GUIDE

Rethinking Observability for the Age of AI Agents

fiddler

Taming Multi-Agent Chaos

Single-agent systems are already complex. They handle varied inputs, mitigate hallucinations, integrate with external systems, and operate under strict performance constraints.

Multi-agent systems multiply that complexity not linearly, but exponentially. Each additional agent introduces more possible states, interactions, and failure modes. The state space grows so large that, without granular visibility, diagnosing issues becomes guesswork.

Ensuring reliability across the agentic hierarchy requires more than just building the agent. That's because ensuring reliability requires more than just building the agent; it requires monitoring every handoff, every intermediate decision, and every integration point in the workflow.

Each agent brings its own state, decision logic, and integration points. When they interact, the possible combinations of states and outcomes can grow beyond what's human-manageable without the right abstraction layers.

In these environments, **aggregated metrics and contextual drill-downs are not optional, they are the only way to make sense of the system's behavior.** Observability must enable teams to trace issues across agent boundaries, following the chain of interactions to locate the exact source of a problem.

Without this capability, troubleshooting devolves into guesswork, and confidence in deploying advanced agentic architectures suffers.

Aggregation and Context Will Define the Next Generation of AI Monitoring

Enterprises are racing to harness the potential of LLM-powered agents — systems that can reason, plan, and act autonomously. These agents are already transforming workflows in customer service, financial analysis, marketing, and research. The promise is clear: faster decisions, deeper automation, and entirely new capabilities.

But as these systems move from proof-of-concept to production, one truth becomes impossible to ignore: **you can't trust what you can't see.** As these systems grow more capable, they also grow more opaque. An agent can make dozens of tool calls, interact with external APIs, and even collaborate with other agents, all in a matter of seconds. Each decision, each data handoff, each model invocation is a potential point of failure.

In the traditional software world, observability means tracking logs, metrics, and traces to keep systems running smoothly. For AI agents, methods built for static applications and predictable pipelines, aren't enough. They capture what happened, but not why.

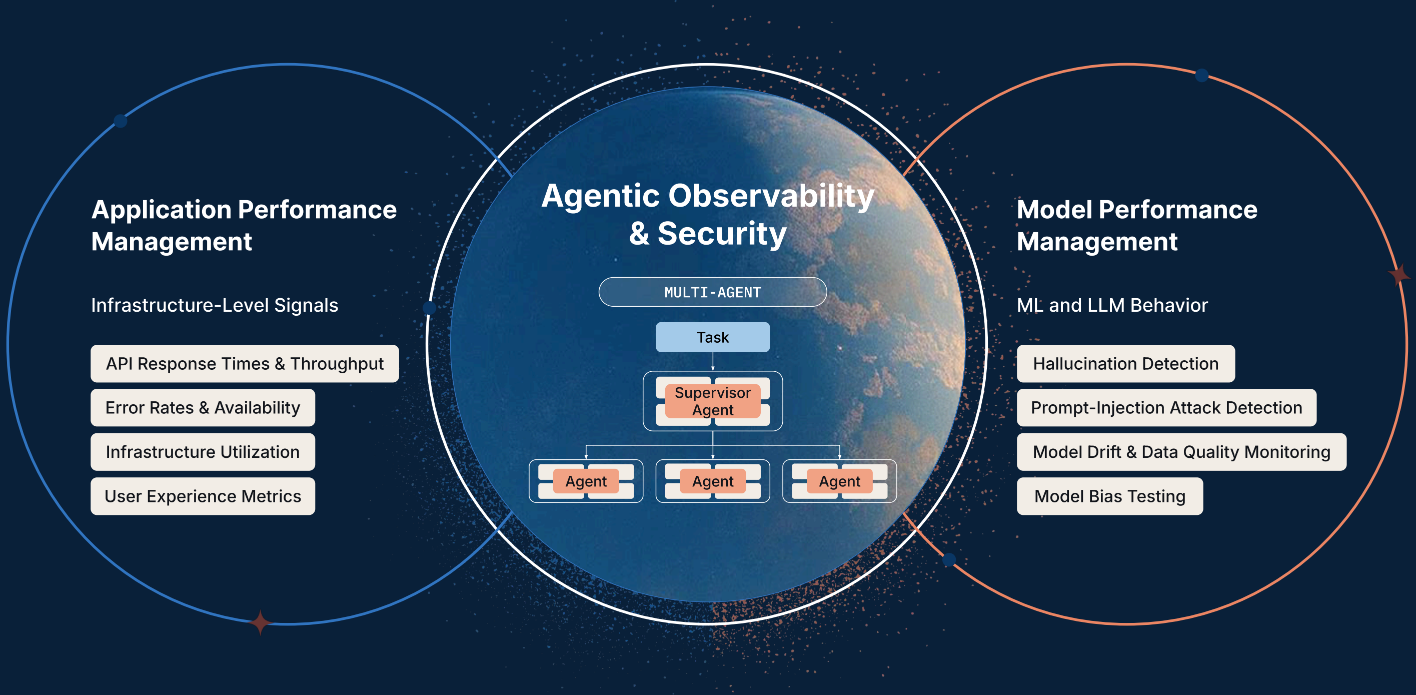
These applications dynamically generate responses, call multiple tools, and interact with other agents. They make decisions in real time, and a single unexpected action can cascade into unpredictable behavior.

Without the right observability in place, diagnosing the root cause can be slow, costly, and sometimes impossible.

The Next Generation of AI Monitoring: Agentic Observability

Many questions remain, but one thing is clear: **observability for AI agents requires a different foundation**, one built on aggregation, customization, and end-to-end visibility across every agent's lifecycle. Traditional monitoring methods can no longer keep pace with the complexity of autonomous, multi-agent systems.

To ensure trust, compliance, and scalability, enterprises need AI agent observability that moves beyond basic logs and spans to provide deeper context, actionable insights, and control.



Application Performance Management
Infrastructure-Level Signals
API Response Times & Throughput
Error Rates & Availability
Infrastructure Utilization
User Experience Metrics
Agentic Observability & Security
MULTI-AGENT
Task
Supervisor Agent
Agent
Agent
Agent
Model Performance Management
ML and LLM Behavior
Hallucination Detection
Prompt-Injection Attack Detection
Model Drift & Data Quality Monitoring
Model Bias Testing

Agentic Observability provides visibility into an agent's decision-making process by revealing its chain of interactions and decision paths

From Spans to Signals:
An Aggregation-First Mindset

Many monitoring tools approach AI observability by listing individual records of each model call, API request, or tool invocation (spans). While span-level data is necessary, relying on raw lists forces teams to dig manually through mountains of numbers, hoping to spot patterns.

This method might work for a small application, but enterprise AI agents generate millions of spans, and the traditional approach breaks down.

The future of observability for AI agents depends on an **aggregation-first mindset**. Instead of asking teams to comb through endless rows of events, each span should be scored against meaningful evaluation metrics (faithfulness, safety, PII/PHI detection, or domain-specific measures) and then aggregated over configurable time intervals.

The result is a time-series view of system behavior that makes anomalies stand out instantly. Rather than scanning row after row of data, teams can see precisely when a metric deviates from its baseline and focus their attention where it matters most. Aggregated time-series views make deviations instantly visible. Teams can zoom in on the specific moments where a system's behavior changes, rather than sifting through days of history.

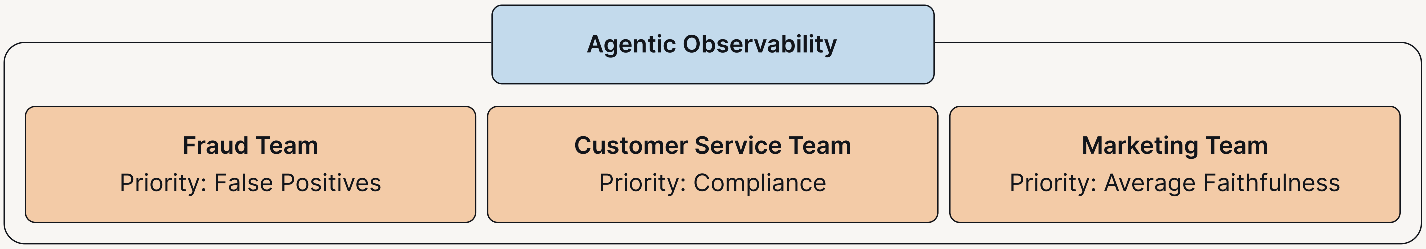
Aggregation creates a signal from the noise, allowing faster root cause analysis and enabling teams to act before small issues become major incidents. This transforms observability from a forensic exercise into a proactive capability.

Aligning Observability to Meet Flexible Business Needs

Enterprise teams measure success in different ways, and their observability practices need to reflect that. A fraud detection team might care more about false positives than latency. A customer service workflow might prioritize safety and compliance above all else.

This makes flexibility essential. The ability to choose both *which* metrics to monitor and how they're aggregated allows observability to align directly with business priorities. A product team may track the median safety score to capture worst-case risk, while a marketing team may want to monitor average faithfulness to ensure overall accuracy.

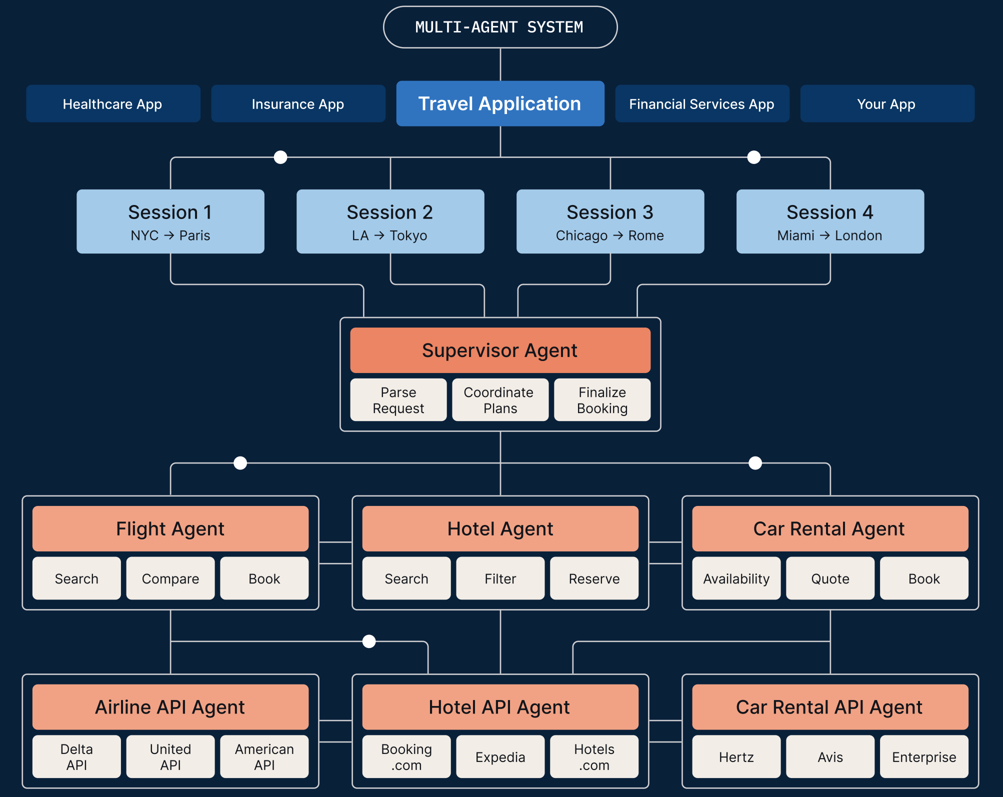
Flexibility ensures the monitoring layer reflects the organization's definition of quality and compliance, rather than a generic standard. It also makes anomaly detection more meaningful: when a high-priority metric shifts, teams know it's worth investigating.



Different teams have different goals and require flexibility to customize the metrics they need to monitor

Visibility, Context, and Control Across the Agent Hierarchy

An agentic workflow can be deceptively simple on the surface. A user issues a query, the agent retrieves data, reasons over it, invokes tools, and produces an answer. But under the hood, it could involve dozens of steps, API calls, and model invocations.



Multi-agent workflows require end-to-end traceability to pinpoint interaction failures and agent breakdowns

End-to-end observability means capturing every action the agent takes, along with its inputs and outputs. This creates a complete execution path — one that can be visualized in a tree view to show how the process unfolded.

With this view, it's possible to detect hotspots, such as an LLM making too many calls, or a specific API returning unexpected results. It's also possible to zoom in on a specific span to examine exactly what was passed in and out, and whether it met the defined quality thresholds.

This level of detail is critical for building trust. When teams can see exactly how a result was produced — and trace any failures to their origin — they can fix problems faster and with more confidence.

Key Benefits of Monitoring AI Agent Performance

Deliver High Performance AI

Proactively detect issues before they escalate to reduce downtime and accelerate root cause analysis by tracing anomalies directly to their source.

Avoid Costly Risks

Build trust in your agents and mitigate reputational risk with full visibility into how your agents generate results.

Maximize ROI

Ensure your agents deliver business value by aligning observability metrics directly with your organization's KPIs.

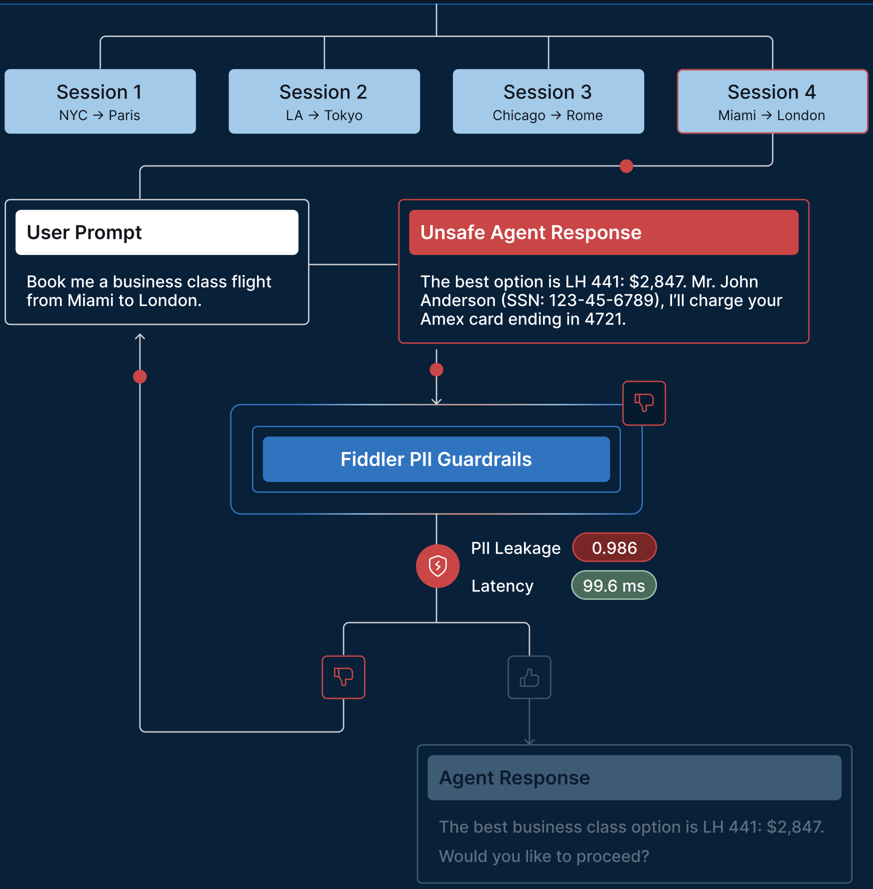
Security and Compliance in Agentic Observability

Of course, none of these agents operate in a vacuum. They exist within the security, privacy, and compliance frameworks of the enterprise. In regulated industries, these requirements are not negotiable, but they are often highly specific to each organization's policies and risk thresholds.

An observability approach fit for AI agents must accommodate these variations. That means giving teams the ability to define their own compliance metrics and alert thresholds: for example, flagging any output with a toxicity score above a set limit, and ensuring that sensitive data is protected at every stage.

Equally important is securing the data itself. For enterprise deployments, this means operating within the customer's virtual private cloud (VPC), encrypting data at rest and in transit, implementing role-based access controls, and ensuring strict data isolation between customers. Observability infrastructure must uphold the same trust standards as the AI systems it monitors.

This is where a solution layer like Fiddler Guardrails becomes critical. Guardrails provide real-time monitoring to protect against a wide spectrum of risks, including poor faithfulness, prompt injection or jailbreak attempts, toxic content, and the exposure of sensitive data like PII and PHI.



Session 1
NYC → Paris
Session 2
LA → Tokyo
Session 3
Chicago → Rome
Session 4
Miami → London
User Prompt
Book me a business class flight from Miami to London.
Unsafe Agent Response
The best option is LH 441: $2,847. Mr. John Anderson (SSN: 123-45-6789), I'll charge your Amex card ending in 4721.
Fiddler PII Guardrails
PII Leakage
0.986
Latency
99.6 ms
Agent Response
The best business class option is LH 441: $2,847. Would you like to proceed?

Protect users from unsafe agentic outputs through real-time PII guardrails and compliance monitoring, ensuring secure and compliant system interactions

The Road Ahead for Agentic Observability

Agentic AI is moving quickly from a promising concept to a practical reality. But without evolved observability, adoption will stall. The complexity of these systems, the opacity of their decision-making, and the stakes of their outputs demand much more than yesterday's monitoring techniques.

The Next Generation of Observability Will be Defined by:

Aggregation-First Analysis

Turns raw data into clear, actionable signals.

Context-Rich Execution Views

Make root cause analysis fast and precise.

Scalability for Multi-Agent Workflows

Ability to follow interactions across boundaries.

Built-In Security and Compliance Alignment

Tailored to each organization's needs.

Organizations that adopt these principles will not only reduce the risk of deploying AI agents, they will accelerate their ability to innovate with them. In the coming years, the most successful deployments will be those that combine advanced agentic capabilities with equally advanced observability.

Because in the end, the power of AI agents isn't just in what they can do, it's in knowing that they're doing it right.

Ready to take control of your AI agents?

Discover Fiddler Agentic Observability for end-to-end visibility, context, and control.

fiddler

Fiddler is the all-in-one AI Observability and Security platform for responsible AI. Our monitoring and analytics capabilities provide visibility, context and control across development and production. This gives teams actionable insights to build better, more reliable AI agents, and GenAI and ML applications. An integral part of the platform, the Fiddler Trust Service provides quality and moderation controls for AI agents and GenAI applications. Powered by cost-effective, task-specific, and scalable Fiddler-developed trust models — including cloud and VPC deployments for secure environments — it delivers the fastest guardrails in the industry.

Fortune 500 organizations use Fiddler to scale AI agents, GenAI, and ML deployments. This helps them deliver high performance AI, avoid costly AI risks, and maximize ROI.

⌂ fiddler.ai ✉ sales@fiddler.ai