Operationalize generative AI with confidence using Datadog

As public sector organizations scale generative artificial intelligence (AI) from proof-of-concept to production, they encounter new complexities in monitoring AI agent performance.

Traditional metrics like uptime, latency, and vulnerabilities are no longer enough to evaluate success. Teams also need real-time visibility into large language model (LLM) behavior, including accuracy, output quality, costs, and potential hallucinations. Integrated generative AI observability helps organizations detect and address LLM performance issues as they happen, while also generating audit trails to meet federal, state, and local compliance requirements.

Datadog, an Amazon Web Services (AWS) Advanced Tier Partner, offers a comprehensive observability platform that unifies infrastructure monitoring, application performance, real user monitoring, security, and compliance. With its LLM Observability application, Datadog helps public sector teams track model behavior, correlate it with underlying infrastructure performance, and streamline regulatory compliance to support confident, scalable generative AI solutions.

Why Datadog?

Full-stack observability in one platform

Public sector IT teams often manage complex environments with constrained resources. Datadog LLM Observability unifies infrastructure, applications, and LLM behavior in a single platform, eliminating the need to switch between traditional and specialized AI monitoring solutions.

Zero-code stack integrations

Public sector organizations often rely on a mix of vendors, LLMs, AI frameworks, and data sources. Datadog's mature tracing framework and auto-instrumentation works across all of them without requiring custom code, accelerating observability without time-consuming integration work.

Built-in compliance and security controls

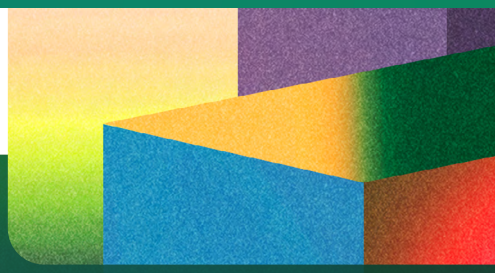
Datadog is FedRAMP authorized and maintains compliance with HIPAA, SOC2, PCI, and other security frameworks. Built-in features like role-based access control (RBAC) and audit logging tools help streamline oversight and simplify compliance reporting.

Bird's-eye view of LLM and user behavior

Without observability into how users interact with LLMs, organizations miss opportunities to identify user questions, misinformation, and even data breach attempts. Datadog LLM Observability groups prompts and responses with semantic clustering by theme—providing insight into user intent and where information gaps may exist—to deliver better services.

Solution traceability and explainability

Granular insight into LLM responses, including data sources, speed, and cost, helps improve transparency, boost cost-efficiency, and identify root causes of issues.



DATADOG *in action*

One of the largest public universities in the United States uses Datadog to monitor a student-facing generative AI chatbot that helps with coursework, financial aid questions, and more.

When response accuracy suddenly dropped, Datadog's LLM Observability platform quickly alerted the team and identified the root cause: an underlying data source in their retrieval-augmented generation (RAG) pipeline hadn't been updated. Using Datadog's full traceability and alerting tooling, the university quickly resolved the issue to continue providing trusted information to its students.

"With better traceability and explainability…customers usually find [Datadog] at least 30% faster in terms of reduction in MTTR [mean time to resolution]."

– Shri Subramanian, Group Product Manager, Datadog

Building confidence with generative AI observability

As public sector leaders scale generative AI pilots quickly across their organizations, effective monitoring is essential. Datadog recommends treating observability as a foundational development step, not an add-on, to establish responsible adoption of generative AI. To get started:

1 **Embed LLM observability early** in the proof-of-concept phase. Don't wait until production to implement monitoring. Early visibility helps teams identify any operational or behavioral issues before they scale.

2 **Establish guardrails and golden datasets.** "With generative AI, there's a lot of unpredictability… You need a good set of test prompts to cover edge cases, guardrails, and real-world variation," says Jeetendra Soneja, Director of Engineering for LLM Observability at Datadog. Define structured prompts, edge cases, and reference outputs early to set expectations for model behavior. These golden datasets provide a consistent baseline to test against during development and help catch regressions before they reach production.

3 **Continuously refine based on real-world behavior.** The ability to observe how models perform in production and iterate quickly is essential. Monitor operational performance metrics (like model latency and cost) along with generative AI-specific evaluations such as sentiment analysis, output relevance scoring, and hallucination tracking, or even custom-defined evaluations like user satisfaction scores. Use these insights to experiment with your prompts, policies, and models to improve performance, quality, and costs.

Is Datadog right for my organization?

Ask yourself these questions to assess whether your teams could benefit from Datadog's generative AI observability platform:

- *Is your organization launching AI pilots without visibility across all of your AI agents?*
- *Do you lack insight into LLM inputs, outputs, and prompt performance?*
- *Does your organization struggle to meet compliance and audit requirements for generative AI workflows?*
- *Are your observability tools siloed across different systems, applications, and AI agents?*

"Having a single observability platform reduces audit overhead and accelerates incident response, which is a double win for public sector teams."

– Shri Subramanian, Group Product Manager, Datadog

AI success with AWS and Datadog

End-to-end system visibility

Datadog integrates with AWS services like AWS Lambda and AWS Step Functions, along with Amazon Bedrock, Amazon SageMaker, and more, to give organizations a unified view of LLM behavior, application workflows, and the infrastructure that supports them.

Zero-code integrations with Amazon Bedrock

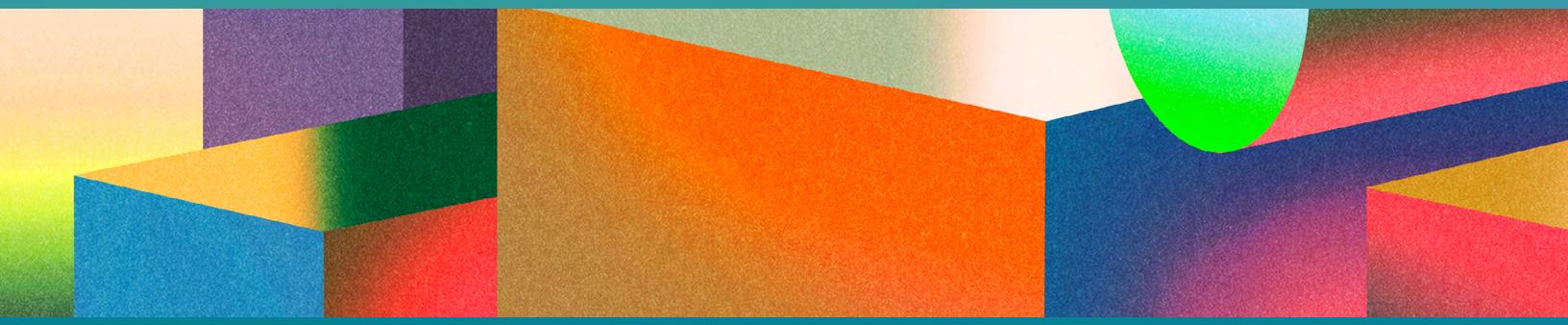
With native support for Amazon Bedrock and integrations for Amazon Bedrock Agents, Datadog offers seamless observability across multiple AI agents with no coding required.

Enhanced performance visibility

By connecting directly to Amazon SageMaker, Datadog provides real-time insights into model performance, cost trends, and resource utilization to support better decision-making in production environments.

Leading security and compliance

AWS supports 143 security standards and compliance certifications, which Datadog builds on to meet FedRAMP, HIPAA, SOC2, and PCI standards. Using AWS logging and monitoring services, Datadog helps organizations centralize security and compliance data, streamline auditing processes, and maintain consistent compliance reporting.



aws | DATADOG

Learn more at datadog.com/gov or open a trial account at app.ddog-gov.com/signup →