LLM observability: A must for scaling secure generative AI

Public-sector organizations are seeing early wins with generative artificial intelligence from applications in document processing, learning and development programs, citizen services, and more.

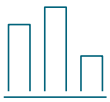
But as pilots grow into enterprise solutions, organizations need deeper visibility into how large language models (LLMs) perform in production environments. Observability is integral to making sure these systems scale with reliability, efficiency, and accountability.

Amazon Web Services (AWS) and Dynatrace work together to help public-sector teams move beyond proof of concept with real-time observability across the entire generative AI stack. From monitoring response times and usage patterns to detecting drift, bias, or anomalous behavior, Dynatrace uses AWS to provide the operational intelligence needed to scale responsibly, manage costs, and stay aligned with mission goals.

How Dynatrace AI observability helps public-sector organizations succeed

Built for AI-specific security and compliance

Generative AI introduces security and compliance considerations that traditional monitoring tools aren't designed to address, such as prompt injection attempts, model behavior anomalies, and data lineage tracking. Dynatrace addresses these gaps by enriching standard security logs with AI-specific context and providing the specialized monitoring that organizations need for compliance reporting.



Track and optimize AI costs at scale

LLM-based applications can generate unpredictable costs due to token usage and model queries. Dynatrace provides continuous visibility into AI spending, automatically flags anomalies, and helps teams fine-tune workloads to stay within budget.

Maintain trust with real-time explainability and governance

As generative AI scales, model behavior must remain reliable and auditable. Dynatrace continuously analyzes LLM outputs to detect hallucinations, drift, or bias, catching issues before they affect services. Using Model Context Protocol (MCP), Dynatrace enriches models with real-time operational context, making AI decisions easier to trace, explain, and govern across live environments.

"Faster root-cause analysis and faster time to resolution means you can tune your models faster and get services out to the field or constituents more quickly. Time to mission is a real thing."

– Willie Hicks, Public Sector Chief Technologist, Dynatrace

Is Dynatrace right for my organization?

Is compliance delaying your AI initiatives? *Meeting authorization requirements can slow down deployment. Dynatrace simplifies the process with continuous monitoring, detailed audit trails, and real-time documentation of system behavior to help compliance teams accelerate ATO and reduce review cycles.*

Do you lack full visibility into your generative AI workflows? *Many organizations monitor infrastructure but can't track their generative solutions from end to end. Dynatrace lets organizations monitor their full AI stack across their entire infrastructure.*

Do you need clearer AI ROI? *Executives want more than usage stats. Dynatrace delivers operational intelligence that connects technology performance to organizational outcomes.*

Do you need to demonstrate AI value and governance? *Public-sector transparency demands evidence of responsible AI use and measurable results. Real-time dashboards and automated reporting help document performance improvements for oversight requirements.*

How to get started with generative AI observability

To scale effective generative AI solutions across the public sector, leaders need clear insight into how LLM systems behave, what they cost, and how they comply. Willie Hicks, public sector chief technologist at Dynatrace, recommends these four foundational steps to help public-sector leaders approach adopting an effective generative AI observability strategy:

- **Identify visibility gaps.** Take stock of your current monitoring tools. Where do you have gaps in visibility across your AI technology stack? Blind spots in model inputs, outputs, or behavior often become blockers when moving from pilot to production.
- **Establish baseline performance before scaling.** Record what your pilot AI systems cost to run, how fast they respond, and what security events they generate. This provides a reference point for scaling responsibly.
- **Put guardrails in place early.** Set up automated alerts for signs of drift, hallucinations, cost spikes, or prompt injection attempts. Logging this activity early supports faster mitigation and future audits.
- **Unify AI observability and debugging.** Look for tools that make it simple to integrate real-time LLM monitoring with live debugging so teams can trace AI behavior back to its context and code to troubleshoot anomalies quickly.
- **Work with vendors who know federal requirements.** Generic AI tools often aren't equipped to handle government compliance needs. Find partners who understand authority-to-operate (ATO) processes and can operate in your specific regulatory environment.