Governing Autonomous AI Agents in the Enterprise

Empowering IT and CSO Teams with Centralized Control in a Multi-Agent World

Executive Summary

Leading platforms now deliver autonomous AI agents that support agent-to-agent (A2A) communication and enable efficient, end-to-end workflows across the enterprise. Providers such as Salesforce AgentForce, ServiceNow AI Agents, AWS Bedrock AgentCore, Microsoft Azure AI Agent Service, Google Cloud Vertex AI Agent Builder, Snowflake Cortex Agents, and Atlassian Rovo drive this transformation.

These agents integrate tightly with enterprise identity and access management systems to enforce secure data access. Their ease of creation and customization accelerates adoption across business units. The open A2A protocol, launched in April 2025 by Google and donated to the Linux Foundation in June 2025, is now hosted by the Linux Foundation with broad industry support—including Amazon Web Services, Cisco, Google, Microsoft, Salesforce, SAP, and ServiceNow—and provides standardized interoperability. Agents from different vendors can discover capabilities, exchange context, and collaborate securely.

The growing need for agents to collaborate across platforms creates significant governance challenges. Siloed policy management, manual discovery processes, and fragmented observability expose organizations to risks of unauthorized actions, compliance violations, and operational failures. IT and Chief Security Officer teams must establish centralized control to manage these risks while sustaining innovation.

Introduction: The Rise of Autonomous AI Agents

Major vendors provide accessible tools for building and deploying AI agents, with widespread adoption of the A2A protocol for cross-platform collaboration:

- **Salesforce AgentForce** supports multi-agent orchestration, low-code customization, and native A2A/MCP integration via Agent Exchange.
- **ServiceNow AI Agents** use AI Agent Fabric to orchestrate tasks, with A2A support enabling communication among ServiceNow and third-party agents.
- **AWS Bedrock AgentCore** provides a secure runtime with A2A protocol support for multi-agent workflows.
- **Microsoft Azure AI Agent** Service enables multi-agent systems with A2A and MCP support for cross-runtime collaboration.
- **Google Cloud Vertex AI Agent Builder** uses the Agent Development Kit with A2A for complex multi-agent applications.
- **Snowflake Cortex Agents** unify the analysis of structured and unstructured data, orchestrating tools such as Cortex Analyst and Search.

- **Atlassian Rovo** embeds agents in Jira and Confluence, with emerging A2A interoperability support.

AWS AgentCore	Azure AI	Google Cloud Vertex	Salesforce Agentforce	ServiceNow AI Agents	Snowflake Cortex

The Promise: Multi-Agent Workflows in Action

Agents deliver the greatest value through cross-system collaboration. Real-world examples include:

- Cybersecurity incident response: A network operations center agent detects a threat, opens a ServiceNow ticket, instructs an endpoint agent to isolate the device, and notifies administrators.
- Customer support remediation: A service agent logs an issue, verifies the warranty, and automatically orders a replacement part.
- Financial forecasting: A Salesforce agent projects revenue using historical data from a Snowflake Cortex agent.
- Insurance claims processing: Agents verify coverage, confirm external factors such as weather, detect fraud, and calculate payouts.
- Employee onboarding: HR, IT, and facilities agents coordinate provisioning, setup, training, and access.

These collaborative workflows eliminate manual steps, accelerate resolution, and extend expertise organization-wide. Enterprises using AI agents for inventory management and supply chain optimization have reported up to 30% efficiency gains.

When to Prioritize Centralized Governance: Key Triggers from an IT Perspective

As AI agent adoption accelerates, with surveys indicating that approximately 79–85% of enterprises have implemented or plan to implement AI agents by the end of 2025 and projections of over 1.3 billion AI agents globally by 2028, organizations must recognize the point at which native platform controls become insufficient.

From an IT and security viewpoint, prioritize centralized governance when any of these triggers appear:

- Agent count reaches double digits (10+ agents). Beyond initial pilots, typically 5–10 agents, manual tracking, policy application, and cross-platform coordination become unsustainable. Risks of agent sprawl, redundant capabilities, and ungoverned autonomy escalate rapidly as interactions grow exponentially.
- Cross-platform A2A interactions emerge. Native tools handle intra-environment orchestration well, but A2A across vendors, such as Salesforce agents querying Snowflake Cortex agents, requires a unified approach to discovery, authentication, policy enforcement, and observability to avoid silos and inconsistencies.
- Regulated industry operations. In finance (SOX compliance for financial reporting, ECOA for fair lending requiring traceable decisions), healthcare (HIPAA mandating audit trails for PHI

access and prohibitions on autonomous clinical advice), or public companies (scrutiny over financial agent interactions), immutable logs, decision rationales, and PII redaction are mandatory. The EU AI Act classifies many multi-agent systems as high-risk, demanding lifecycle risk assessments, transparency, and human oversight.

- Audit and compliance demand increase. When agents influence regulated outcomes, such as loan approvals, patient data handling, or financial disclosures, end-to-end traceability, including action rationales and data provenance, becomes essential under frameworks like NIST AI RMF.
- Autonomy impacts sensitive actions. Agents performing approvals, data sharing, or multi-step decisions introduce emergent risks such as cascading errors and bias amplification, necessitating real-time monitoring and escalation controls.

Delaying centralization leads to agent sprawl, shadow deployments, inconsistent policies, and heightened exposure to incidents. Proactive IT leaders conduct early agent inventories and enforce policies prohibiting ungoverned cross-platform A2A until centralized mechanisms are in place.

AI Agent Governance Maturity Model

Assess your organization's readiness:

- Level 1: Siloed Adoption. Agents in isolated platforms; no cross-agent interactions. (Low risk, limited value.)
- Level 2: Initial Collaboration. 5–10 agents with ad-hoc A2A; manual policies. (Emerging risks; governance planning needed.)
- Level 3: Scaling Multi-Agent. >10 agents with frequent A2A; platform-native controls only. (High risk; centralized enforcement required.)
- Level 4: Governed Ecosystem. Unified policies, discovery, and observability across all agents. (Mature; safe scaling enabled.)

If you are at Level 3 or moving toward it, centralized governance is no longer optional.

Native Platform Controls vs. Centralized Enforcement

Aspect	Native Platform Controls	Centralized Enforcement
Policy Management	Siloed per platform; inconsistent application	Universal rules with per-agent refinements; consistent across ecosystems
Discovery	Strong internally; manual cross-platform	Automated, health-monitored discovery via A2A proxy
Enforcement	Platform-specific thresholds/bans	Real-time checks on security, compliance, risk; enterprise-wide blocking
Observability	Logs trapped in one environment	Immutable end-to-end tracing with correlation IDs and rationales
Risk Coverage	Limited to intra-platform interactions	Addresses emergent multi-agent risks like cascades or autonomy failures

The Governance Challenge: Core Limitations of the Current Approach

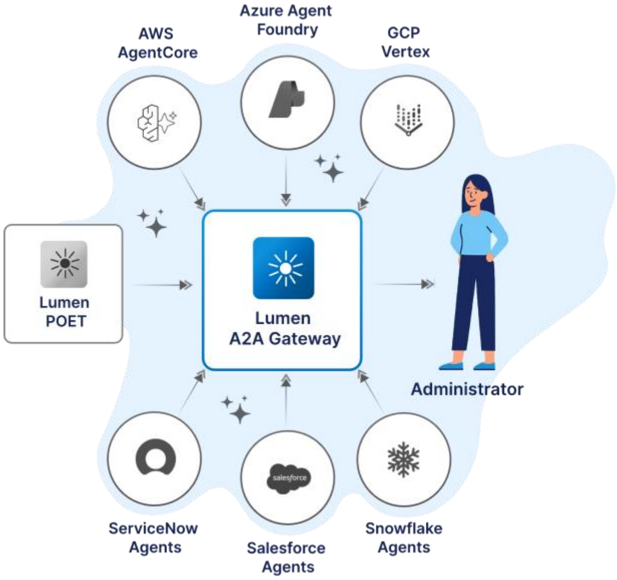
Even with robust native controls, rapid agent proliferation exposes fundamental gaps:

- Discovery and orchestration are strong within a single platform but rely on manual or ad-hoc processes across environments.
- Security, compliance, and business policies, such as "Require approval for invoices over $5,000," must be configured separately in each platform, creating inconsistency and administrative overhead.
- Thresholds, bans, and disabling capabilities remain platform-specific, preventing unified IT oversight.
- Observability and immutable logging are often confined to one environment, making enterprise-wide auditing and incident response difficult.
- Multi-agent autonomy can produce unintended emergent behaviors, coordination failures, or cascading issues without consistent guardrails.

Without centralized governance, organizations face unauthorized actions, regulatory non-compliance (GDPR, HIPAA, SOC 2, EU AI Act), and operational disruptions. Real-world examples highlight the stakes: AI-related deceptive practices and misuse have driven surges in fraud and regulatory actions in 2025. Under the EU AI Act, non-compliance can result in fines of up to €35 million or 7% of global turnover.

IT and security teams have observed rapid adoption across business units. They must now shift from observers to active orchestrators to enable innovation while maintaining control over security, privacy, risk, and operations.

Closing the Gap with Lumen A2A Gateway



Enterprises require a purpose-built zero-trust gateway that enforces governance across the entire multi-agent ecosystem while preserving the strengths of native platforms.

Lumen A2A Gateway from Backflipt serves as this vendor-agnostic enforcement layer for A2A interactions. It provides centralized, real-time control that complements existing investments in

Salesforce, ServiceNow, AWS Bedrock, Azure, Google Vertex AI, Snowflake Cortex, Atlassian, Zendesk, and other platforms.

Executives prioritize three outcomes delivered by Lumen:

- Unified policy enforcement. A single engine applies context-aware policies across security (identity, scopes, prohibitions), compliance/privacy (redaction, DLP), business/risk (thresholds, escalations), and operational assurance (quality gates, alerting), with global rules and per-agent refinements.
- Enterprise-wide visibility. Immutable logs, correlation IDs, recorded rationales, dashboards, and alerts deliver complete traceability and rapid issue resolution.
- Scalable deployment. Automatic agent discovery and registration, health monitoring, and flexible Kubernetes-based deployment (multi-cloud, on-premises, or air-gapped) support thousands of agents in alignment with open standards.

Lumen enables organizations to scale multi-agent automation confidently, reducing risk and administrative burden while accelerating business value.

Lumen A2A Gateway Policy Framework

Lumen A2A Gateway implements governance through a structured, four-category policy framework that enterprises can deploy immediately.

Policies are layered for maximum effectiveness:

- Top-Level (Global). Universal rules applied consistently to all agents across platforms.
- Per-Agent. Targeted refinements based on role, platform, industry, or risk profile.

1. Security Policies

Purpose: Protect agent identity, communication, access boundaries, and prevent abuse.

Top-Level (Global):

- Strong authentication for all A2A interactions (verified credentials, mutual TLS).
- Universal prohibitions on high-risk autonomous actions.
- Rate limiting and anomaly detection.
- Real-time blocking of unverified interactions.

Per-Agent:

- Granular READ/WRITE/OFF-LIMITS scopes.
- Customized "Cannot Do" lists and rate limits.

2. Compliance & Privacy Policies

Purpose: Ensure lawful data handling and regulatory adherence.

Top-Level (Global):

- Automatic redaction of regulated data.
- DLP scanning, data minimization, and retention rules.
- Regional compliance checks.

Per-Agent:

- Industry-specific boundaries (e.g., HIPAA no-diagnosis).
- Role-based access and metadata overrides.

3. Business Policy & Risk Policies

Purpose: Align actions with organizational objectives and risk tolerance.

Top-Level (Global):

- Universal escalation triggers (e.g., legal threats).
- Basic risk scoring.

Per-Agent:

- Transaction thresholds, temporal rules, and action-specific escalations.

4. Operational Assurance Policies

Purpose: Maintain reliability, quality, and monitoring.

Top-Level (Global):

- Immutable logging with timestamps, IDs, and rationales.
- Real-time alerting and performance gates.

Per-Agent:

- Availability constraints, confidence thresholds, session limits, and audit sampling.

This framework, built into Lumen A2A Gateway, addresses current requirements while remaining adaptable to future regulatory and technical evolution (EU AI Act systemic assessments, NIST ongoing risk management).

The Business Case: Risk Reduction and Value Enablement

Centralized governance of autonomous AI agents delivers a compelling risk-adjusted return through downside protection and upside acceleration.

1. Downside Risk Mitigation (Cost Avoidance)

   - Regulatory and Compliance Penalties. The EU AI Act imposes fines up to €35 million or 7% of global annual turnover for non-compliance in high-risk AI systems. This category includes most multi-agent workflows handling personal data, financial decisions, or critical operations. Similar exposure exists under GDPR, HIPAA, and SOX.

   - Incident and Breach Costs. Ungoverned agent interactions can lead to unauthorized actions, data leakage, or cascading failures. The global average cost of a data breach in 2025 is $4.44 million (IBM Cost of a Data Breach Report 2025), with AI-amplified incidents potentially higher due to autonomous propagation and shadow AI adding hundreds of thousands in extra costs.

   - Reputational and Operational Disruption. AI misuse has led to multimillion-dollar settlements and customer losses in recent cases.

2. Upside Value Enablement (Productivity and Scale)

   - Accelerated Workflow Efficiency. Organizations deploying multi-agent systems report 20–40% reductions in process cycle times for use cases such as incident response, claims processing, and customer remediation, translating into labor savings and faster revenue recognition.

   - Safe Scaling of Agent Initiatives. Centralized governance removes barriers to broader adoption, enabling confident deployment of cross-platform projects and unlocking ROI from automation beyond isolated pilots.

- Reduced Administrative Overhead. Eliminating duplicated policy configuration and manual coordination saves significant IT/security resources as agent counts grow.

3. Typical Payback Profile Enterprises implementing centralized agent governance typically achieve break-even within 12–18 months through avoided incidents, reduced compliance effort, and accelerated benefits, with 3–5x return over three years when combining risk avoidance with 25–35% average efficiency gains in targeted workflows.

In regulated sectors, governance investments, often a fraction of a single major incident, protect the balance sheet while enabling the full transformative value of agentic AI.

Conclusion: Preparing for the Agentic Era

Autonomous AI agents, empowered by A2A interoperability, offer transformative potential, but only when appropriately governed. The current fragmented approach leaves critical gaps in control and visibility.

IT and CSO teams must establish centralized mechanisms for discovery, consistent policy enforcement, and end-to-end monitoring. Platforms such as Lumen A2A Gateway address these gaps, turning risk into opportunity while unlocking substantial productivity gains.

Proactive investment now, beginning with an agent inventory and governance roadmap, prevents reactive crises later and positions the organization to scale multi-agent workflows safely and competitively.