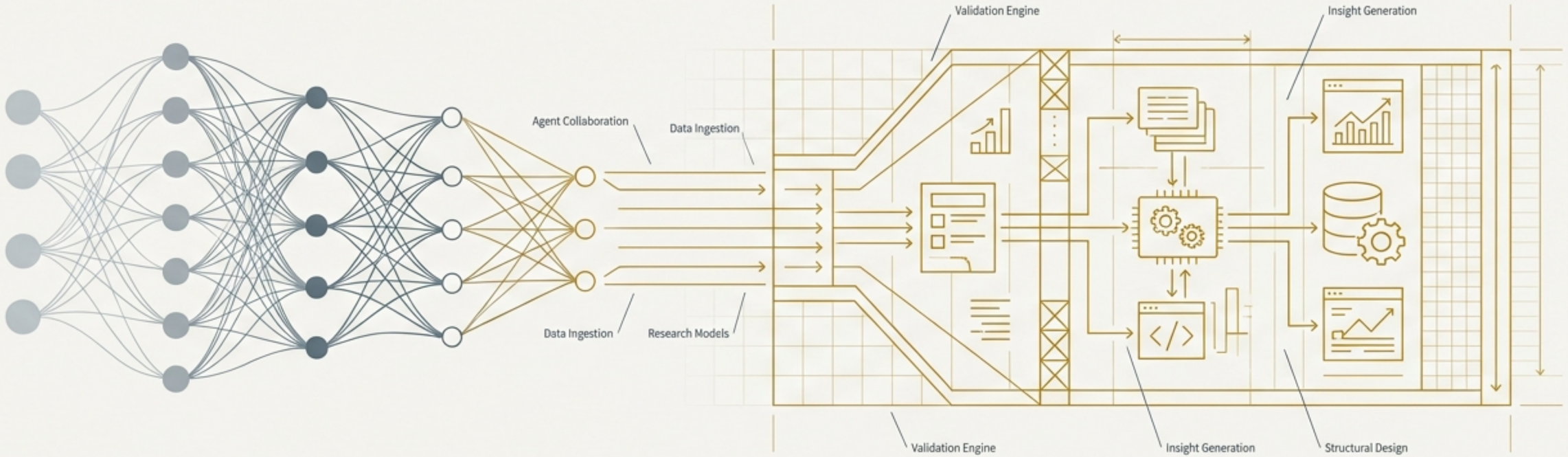
Architecting for Insight

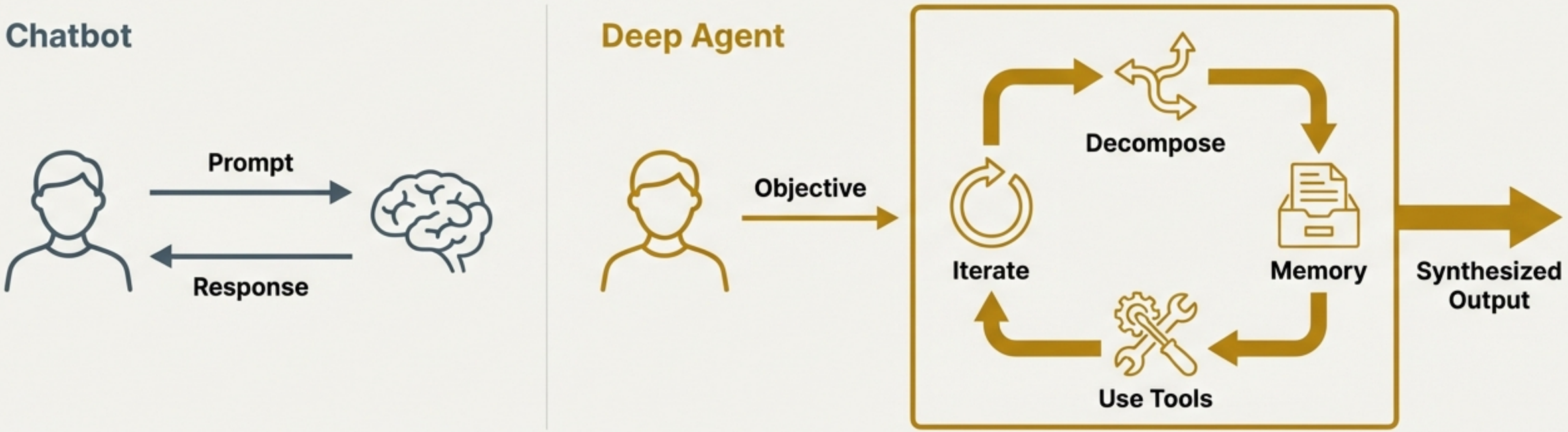
A Blueprint for Custom Multi-Agent Financial Research



The Industry is Moving Beyond Chatbots to ‘Deep Agents’

We define Deep Agents as long-horizon AI systems that function like junior analysts, not just conversationalists. Unlike simple chat, they are designed to:

- Decompose complex problems into sub-tasks.
- Maintain intermediate state and memory.
- Use tools and files explicitly.
- Iterate, refine, and self-correct their work.



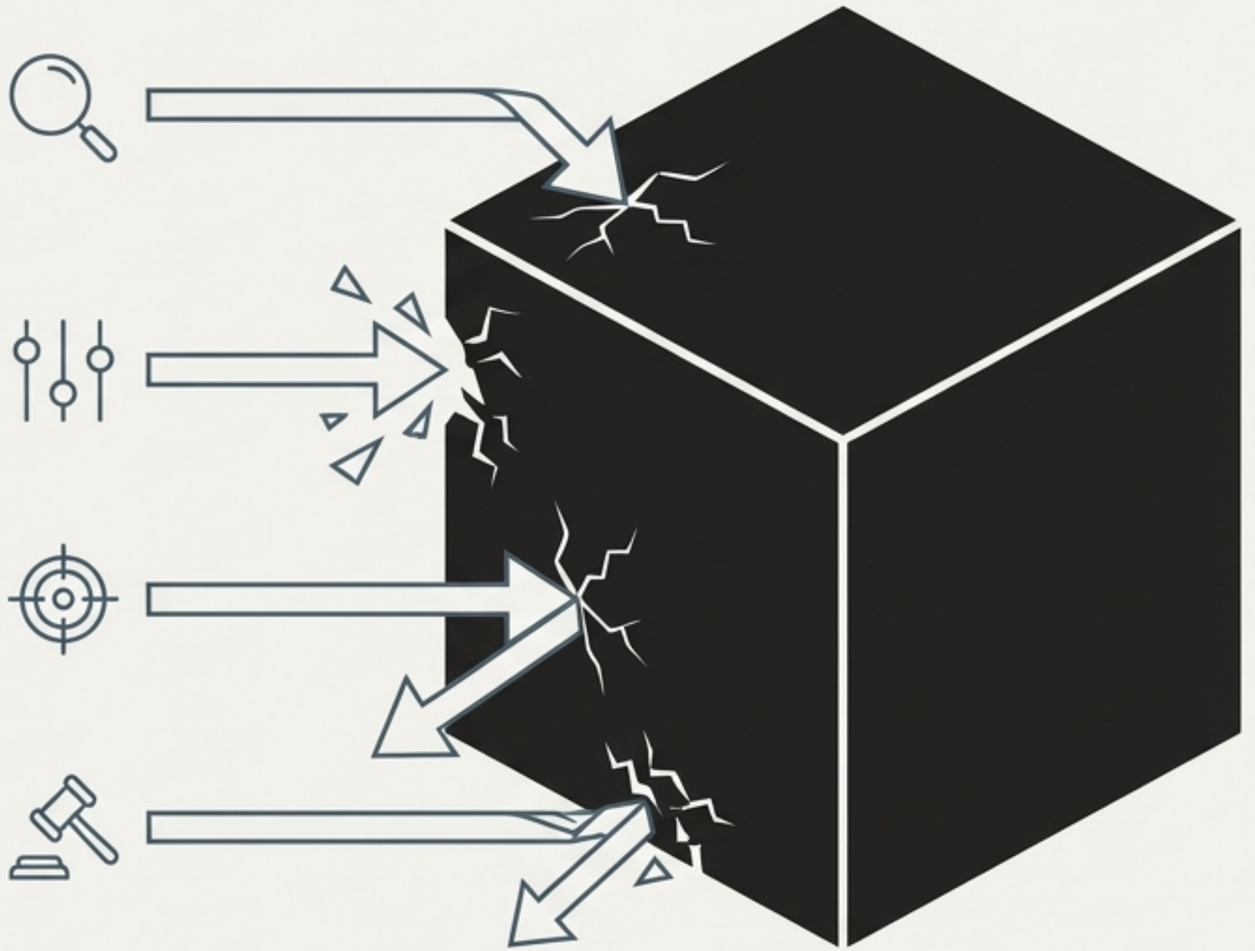
Prebuilt Deep Agent Systems Validate This Shift

Category	Example Systems	Core Focus
Framework-based	LangChain Deep Agents	Tool-driven multi-step workflows
Autonomous Research	ChatGPT Deep Research	Long-form, web-heavy investigation
Coding-centric	Claude Code	Deep code understanding and refactoring

*The emergence of these systems signals a critical **shift** from **prompt engineering** to building agentic workflows.*

But Prebuilt Solutions Are Unsuitable Black Boxes for Finance

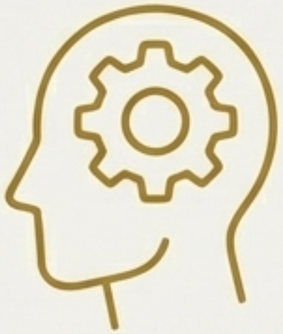
- **Opacity:** Internal orchestration and reasoning steps are hidden from view.
- **Lack of Control:** There is limited ability to influence memory, persistence, and tool usage.
- **Generic Nature:** They are difficult to optimize for specific financial research workflows and terminology.
- **Audit & Compliance Risk:** The ‘black box’ nature makes it impossible to produce a verifiable audit trail.



A Robust System Begins with a Clear Philosophy

1. LLMs are for Reasoning, not Storage

Use models for their unique ability to reason, plan, and synthesize.



2. Files are for Memory, not Models

The filesystem is the only reliable source of truth for long-term state.

3. Agents are Specialized, not General-Purpose

Assign discrete, single-purpose roles for more reliable and predictable behavior.

4. Coordination is Explicit, not Implicit

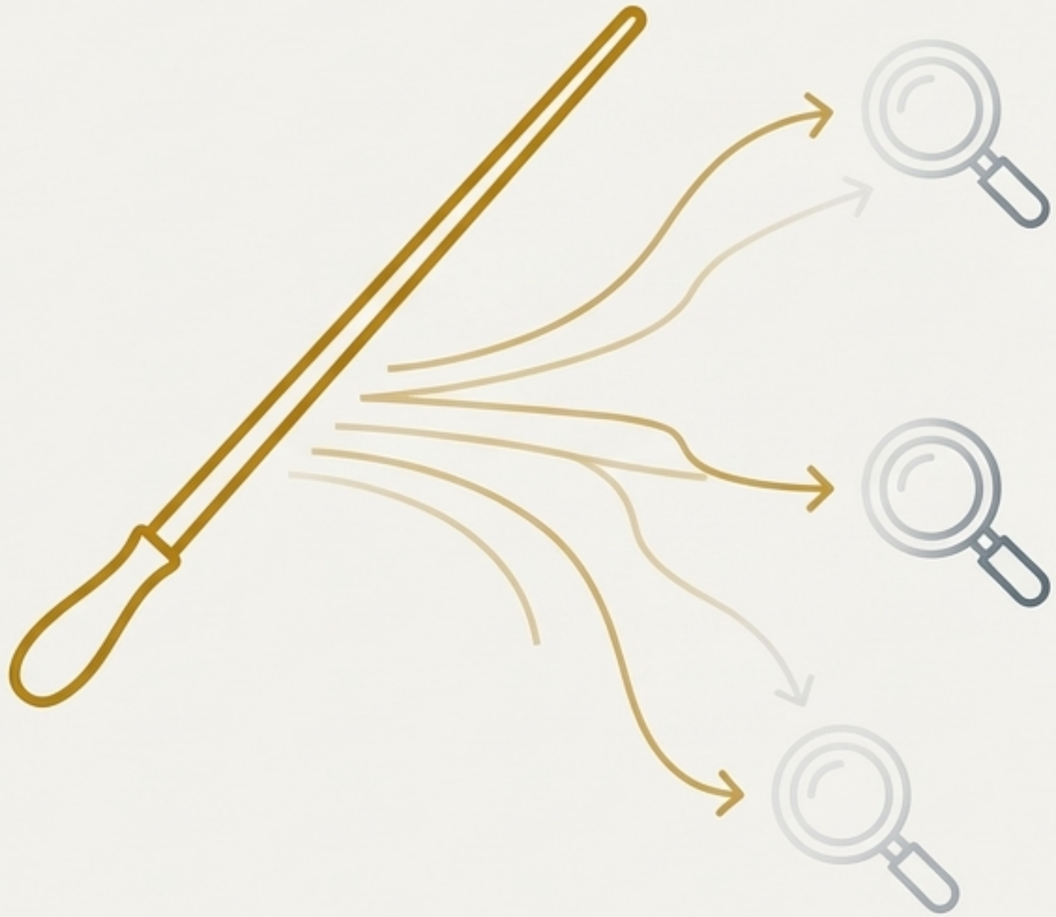
Workflow is managed by a dedicated orchestrator, not emergent behavior.

Our Architecture Mimics an Expert Research Team

We implement a hierarchical, multi-agent system where responsibilities are clearly separated, just like in a real-world team.

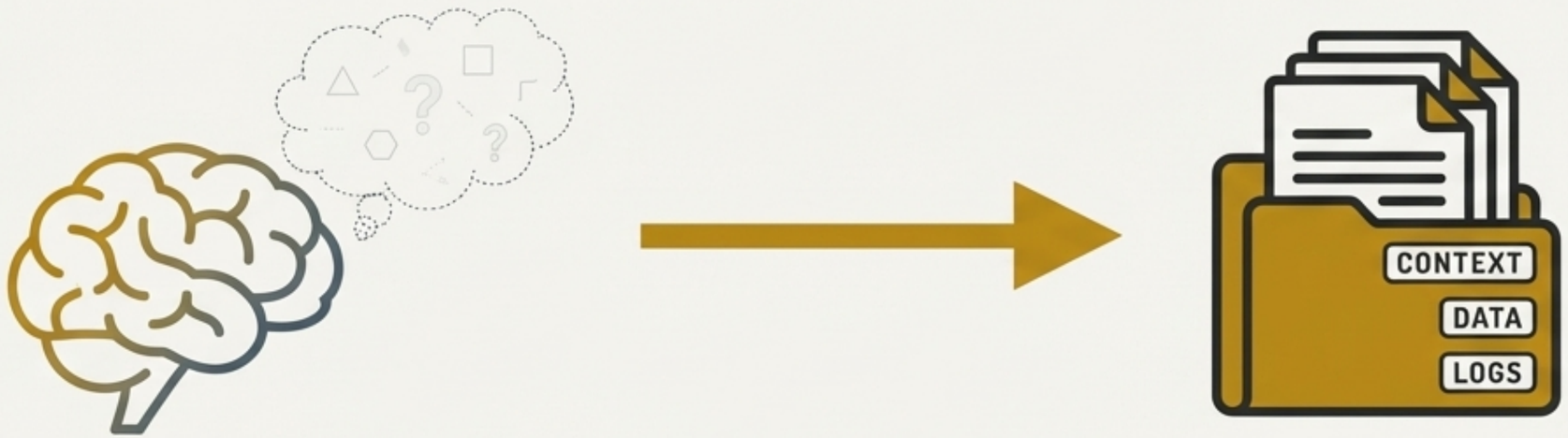
The Orchestrator Plans and Supervises, It Does No Research Itself

- Interprets the user's initial query.
- Decomposes complex requests into thematic research questions.
- Assigns each theme to a dedicated Researcher agent.
- Manages workflow, handling retries, and triggering the final synthesis.



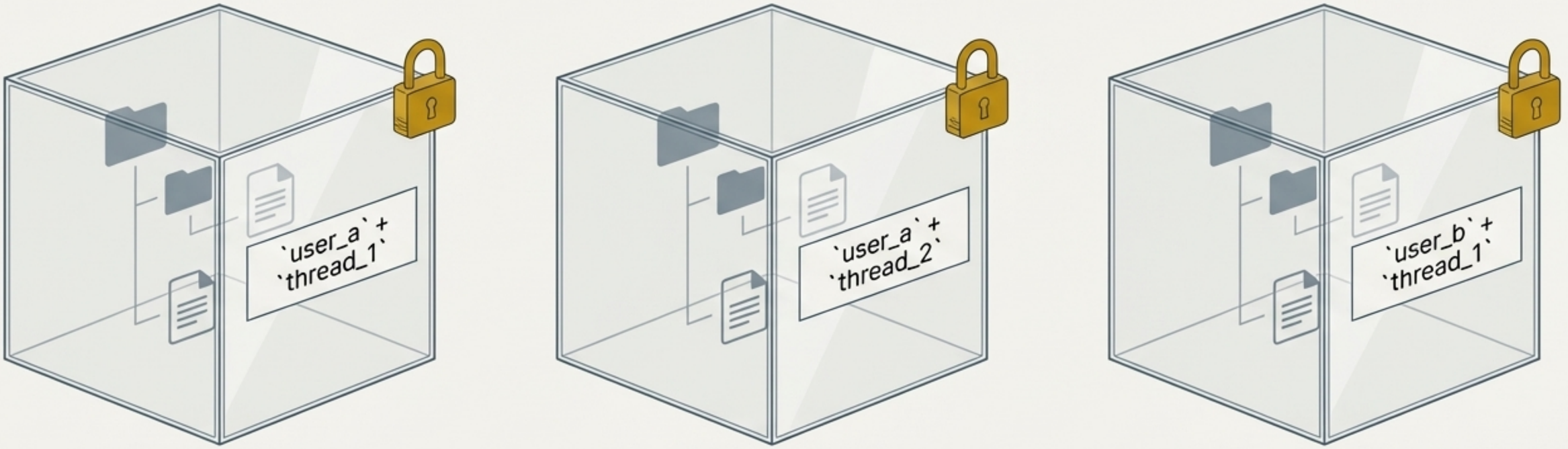
Agents think, but files remember.

**Problem*: Context limits, truncation, and hallucination make in-memory approaches unreliable for deep research.
**Solution*: We treat the filesystem as the system's durable, persistent memory. Every meaningful output is intentionally written to disk.
**Benefit*: This guarantees persistence, auditability, and the ability to reason over contexts far larger than any single model can handle.



Every Research Thread Gets Its Own Isolated Workspace

Zero Data Leakage: Information cannot cross between users or their distinct research threads.
Full Reproducibility: Any research task can be perfectly reproduced from its file-based artifacts.
Clean Resumption: Interrupted workflows can be cleanly resumed via checkpointing.



A Disciplined, Auditable Flow from Question to Report

1 Query

User submits a complex research question.

2 Plan

The Orchestrator creates a research plan with distinct thematic questions.

3 Execute

Researcher agents execute each theme, writing findings and sources to their workspace on disk.

4 Review

The Editor agent reads all artifacts from the workspace.

5 Synthesize

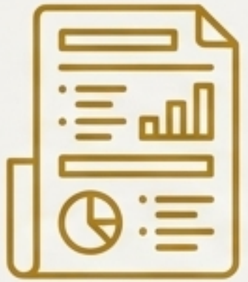
A final, synthesized report is generated with source citations.

An Architecture Designed for the Demands of Financial Analysis

Accuracy & Traceability: Enforced by requiring every finding to be written to a file with sources.

Long-Context Reasoning: Enabled by using the filesystem as an unlimited, durable memory.

Complete Auditability: Every step in the process—from planning to execution—leaves a digital paper trail.



Structured Outputs: The specialized Editor role ensures consistent, high-quality, and properly formatted reports.

The Advantage: Full Control, Full Transparency

Dimension	Custom Deep Agent (This System)	Prebuilt Deep Agent
Control	✅ Full control over planning, tools, and memory.	⚠️ Limited or opaque.
Explainability	✅ Fully auditable via file-based steps.	⚠️ Partial or hidden.
Domain Adaptation	✅ Designed specifically for financial workflows.	⚠️ Generic, one-size-fits-all.
Cost Management	✅ Predictable, tool-scoped LLM usage.	⚠️ Often opaque and unpredictable.
Reliability	✅ Deterministic, checkpointed workflows.	⚠️ Probabilistic, often stateless behavior.

A Fundamental Shift in Perspective

“Prebuilt deep agents show **"what is possible.”**

Building your own deep agent teaches **"why and how it works.”**

This is a Blueprint for a Production-Grade Deep Agent

This system demonstrates how to build a reliable, auditable, and domain-specific research tool, moving beyond the limitations of a black-box assistant.