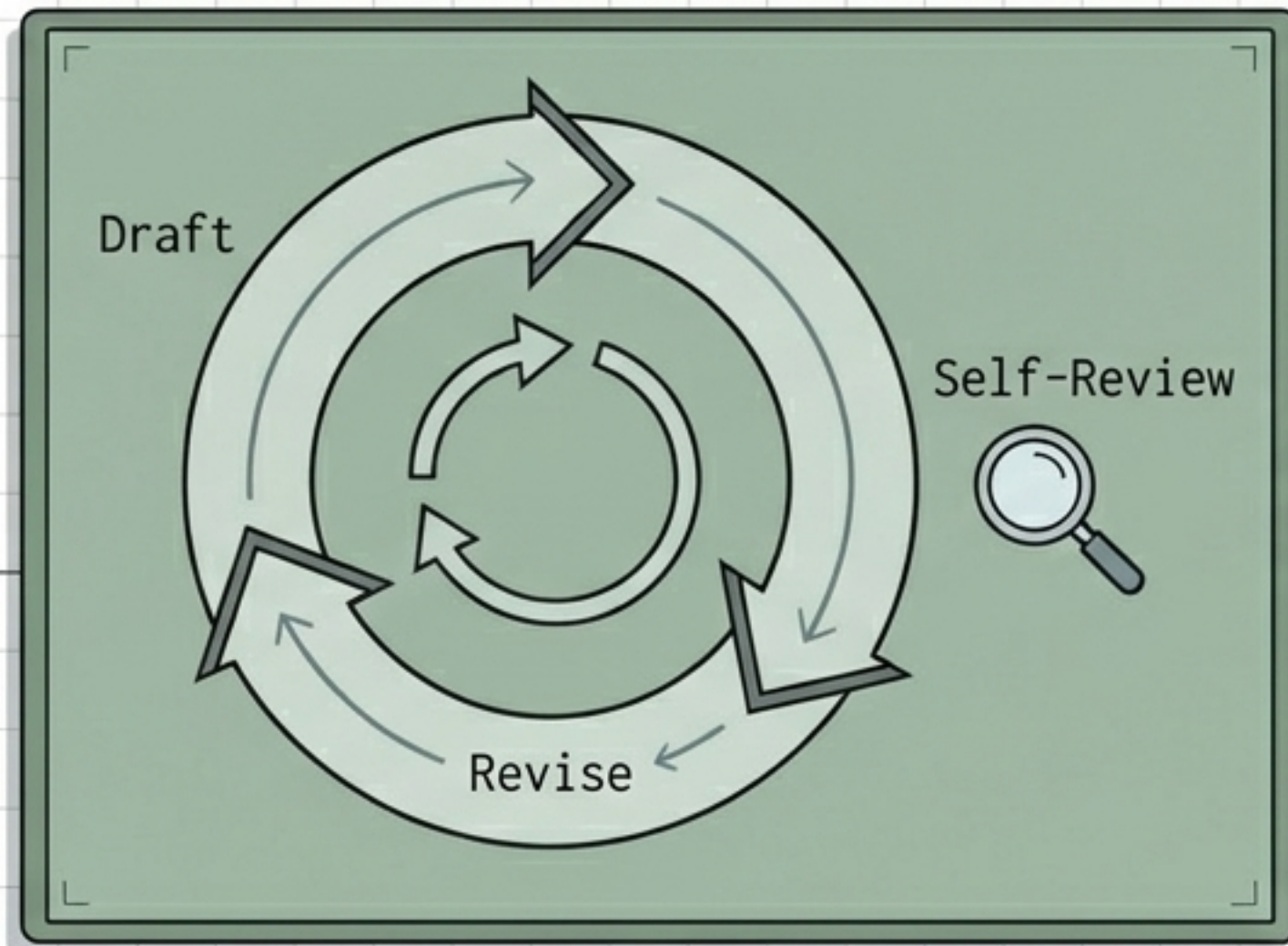


When SMEs disagree, the IC Panel triggers a conditional debate. It forces evidence from both sides. If evidence is weak or mitigated by another SME, a CRITICAL concern is formally downgraded.

# The Synthesis Funnel



# Report Agent: Draft, Critique, Revise



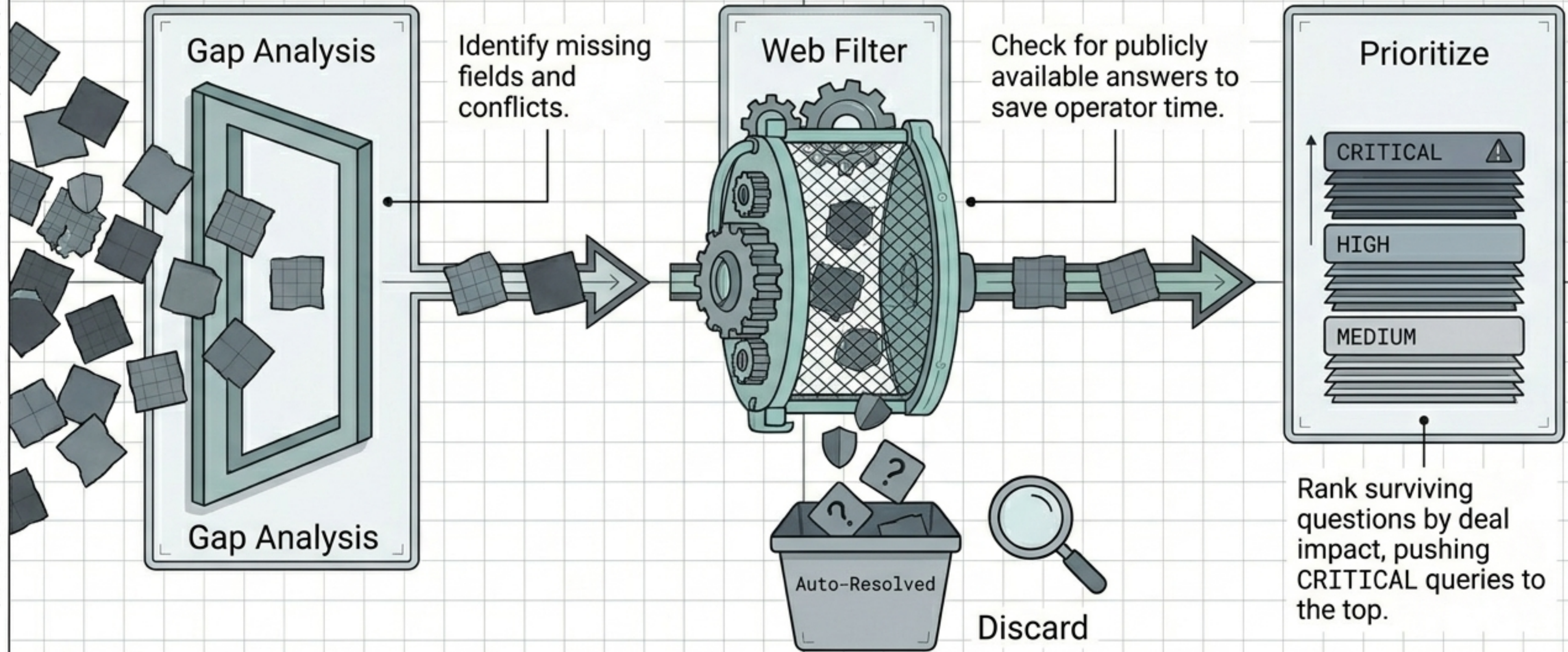
Focuses heavily on IRR and upside scenarios.



Focuses on credit rationale and covenant analysis.

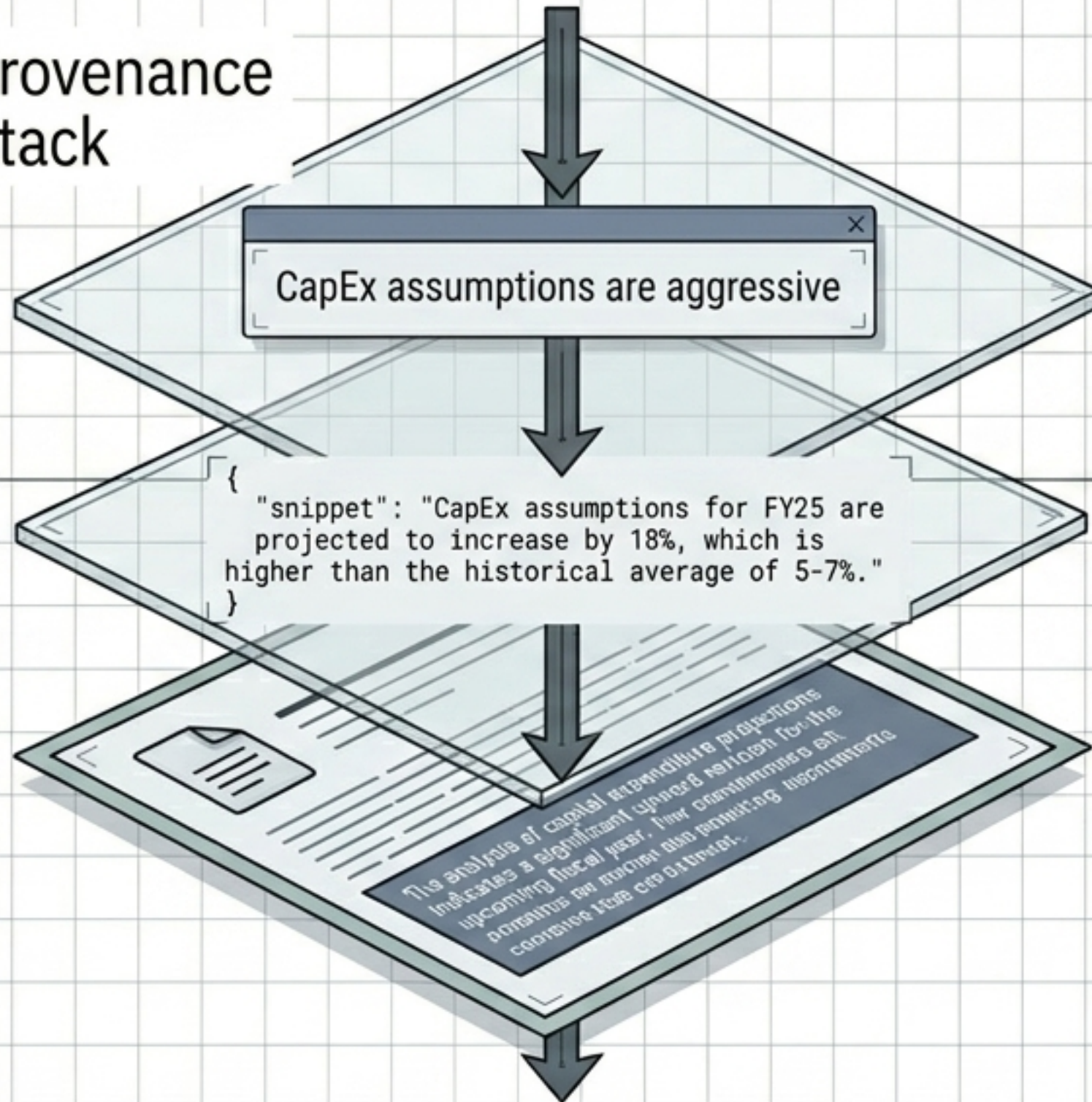
Reports are no longer one-shot generations. The model reviews its own draft to ensure concerns are addressed, numbers are sourced, and tone aligns with the audience.

# Question Agent: Ask Only What You Can't Google



# Evidence Provenance

## Provenance Stack



## Core Principle

Every finding traces its exact lineage.  
No unverified hallucinations.

## Trace Architecture

1. Source Document File
2. Specific Document Location / Page
3. Exact Text Snippet

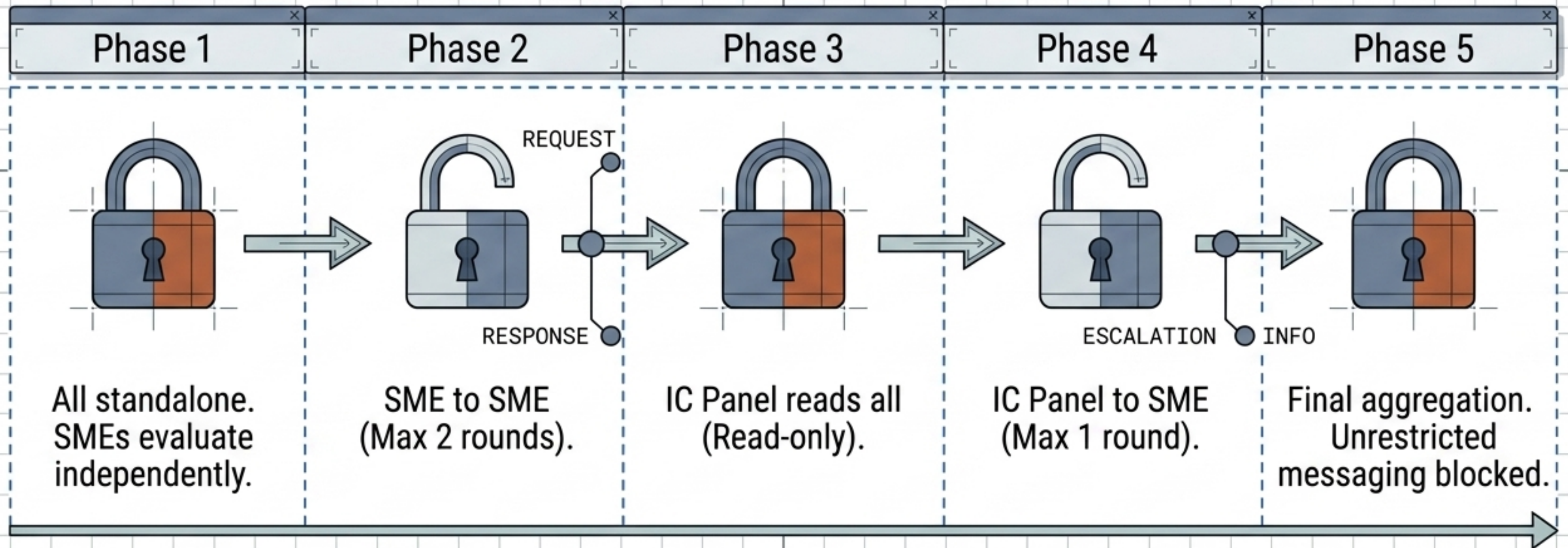
## Severity Matrix



Upgraded from binary (pro/con) to a 4-level scale: **CRITICAL / HIGH / MEDIUM / LOW**, paired with generated mitigations.

# Preventing Chaos: Phased Communication

Unrestricted messaging creates infinite loops. The V3 bus uses a strict phased firewall.  
Message types allowed: REQUEST, RESPONSE, CHALLENGE, ESCALATION, INFO.



# Execution Metrics: Arkane Cloud Production Data

| Command Line Interface |        |          |
|------------------------|--------|----------|
| Agent                  | Tokens | Duration |
| -----                  |        |          |
| SME (gpu_engineer)     | 75K    | ~5 min   |
| IC Panel               | 20K    | ~30 sec  |
| Report (equity)        | 95K    | ~3 min   |
| Question Agent         | 21K    | ~1 min   |
| -----                  |        |          |
| Total (single SME)     | ~211K  | ~10 min  |

Full pipeline with all 11 SMEs: Just a few dollars per deal.

Robustness: 63 tests (58 mock/integration, 5 end-to-end API) written in a single Claude Code session.

# Intelligence is a System Property.

The underwriting pipeline didn't get smarter by swapping in a larger LLM. It got smarter because the architecture granted it the human-like abilities to question itself, demand evidence, and debate conflicts.

Observe → Think → Act → Reflect