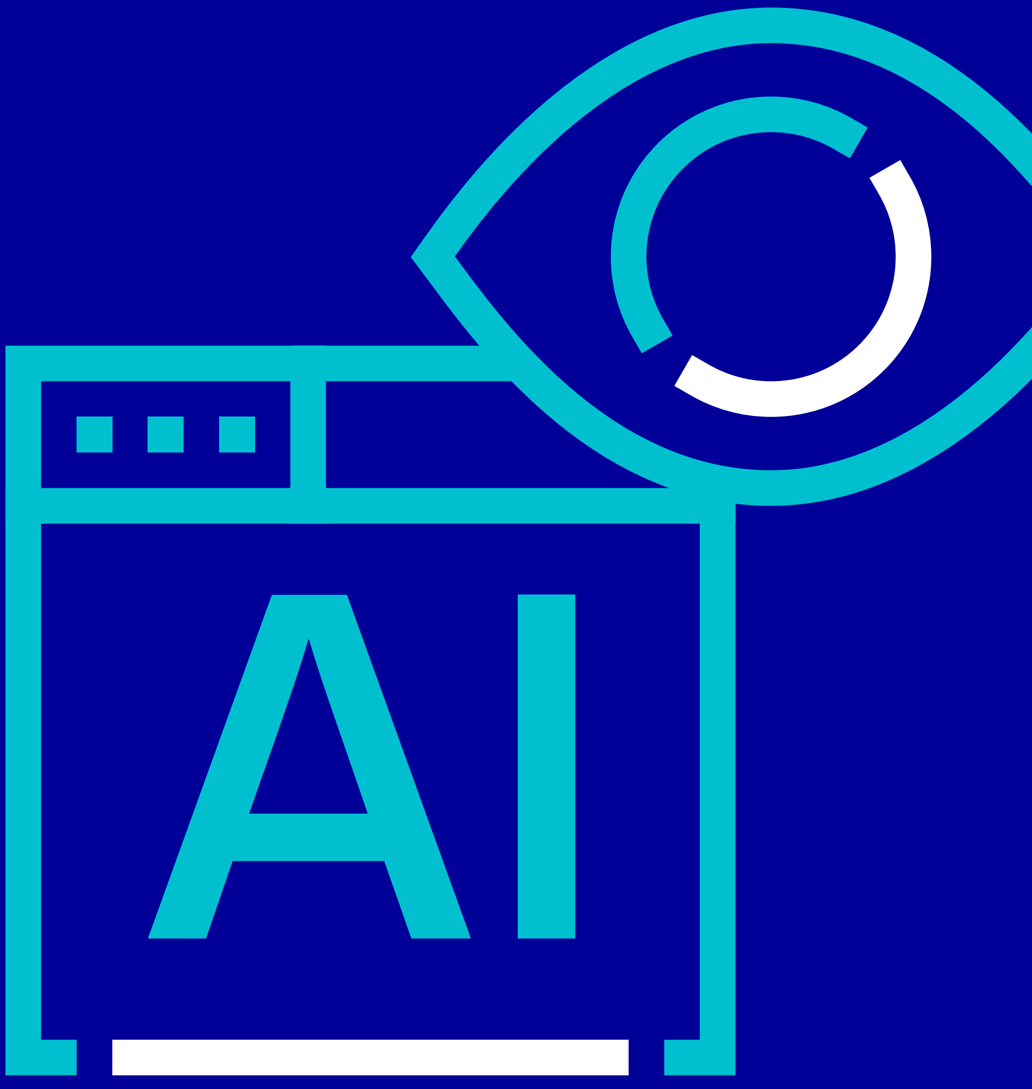
sumo logic

WHITE PAPER

Governing AI in the age of agentic systems and Model Context Protocol

David Girvin, Sumo Logic
October 2025



Executive Summary

Digital organizations face an incredible new pace of innovation; artificial intelligence is moving faster than almost any other technology in history. It can now autonomously detect threats, investigate incidents, and execute responses across complex cloud environments. We are no longer dealing only with static models that produce recommendations; this is an era of agentic AI systems that can act on behalf of users, alongside the Model Context Protocol (MCP) that standardizes how agents, tools, and models exchange context.

At the same time, enterprises need a Model Control Plane to enforce policy, security, and observability across these components. Organizations that embed governance into their AI architecture gain a decisive advantage: the ability to innovate at scale while maintaining security, compliance, and operational reliability.

Governance models of the past—checklists, ad hoc oversight, or isolated audits—are not enough. AI systems evolve dynamically, interact with sensitive data, and operate across borders and regulatory regimes. For organizations to succeed, governance must be adaptive, transparent, and embedded directly into the architecture. This paper presents a forward-looking framework that blends global regulatory developments, academic insights, and industry best practices—tailored for enterprises seeking to innovate safely and responsibly.

Note on terminology: In this paper, **MCP** refers to the **Model Context Protocol** (interoperability and context exchange). The **Model Control Plane** is the enterprise governance/orchestration layer that applies policy, security, and observability.

1 The global AI governance landscape

AI governance is no longer hypothetical. Governments, regulators, and industry groups are actively building frameworks to ensure AI is deployed responsibly based on core principles: transparency, accountability, and adaptability. Understanding this landscape enables organizations to design governance that satisfies compliance requirements while enabling innovation.

Regulatory highlights

- NIST AI RMF (2023–)
  Voluntary U.S. framework defining core functions (Govern / Map / Measure / Manage) to embed trustworthiness across the AI lifecycle.

- European Union – AI Act (2024)
  The first comprehensive, binding law on AI, tiering systems by risk. Unacceptable-risk systems are banned; high-risk systems face strict obligations; minimal-risk systems have light oversight. Enforcement rolls out 2025–2027.

- Council of Europe AI Convention (2024)
  A legally binding treaty emphasizing transparency, accountability, and human rights.

- California's AI Working Group (2025)
  "Trust but verify" approach with whistleblower protections, incident reporting, and oversight mechanisms.

- India's RBI & SEBI Initiatives (2025)
  Sector-specific frameworks for finance and securities focusing on fairness, robustness, and disclosure.

- China's Global AI Proposal (2025)
  Calls for a UN-led governance framework with shared research and openw-source principles.

2 Academic and policy insights

- **Adaptive governance:** Static rules fail when tech evolves. Build feedback loops, continuous monitoring, and the ability to adjust rules as models/agents change.

- **Model registries:** "Passports" for models/agents (architecture, data lineage, intended purpose, resource use, permissions) to enable auditability without full IP disclosure.

- **Multilevel governance:** Governments, corporations, and civil society each play a role—effective oversight is shared oversight.

- **EPIC framework (Education, Partnership, Infrastructure, Community):** Pragmatic pillars for implementing responsible governance.

3 Why agentic AI and MCP demand new thinking

Agentic AI systems

Agentic AI systems shift how digital organizations detect, investigate, and respond to threats. Rather than merely surfacing recommendations, these systems can autonomously execute security playbooks, contain threats, and remediate vulnerabilities—transforming operational speed and scale. This autonomy delivers measurable business value but requires governance that balances that speed with control:

- **Oversight vs. autonomy:** Define which actions require human approval and which can run autonomously.

- **Error handling:** Establish escalation, rollback, and remediation paths.

- **Explainability:** Document decision chains so actions can be audited and corrected.

Model Context Protocol (MCP)

MCP standardizes **interoperability and context exchange** between models, tools, and agents for collaboration. For security and operations teams, MCP enables unified visibility across disparate systems, allowing agents to share threat intelligence, coordinate responses, and maintain consistency across complex cloud environments. Key capabilities include:

- **Structured context sharing:** Consistent formats for prompts, tools, and state.

- **Composable systems:** Easier integration of new tools/models without bespoke glue.

- **Policy-aware messaging:** Context can carry labels/metadata that the control plane can validate/enforce.

Important: MCP **does not** replace governance. It enables interoperable communication under policies enforced by the **Model Control Plane**.

Together, agentic systems and MCP move governance from reactive to architectural: MCP connects the parts; the **control plane** governs their behavior.

4 Industry frameworks and tools

Organizations can draw from:

- **COBIT for AI Governance (2025):** Extends an established IT governance model to AI.
- **Generative AI Governance Principles:** Emphasize transparency, predictability, accountability.
- **Palo Alto Networks and WEF Guidance:** Visibility, monitoring, and proactive controls as foundations.

These are building blocks; they must be adapted to AI-native, agent-centric environments.

5 Sumo Logic AI governance blueprint

The Sumo Logic AI governance blueprint provides a 12-pillar framework for governing agentic AI systems in security and operations environments. This framework enables digital organizations to deploy AI with confidence—unifying visibility, embedding policy enforcement, and moving from reactive compliance to proactive readiness.

1. **Risk-tiered mapping** – Classify systems by risk and map to security playbooks. High-risk agentic actions (e.g., containment workflows) require human-in-the-loop validation and enhanced monitoring; lower-risk systems use lighter controls.

2. **Adaptive oversight** – Embed continuous feedback loops, evaluation metrics, and drift monitoring into the control plane. Include performance/fairness reviews, alignment testing, and traceable policy updates as capabilities evolve.

3. **Model/Agent registries** – Maintain internal "passports" for models and agents: training lineage, data sources, permissions, allowed operations, and operational contexts. Enable rapid revoke/quarantine of compromised assets.

4. **Lifecycle guardrails** – Secure design → train → deploy → monitor → retire. Apply input/output filtering, ethical boundaries, and automated rollback/sandboxing if agents misbehave. Capture real-time telemetry, confidence signals, and anomalies.

5. **Agent governance** – Formalize prompt → action → escalation → rollback workflows. For SecOps, prevent agents from closing or suppressing critical alerts without human approval.

6. **Model Control Plane enforcement** – Use the Model Control Plane as the enforcement layer for all AI use. Route every model or agent invocation through it for authentication, authorization, policy application, and logging. Unified policy means unsafe behavior can be detected, blocked, or isolated with full observability.

7. **MCP Integration (Interoperability)** – Use MCP to standardize context exchange among agents, models, and tools. Carry policy labels/metadata in the context envelope so the control plane can validate and enforce them consistently.

8. **Incident capture and reporting** – Comprehensive telemetry, AI observability dashboards, and continuous red-team exercises against agents and control-plane components. Establish structured incident reporting and responsible disclosure channels.

9. **Evaluation and continuous assurance** – Ongoing evaluation (benchmarks, adversarial sims, bias/drift detection, hallucination analysis, alignment scoring). Feed results into control-plane dashboards and risk tiers to close the loop.

10. **Cross-functional oversight** – Involve security, data science, compliance, and operations—plus red-teamers, incident responders, and ethics reviewers—to scrutinize both agents and the control plane.

11. **Ethical principles** – Build fairness, transparency, and explainability into design and operations. For SecOps, ensure explainable playbooks and traceable agent actions.

12. **Education and culture** – Train leaders and analysts to "trust but verify." Interactive environments plus observability dashboards build confidence and literacy.

6 Implementation roadmap

Phase one – Foundations: Stand up registries, control-plane policy enforcement, and risk mapping. Launch initial red-team exercises against agents and control-plane surfaces.

Phase two – Scaling: Add adaptive oversight, automate approvals/workflows, and deepen incident capture. Deliver agent-level observability dashboards (Sumo Logic Dojo AI).

Phase three – Validation: Conduct external audits, map to multi-jurisdictional requirements, and prove controls. Deploy guardrail automation so the control plane can automatically halt/sandbox policy-breaking agents.

Phase four — Continuous Evolution: Maintain continuous evaluation

7 Why governance is a business advantage

- **Builds trust:** Customers want safe, secure, and fair AI.
- **Reduces risk:** Minimizes fines, reputational harm, and operational disruption.
- **Accelerates innovation:** Confidence to scale beats endless pilots.
- **Differentiates:** Transparent, auditable governance wins deals.

Conclusion

Effective AI governance is no longer optional—it is a competitive necessity for digital organizations operating in an AI-driven world. As agentic AI systems become central to security and operations, governance must be architected into every layer: from standardized interoperability through MCP to unified policy enforcement through control planes to continuous monitoring and adaptation.

Organizations that embed governance into their AI operations gain decisive advantages: faster threat detection and response, greater operational reliability, stronger regulatory compliance, and the confidence to scale AI initiatives across the enterprise.

Sumo Logic helps digital organizations make the AI era secure, fast, and reliable. Built on the Sumo Logic Intelligent Operations Platform, our approach unifies critical security and operational signals, applies intelligent governance at scale, and empowers teams to move from reaction to readiness. By treating governance as both a security safeguard and a business accelerator, digital organizations can unlock AI's full potential—responsibly, transparently, and with confidence.

White paper based on David Girvin's independent research s u

m o

sumo logic

Toll-Free: 1.855.LOG.SUMO | **Int'l:** 1.650.810.8700
855 Main St., Suite 100, Redwood City, CA 94603

www.sumologic.com