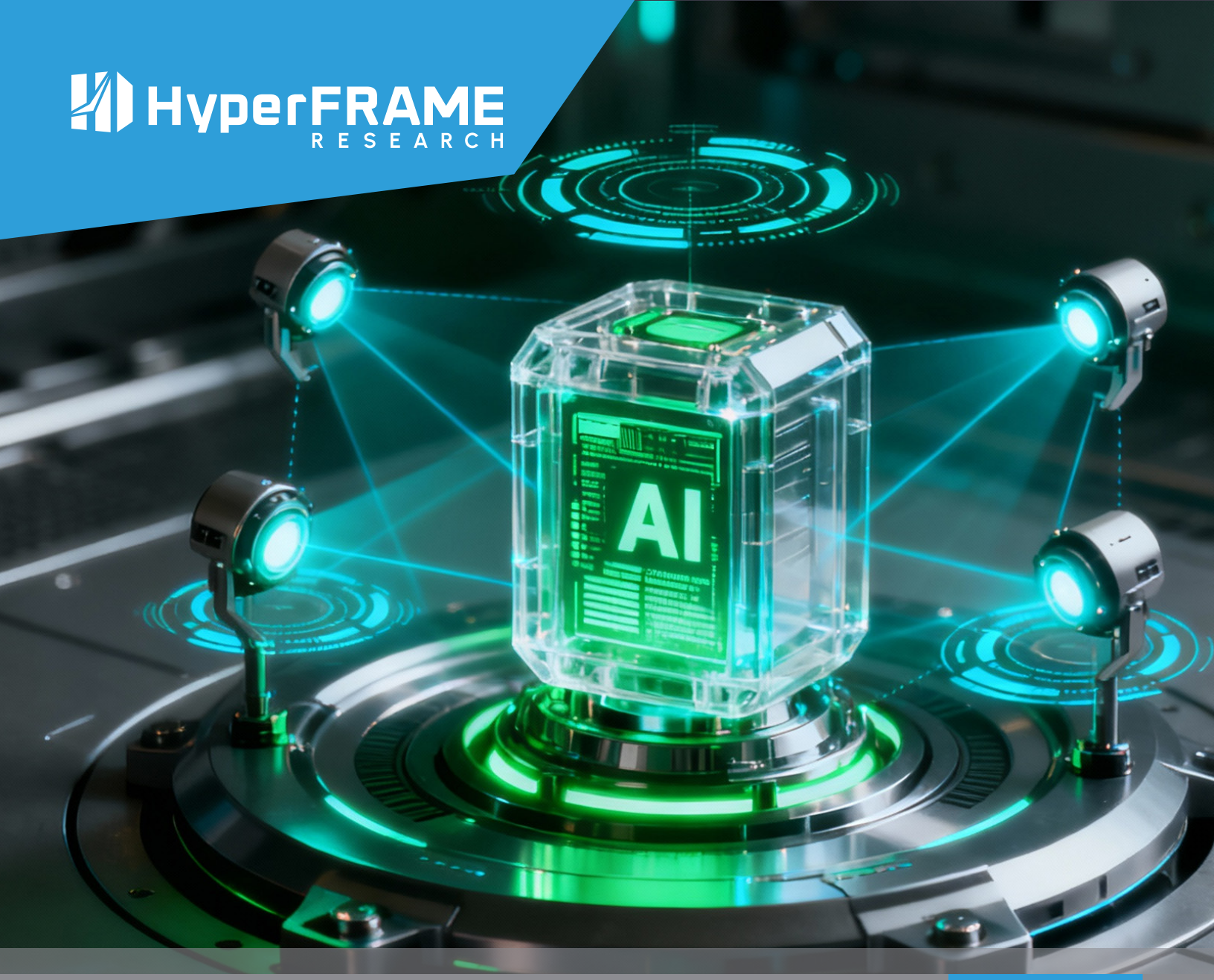
HyperFRAME RESEARCH



WHITE PAPER

Beyond Monitoring: Observability for Generative AI, Agentic AI, and LLM Workloads

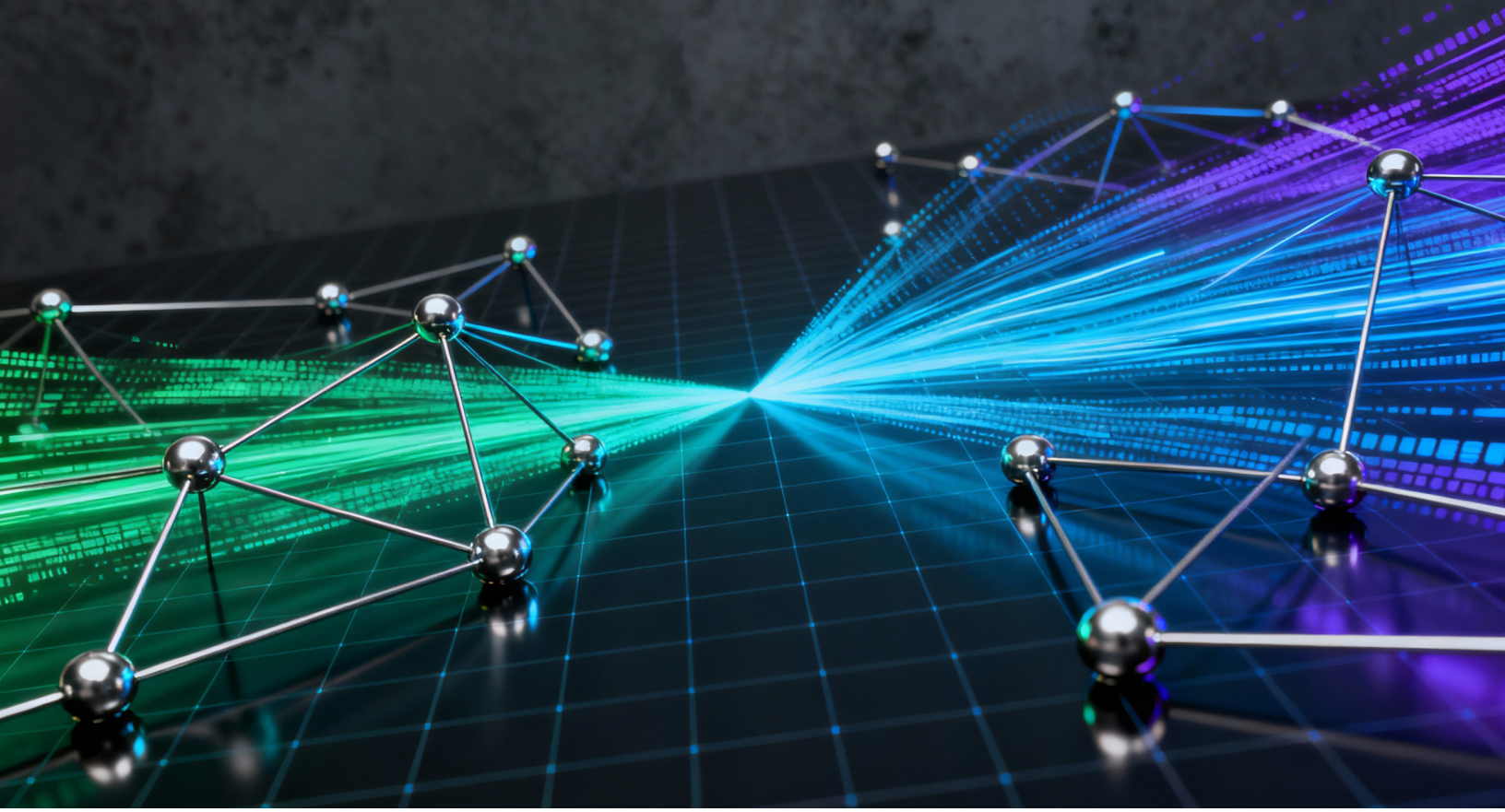
Why Complexity Demands a Modern Platform

Authors:

Stephanie Walter

Practice Leader - AI Stack

OCTOBER 2025

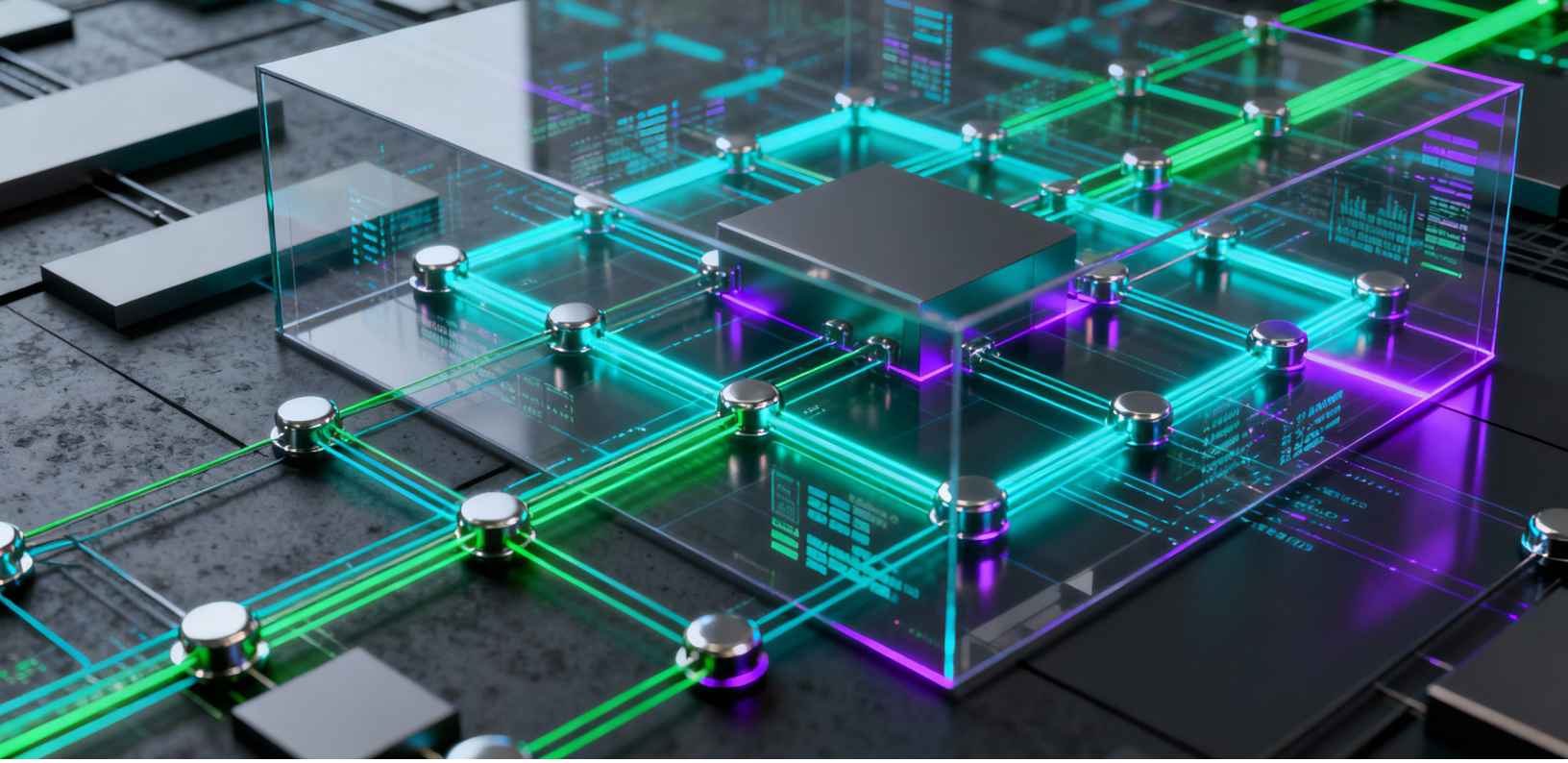


Executive Summary

According to McKinsey's State of AI 2025, 71 percent of business leaders say their organizations now regularly use generative AI in at least one function. Artificial intelligence has made significant inroads into production environments over the last year, however they face challenges in building scalable and reliable applications that can be trusted in production. Companies are piloting agentic AI systems that can act with at least some degree of autonomy, moving from AI assistants that answer simple questions to AI agents that can execute complex and multi-step tasks. The result is both a new paradigm of opportunity, but also new risk vectors.

In the past, monitoring tools were designed for deterministic systems where failures followed predictable paths. AI systems behave differently. These systems can choose alternate routes to solve the same problem. AI-centric systems may deliver one answer today and a completely different one tomorrow. This unpredictability, while powerful, makes failures harder to detect and reproduce. Costs can spike without warning. Latency can fluctuate depending on the path an agent takes. Risks can propagate across entire workflows without visibility into what triggered them. Meanwhile, AI itself has advanced in observability, moving from early machine learning models, to chatbots, to generative AI assistants, and now to fully agentic systems. Each stage has brought greater insight but also more complexity to manage.

That is why observability is no longer optional. It is mandatory. GenAI systems are complex to manage, and often agentic, which means organizations need modern observability platforms that provide early instrumentation, continuous insight, and context-rich root cause analysis. The stakes are high. Without observability, enterprises risk not only downtime, but also runaway costs, brand damage, and regulatory exposure. Based on HyperFRAME Research's analysis, Dynatrace offers one of the most complete answers to this challenge. Dynatrace's mature Davis® AI engine and new AI observability features give practitioners visibility, guardrails, and automation across AI-infused systems. The era of AI observability has arrived and is sorely needed.



The Shifting Landscape of AI in Production

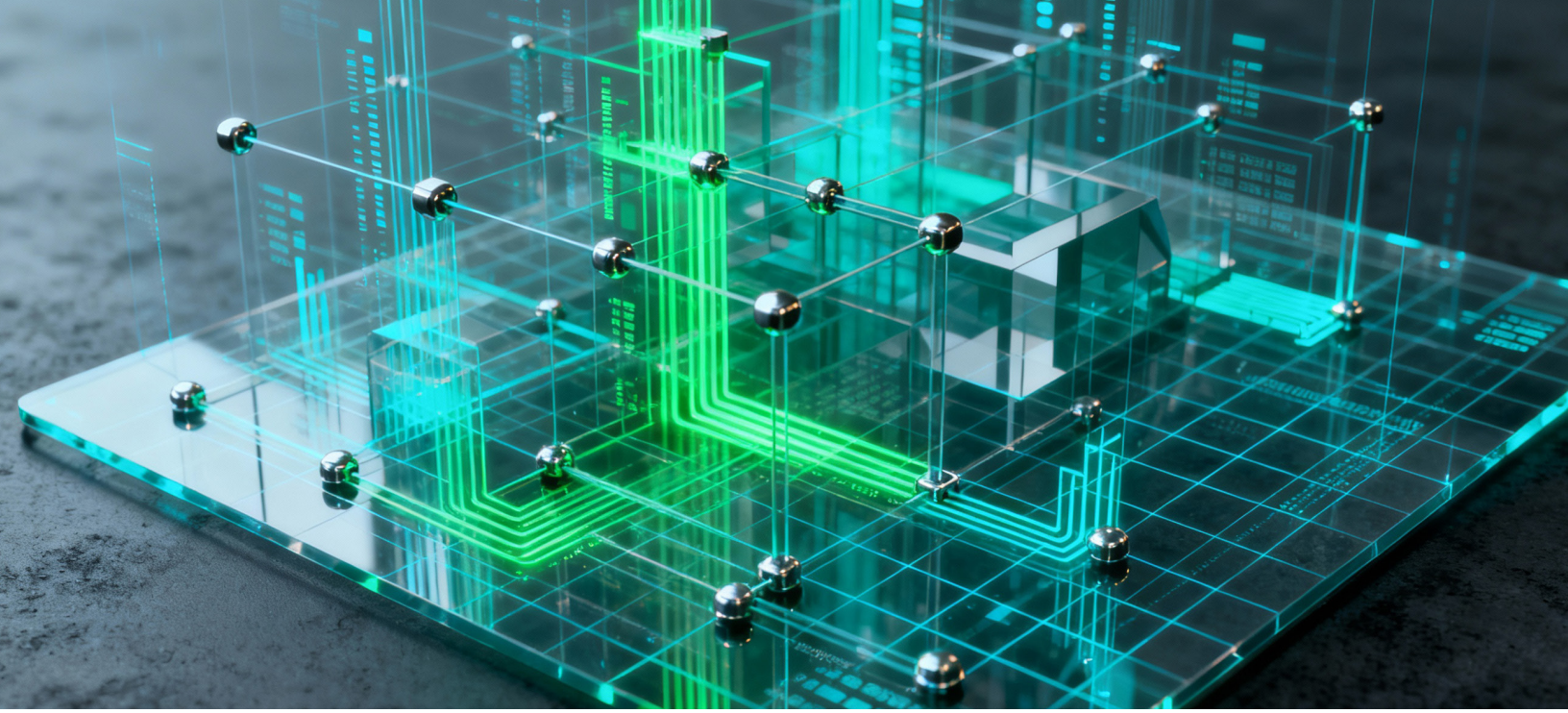
The last two years have brought a wave of AI adoption unlike anything we have seen before. Generative AI has moved from pilot projects to business-critical workloads. Customer service teams are embedding large language models (LLMs) to handle requests. Marketing departments are using them to generate content at scale. Developers are relying on AI assistants to accelerate coding. What began as experimentation is now embedded in daily operations.

Agentic AI is the next stage of this shift. These systems are designed to take action, not just generate text or images. These agentic systems can call APIs, connect to databases, and coordinate with other agents. In practice, this means an AI assistant that not only drafts an email but also updates the CRM system, schedules a meeting, and alerts a manager when needed. The orchestration across multiple agents is where both power and risk emerge.

HyperFRAME Research sees three new risks dominating early deployments:

1. **Poisoning the well:** If one agent behaves poorly, its output can cascade downstream and taint the decisions of other agents.
2. **Hidden cost:** Large-scale deployments consume tokens, compute, and API calls in unpredictable ways, creating surprise bills that can catch leaders off guard.
3. **Operational blind spots:** When LLMs are embedded in workflows, their decision-making is often invisible. Teams may not know why an answer was produced, what data was used, or how it links back to business context.

The implication is clear. Without visibility, enterprises run the risk of runaway systems, spiraling costs, and brand-damaging incidents. The opportunity of AI is massive, but so is the downside of mismanaging it. Observability is the only way to navigate this shifting landscape with confidence.



Observability is Now a Strategic Imperative

Every generation of technology creates new expectations for reliability. With AI, those expectations are complicated by unpredictability and scale. AI systems do not follow linear paths. The same prompt may trigger a completely different execution route, producing variable costs, latencies, or even contradictory outputs. This variability forces organizations to rethink how to gain visibility and control.

Instrumentation is the foundation. Without early instrumentation, teams have no way to trace the decision-making of a large language model or understand why an agent chose a particular API call. For example, a generative AI chatbot may provide a flawless answer one minute and hallucinate the next. Without instrumentation, teams cannot connect that shift back to a model update, a context window limitation, or a data drift event. By the time incidents appear in production, the cost has already been paid in wasted compute cycles, compliance violations, or poor customer experience.

Monitoring is not enough. Traditional health checks can confirm that an endpoint is "up," but they cannot reveal why a model's response quality is declining, why latency is spiking in an agentic workflow, or why costs doubled overnight. Observability provides causal explanations and context, linking telemetry across infrastructure, models, prompts, and orchestration layers. This context is essential when AI-driven systems behave in non-deterministic ways.

The risks extend beyond performance. AI systems introduce hallucinations, propagate bias, and drift over time as data changes. They can introduce hidden costs when multiple agents call external APIs in unpredictable ways. A single misrouted agent may even propagate flawed outputs downstream, poisoning entire workflows. These risks multiply as enterprises scale across applications, regions, and business units.

The aviation analogy applies. Monitoring tells you only that the plane is in the air. Observability gives you altitude, heading, fuel levels, and weather conditions. In the AI era, it is as if the plane may choose a new route every time it takes off. Without instrumentation, enterprises are flying blind, and the consequences are far greater than a delayed landing: regulatory fines, reputational damage, and runaway costs.

AI Observability Is Harder Than You Think

Many organizations underestimate what it takes to observe AI in production. Enterprises assume the same dashboards and monitoring scripts that worked for cloud workloads will stretch to cover AI. They will not. The reality is that AI introduces layers of unpredictability, complexity, and business risk that break the limits of legacy approaches.

Unpredictability is the new baseline. Agentic AI systems are non-deterministic by design. Enterprises can solve the same task in different ways on different runs. An agent asked to generate a customer email may route through one API chain today and a completely different one tomorrow. From the outside, the result may look fine, but the path taken, the costs incurred, and the time required can vary widely. Failures are hard to reproduce because the execution path constantly shifts. True observability requires not just knowing the outcome but tracing why an agent chose one path over another.

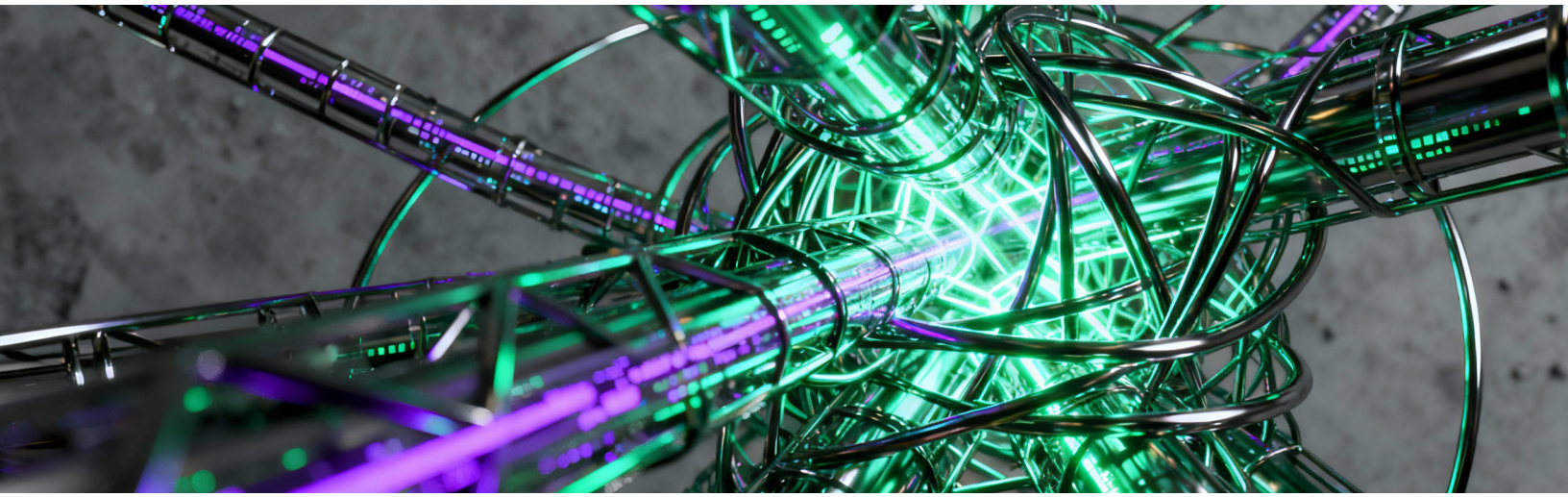
Model complexity adds another layer. Enterprises are not working with a single static model. They use multiple LLMs, fine-tuned variations, prompt engineering techniques, and custom embeddings. Each of these adds behavioral diversity. Tracking accuracy, drift, and reliability across this portfolio demands continuous observability. One model drifting slightly may not cause problems alone, but combined with others in a chain, the effects compound.

System complexity magnifies the challenge. Agentic orchestration often spans multiple services, APIs, and external data sources. A small anomaly in one component can ripple across the system and create unexpected behaviors far downstream. Traditional monitoring tools cannot map these dependencies at the speed or depth required. Business complexity is just as important. Enterprises face compliance, governance, and ethical requirements. They must prove transparency in decision-making, safeguard personally identifiable information, and ensure fairness in outcomes. Costs and service-level agreements also become harder to predict when agents take dynamic paths. Without observability, enterprises cannot meet these obligations with confidence.

Customer experience adds another dimension of difficulty. Metrics such as uptime or latency cannot fully capture how users perceive AI-driven interactions. A model may technically perform within thresholds while still delivering confusing, biased, or low-quality responses that erode trust. Measuring experience requires observability platforms to go beyond infrastructure health and track how AI outcomes affect customer satisfaction, business processes, and brand perception in real time. Without this layer of insight, organizations risk believing systems are performing well while users quietly disengage or lose confidence.

This is why “do it yourself” does not work. Manual dashboards and custom scripts simply cannot keep pace with the velocity of agentic systems. Most enterprises do not have the engineering resources to build observability frameworks capable of spanning models, orchestration, infrastructure, and business rules. Complexity is compounding across every layer.

The emerging reality is sobering. Without observability, enterprises will be overwhelmed. The pace of AI adoption ensures that unpredictability, model diversity, system interdependencies, and business requirements will only grow. Observability is not just harder than expected; it is impossible without a platform approach.

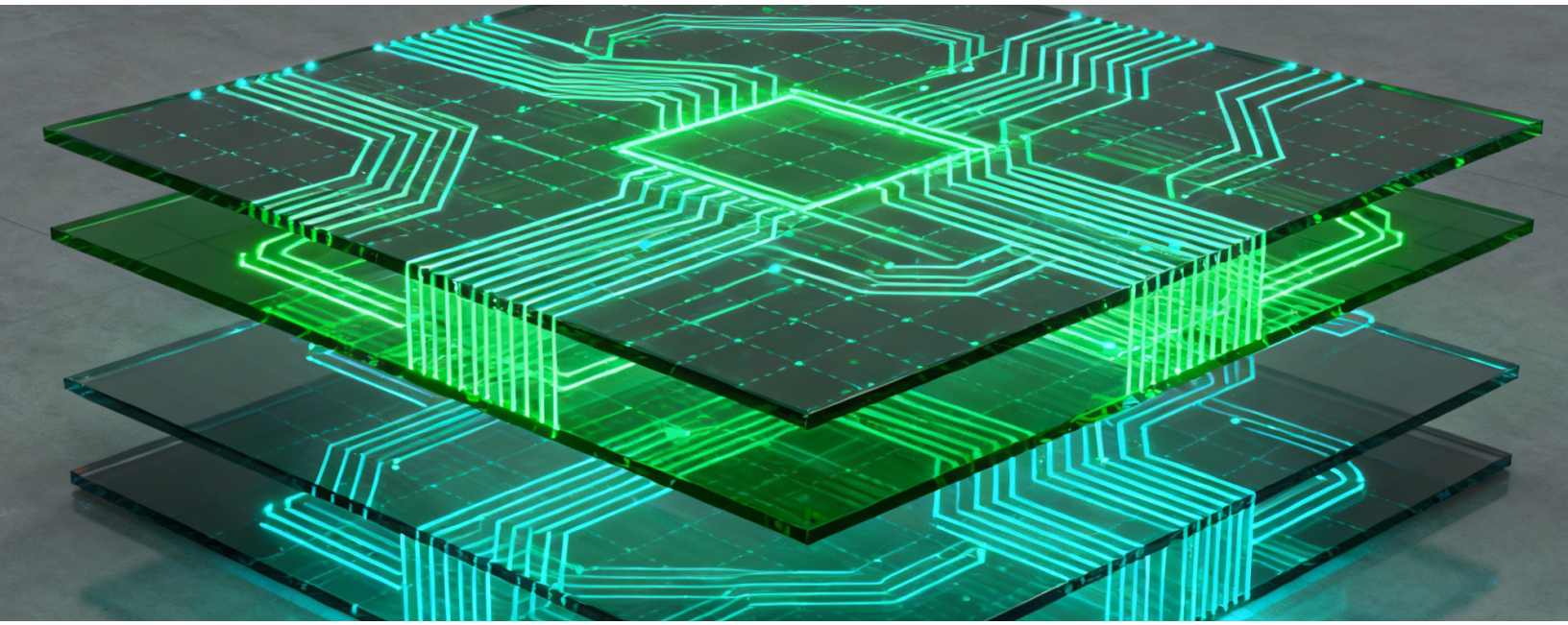


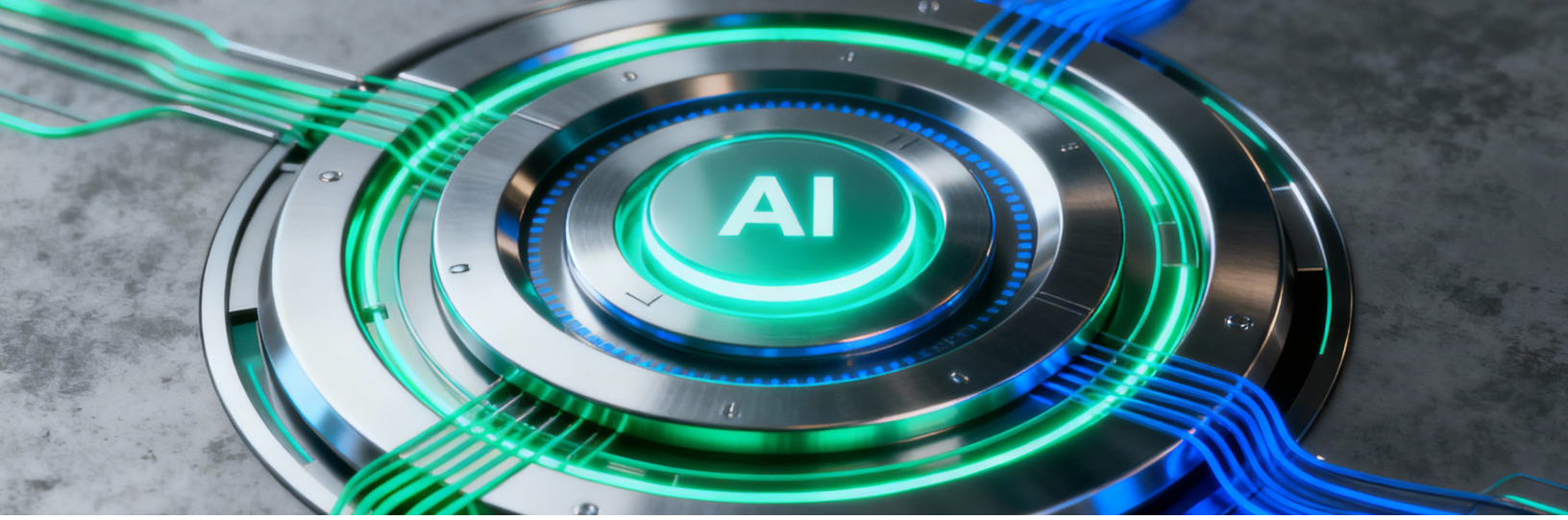
Recommendations for Practitioners

AI complexity is not something to manage after the fact. HyperFRAME Research recommends six practical steps for practitioners who are preparing their organizations for generative and agentic AI.

1. **Instrument early**. Observability should be considered from the start. Waiting until production incidents force a reaction means that valuable time and money have already been lost. Early instrumentation creates the baseline needed to detect drift, anomalies, and inefficiencies as they emerge.
2. **Adopt modern observability platforms**. Organizations deploying AI systems should adopt modern observability platforms specifically designed for dynamic and autonomous workloads. These solutions enable rapid development cycles and ensure operational reliability by providing real-time metrics such as hallucination rates, token consumption, and anomaly detection that are critical for managing performance, safety, and cost at scale.
3. **Prioritize root cause analysis**. Anomaly detection alone is not enough. Teams need causal explanations that show not only what went wrong, but why it happened, and how to prevent recurrence. Focus on explainability, value-driven actions, and layered governance to safeguard systems, build trust, and ensure responsible automation as teams validate and scale AI capabilities.
4. **Evaluate agentic workflows critically**. Orchestration across multiple agents compounds complexity. Instrument systems like Amazon Bedrock AgentCore, Azure AI Foundry Agent Service, Google Agent Development Kit (ADK), and Anthropic Claude Code to trace decisions, tool calls, memory, costs, latency, and safety events across the full workflow. Require per-agent and cross-agent traces, tool-invocation logs, guardrail outcomes, and business KPIs to detect branching into untested paths and emergent behaviors early.
5. **Leverage open standards like OpenTelemetry and OpenLLMetry**. These frameworks are becoming the de facto industry standards, adopted by AWS, Google, OpenAI, and others. Using open standards ensures interoperability, reduces vendor lock-in, and makes it easier to integrate observability across diverse AI workloads.
6. **Recognize the limits of DIY**. Enterprises should be honest about the scale and velocity of AI. Manual dashboards or isolated scripts will not keep pace. Automation must be built into observability, or teams risk being consumed by noise rather than insight.

These practices form a checklist for practitioners. Observability is a requirement for delivering AI systems that are efficient, reliable, and trustworthy.





Dynatrace as a Modern AI Observability Platform

Dynatrace represents a mature, but innovative platform designed for the realities of AI in production. It is cloud native, built for scale, and architected to handle both the depth and unpredictability of modern workloads. What sets it apart is its ability to combine proven core observability with specialized capabilities for generative and agentic AI.

Dynatrace provides a centralized, in-context data lakehouse, Grail, for all telemetry. Davis AI interprets and analyzes telemetry in Grail in real time, translating AI complexity into actionable insights. Its root cause analysis, anomaly detection, natural language explanations, and remediation steps help close reliability gaps. The unified platform breaks down telemetry silos, delivering in-context data that improves visibility early so customers can move into production with confidence.

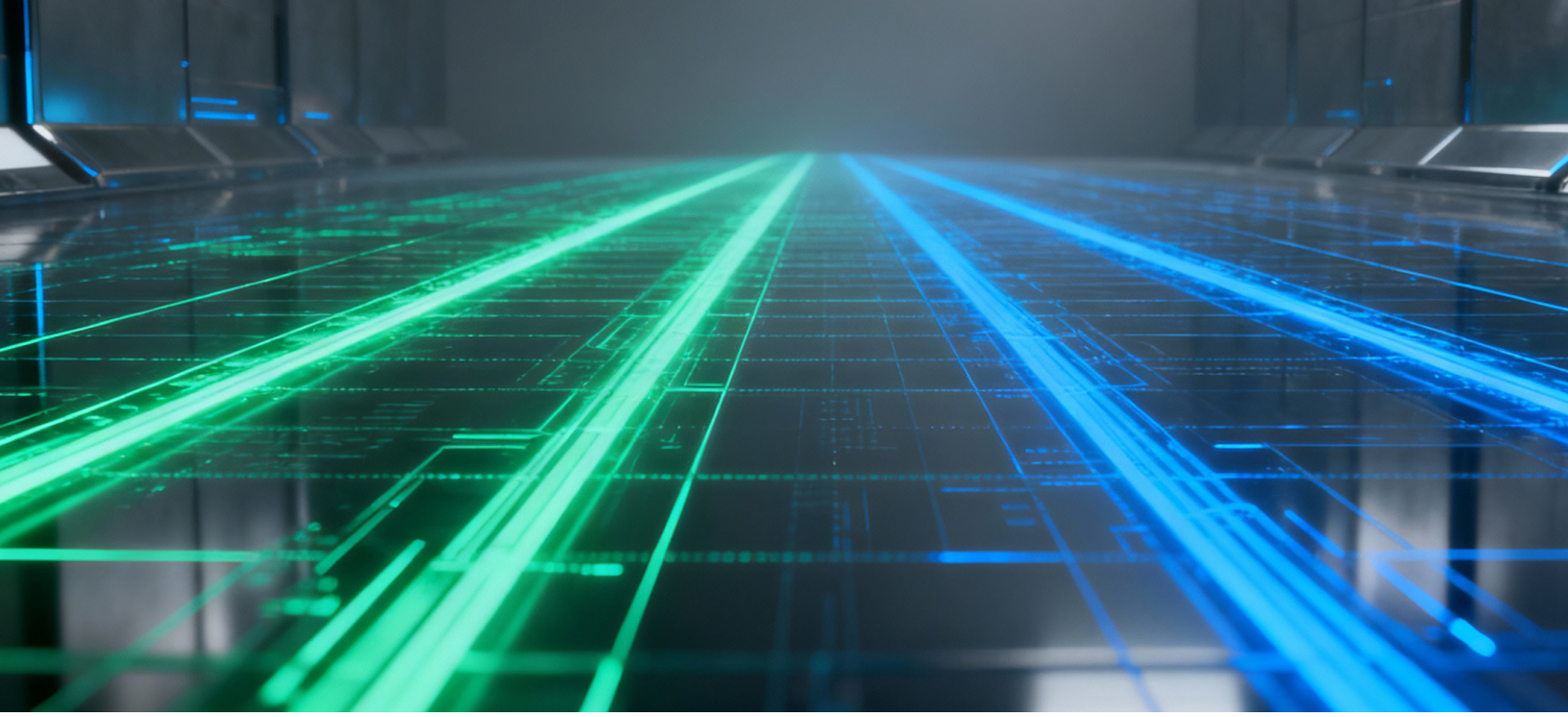
Dynatrace also extends observability into automation. With its Automation Engine and App Engine, the platform enables organizations to define goals, then let AI determine the best way to achieve them. This includes proactive remediation, reactive fixes, and ongoing optimization. Human teams set the boundaries while the system does the heavy lifting.

For AI observability specifically, Dynatrace provides a set of capabilities practitioners can adopt immediately:

- **Visibility into LLM and agent usage**. Track token consumption, latency, errors, and cost in real time with customizable dashboards.
- **Detection of drift and anomalies**. Davis AI continuously monitors fine-tuned models, orchestrated agents, and RAG pipelines for deviations in performance or behavior.
- **Guardrails for generative AI**. Monitor and enforce input/output guardrails to detect hallucinations, prevent malicious prompt injections, stop PII leakage, and filter toxic language.
- **Context-rich tracing and root cause analysis**. End-to-end logging and tracing across application, orchestration, RAG, LLM, and agentic layers enable Davis AI to automatically pinpoint root causes and accelerate remediation.
- **Compliance and audit support**. Capture full lineage of prompts and responses, retain up to 10 years of logs, and provide dashboards that demonstrate transparency and regulatory compliance.
- **Davis CoPilot**. Interact with telemetry and insights in natural language, accelerating troubleshooting, optimization, and decision-making.

The Dynatrace platform supports practical use cases that organizations can adopt quickly. These include AI cost tracking and optimization, drift detection for generative models, monitoring agent workflows for malicious activity, and identifying performance bottlenecks in AI-augmented applications. Each use case provides immediate value while building a foundation for long-term scale.

Dynatrace delivers a rare combination: maturity in observability, proven AI expertise, and purpose-built features for the complexity of generative and agentic AI. For practitioners facing unpredictable workloads, it is not just a tool but a strategic enabler of trust and confidence.



Looking Ahead

Generative and agentic AI promise to transform how enterprises operate, but the complexity they introduce cannot be ignored. These systems behave in ways that are powerful but unpredictable. AI systems can drift, hallucinate, or take entirely new paths through a workflow. They can escalate costs and introduce risks faster than traditional monitoring tools can detect.

The defining challenge of this era is complexity. The defining solution is observability. Without it, organizations are left to guess at the causes of failures, absorb runaway expenses, or suffer reputational harm when AI systems misfire. With it, organizations can gain clarity, control, and confidence.

Practitioners have a clear playbook: instrument early, embrace modern observability platforms, prioritize root cause analysis, and recognize the limits of manual approaches. The responsibility is to be proactive, not reactive.

Dynatrace provides a platform designed for this reality. With its mature Davis AI engine and modern AI observability capabilities, it gives enterprises the visibility and automation they need. Observability is no longer optional. It is the foundation for success in the age of AI.

ABOUT HYPERFRAME RESEARCH:

HyperFRAME Research delivers in-depth research and insights across the global technology landscape, spanning everything from hyperscale public cloud to the mainframe and everything in between. We offer strategic advisory services, custom research reports, tailored consulting engagements, digital events, go to market planning, message testing, and lead generation programs.

Our industry analysts specialize in rigorous qualitative and quantitative assessments of technology solutions, business challenges, market forces, and end user demands across industry sectors. HyperFRAME Research collaborates closely with your Analyst Relations, Product, and Marketing teams to build and amplify your thought leadership, positioning your expertise to enhance brand and product recognition. Through content that engages readers, viewers, and listeners alike, we ensure your voice resonates across channels.

CONTACT HYPERFRAME RESEARCH:

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

Email Address:

steven.dickens@hyperframeresearch.com

Telephone Number:

+1 845 505 1678

X: @StevenDickens3

LinkedIn: Steven Dickens

BlueSky: Steven Dickens

CONTRIBUTORS

Steven Dickens
CEO & Principal Analyst

Stephanie Walter
Practice Leader - AI Stack

INQUIRIES

Contact us if you would like to discuss this report and HyperFRAME Research will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "HyperFRAME Research." Non-press and non-analysts must receive prior written permission by HyperFRAME Research for any citations.

LICENSING

DISCLOSURES

HyperFRAME Research provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.


HyperFRAME
RESEARCH