TRACK 1

Implement generative AI securely into your organization

As public sector organizations begin to scale generative artificial intelligence (AI) across teams and departments, leaders have an imperative to build solutions that operate securely, accurately, and responsibly from day one.

That starts with a strong security posture, which encompasses data encryption, identity-aware access controls, and responsible usage policies grounded in zero-trust principles. These protections should span the entire generative AI workflow, from input and data retrieval to model output and downstream actions for agentic AI solutions.

Enterprise-wide observability also plays a key role in managing and maintaining this secure foundation. Monitoring how large language models (LLMs) continuously perform in production, including prompt behavior, model output and accuracy, latency, and token usage, helps teams detect anomalies, manage cost, and align generative AI performance with mission needs.

Amazon Web Services (AWS) and AWS Partners help public sector teams build this secure operational foundation. From enforcing AI-specific governance controls to monitoring performance, this track highlights how to begin implementing generative AI solutions securely across an organization.

Why choose AWS Partners?

Secure-by-design solutions for generative AI

AWS Partner offerings support zero-trust principles, secure access controls, enterprise-wide AI model governance, and real-time visibility into model behavior so public sector organizations can scale responsibly and reduce operational risk.

Explore curated generative AI governance and monitoring solutions in the Partner Expo.

Why organizations need enterprise-wide governance and monitoring for generative AI

Because generative AI systems interact dynamically with users, data, and infrastructure, maintaining continuous and proactive visibility across the generative AI prompt "lifecycle" is essential for long-term success. Strong generative AI governance strategies and solutions help organizations:

Safeguard applications and data with built-in security

Each layer of the AI stack presents unique risk considerations. Strong encryption and access controls that apply a zero-trust architecture model across each layer support solution integrity at every stage of a generative AI workflow. The Generative AI Security Scoping Matrix helps customers match their AI workloads with the right security and compliance controls to support protection of data and assets.

Track AI usage and performance in real time

Continuous monitoring capabilities can help teams understand how models are being used, how well they are performing, and how they align with an organization's expectations and goals. Monitoring inputs, outputs, latency, and user patterns maintains operational confidence as generative AI solutions grow.

Manage cost and resource utilization

Generative AI workloads can vary depending on user activity, model complexity, and data flow. Access to real-time token usage supports budgeting and helps organizations scale generative AI solutions responsibly.

Maintain consistency and accuracy

By observing system behavior and validating model outputs against established goals or datasets, organizations can make sure their generative AI applications continue to deliver relevant, high-quality responses as conditions evolve. If organizations begin to detect inaccuracies in generative AI model outputs, an effective monitoring system allows teams to identify and troubleshoot the cause—like inaccurate data sources or recent model updates—and quickly remediate to maintain trust.

How AWS supports secure, scalable AI operations

AWS helps organizations to:

- ***Protect data at every layer of the AI stack*** *with built-in security measures that enable isolation, encryption, and granular access controls to support confidentiality, integrity, and availability.*
- ***Standardize governance and auditability*** *across hybrid and multi-account environments with centralized logging and integrated compliance tooling.*
- ***Optimize performance and cost*** *by monitoring compute and model usage in near real time, helping you scale efficiently while staying on budget.*
- ***Build on a secure foundation,*** *using FedRAMP-authorized infrastructure to meet rigorous requirements for security, privacy, and compliance.*
- ***Access a robust network of AWS Partners,*** *who offer enhanced security capabilities tailor-made for public sector organizations.*

How to get started with generative AI governance

1 Develop security-first architecture and data access controls

To build a secure AI foundation, review how your data flows across the solution, who can access it, and whether permissions reflect zero-trust principles. Develop role-based access controls and just-in-time permissions to reduce risk from lateral movement, in which users or systems gain unintended access to adjacent data or environments.

2 Establish a baseline for performance and usage

Before scaling a generative AI pilot to organization-wide production, determine a performance baseline for response times, token costs, and output behaviors. Observing these benchmarks early lets your team identify meaningful trends and better detect issues as they arise.

3 Align your monitoring strategy with your mission priorities

Define what matters most to your organization—faster response times, lower operational costs, or improved service quality—and use those goals to shape your operational monitoring strategy. Set relevant thresholds and alerts to support mission-related outcomes alongside technical metrics.

4 Define responsible usage policies upfront

Set clear policies around user access and acceptable use. Tools like Amazon Bedrock Guardrails can help enforce these policies programmatically across a range of leading foundation models (FMs), so your generative AI solutions align with your organization's standards without requiring a custom-built LLM.

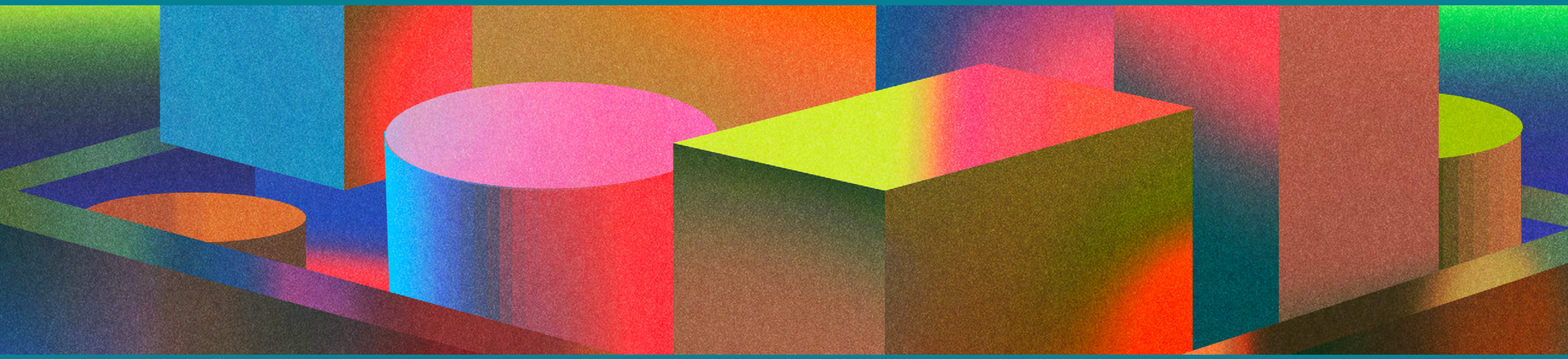
5 Close cross-functional feedback loops early and often

Secure and performant generative AI requires coordination across many teams: IT teams for design and deployment, security teams for threat monitoring and compliance, business owners to align use with mission needs, and end users to ground everything in practical feedback. Create communication channels that connect these core stakeholders early so feedback flows seamlessly across these groups.

Explore featured solutions from AWS Partners

AWS Partners are helping public sector organizations securely operationalize generative AI with enterprise-wide solutions that scale automatically as adoption grows.

Visit the "Partner Expo" to explore how these solutions can support your organization's goals →



KEY CONTRIBUTORS

Mehmet Bakkaloglu | *Principal Solutions Architect for ISV Partners, AWS*
Varun Jasti | *Solutions Architect for Worldwide Public Sector, AWS*