EU AI ACT · NIST AI RMF · OWASP

MCP Governance and Compliance:

What the frameworks now require from the agentic action layer.

EU AI ACT | NIST AI RMF | OWASP LLM TOP 10 | OWASP API TOP 10

Regulators, auditors, and cyber insurers are beginning to ask the same question: how does your organization govern the AI systems that take action on its behalf?

MCP servers are the action layer. They are where AI agents get the ability to read data, write to systems, and execute workflows. Every major AI governance framework now has requirements that map directly to this layer — and most organizations cannot yet demonstrate compliance.

2026
EU AI Act high-risk provisions effective

NIST
AI RMF now referenced in federal procurement

$35M+
Average EU AI Act fine for non-compliance

> *"Compliance pressure is not a future risk. The EU AI Act high-risk provisions take effect in 2026. NIST is already referenced in federal procurement.* ***MCP governance is a current requirement, not a roadmap item."***

WHY THIS MATTERS NOW

MCP governance is no longer optional.

For the past two years, AI governance has been primarily a policy conversation, frameworks published, working groups formed, guidance issued. That phase is ending. The EU AI Act's high-risk provisions take effect in 2026. NIST's AI Risk Management Framework is now referenced in U.S. federal procurement requirements. Cyber insurers are building AI governance questions into policy renewal questionnaires.

MCP servers sit at the intersection of every major AI governance requirement. They are the mechanism through which AI agents access data, take actions, and produce real-world consequences. They are where compliance obligations for auditability, access control, human oversight, and risk management must be operationalized.

Most organizations are not ready. The frameworks are clear on what is required. The gap is between documented policy and demonstrable technical control — and MCP servers are the layer where that gap is widest.

Who this brief is for

This brief is written for CISOs, GRC leaders, legal and compliance teams, and procurement stakeholders who need to understand what the major AI governance frameworks require at the MCP layer, and what demonstrating compliance actually looks like in practice. It maps each major framework to specific MCP control requirements, identifies the gaps most organizations currently have, and explains how Salt addresses them.

FRAMEWORKS COVERED

Three frameworks. One control gap.

EU AI Act

The world's first comprehensive AI regulation. Applies to AI systems deployed in the EU or affecting EU residents. High-risk AI systems — including those that interact with enterprise systems and take consequential actions — face mandatory requirements for risk management, auditability, logging, and human oversight. Penalties up to 3% of global revenue or €15M for violations.

NIST AI RMF

The U.S. federal standard for AI risk management. Organized across four functions: Govern, Map, Measure, Manage. Referenced in federal procurement requirements and increasingly adopted by regulated industries including financial services and healthcare. Provides the governance vocabulary that regulators and auditors use to evaluate AI program maturity.

OWASP LLM + API

The practitioner standard for AI and API security risk. OWASP's LLM Top 10 defines the critical risks in large language model deployments. The API Security Top 10 defines the critical risks in the API layer where agents act. Together they cover the full agentic stack — from model to action — and are used by security teams and auditors as a practical evaluation benchmark.

FRAMEWORK: EU AI ACT

What the EU AI Act requires at the MCP layer.

The EU AI Act classifies AI systems by risk level. High-risk AI systems; those that interact with regulated domains, including employment, credit, healthcare, critical infrastructure, and law enforcement face the most stringent requirements. Agentic systems that take real-world actions through MCP servers will almost universally fall into the high-risk classification in enterprise deployments.

EU AI Act

Regulation (EU) 2024/1689 · High-risk provisions effective 2026

Applies to AI systems placed on the EU market or affecting EU residents. High-risk classification triggered by use case, not technology.

REF	REQUIREMENT	MCP IMPLICATION
Art. 9	Risk management system	A continuous risk management process must be established and maintained throughout the AI system's lifecycle. For agentic systems, MCP servers are a primary risk surface — their tool exposure, authentication state, and connected APIs must be inventoried and assessed as part of this process.
Art. 12	Record-keeping and logging	High-risk AI systems must automatically log events at a level sufficient to enable post-hoc investigation. MCP server tool invocations — including agent identity, parameters, and outputs — must be captured. Systems with no MCP-layer logging cannot meet this requirement.
Art. 13	Transparency and information provision	Users and affected parties must be able to understand how the AI system reached its decisions and took its actions. When an agent acts through an MCP server, that action must be traceable. Ungoverned MCP infrastructure breaks the traceability chain.
Art. 14	Human oversight	High-risk AI systems must be designed to allow human oversight and intervention. Agents operating through MCP servers without access controls or behavioral monitoring cannot provide meaningful oversight capability — actions execute before any human review is possible.
Art. 17	Quality management system	Providers must establish a quality management system covering the full AI system lifecycle. MCP server governance — deployment approval, security review, access control, and decommission procedures — must be documented and operationalized as part of this system.

FRAMEWORK: NIST AI RMF

What the NIST AI RMF requires at the MCP layer.

The NIST AI Risk Management Framework organizes AI governance across four functions: Govern, Map, Measure, and Manage. Each function has direct implications for how organizations must treat MCP servers — the action layer that makes AI risk concrete and consequential.

NIST AI RMF

NIST AI 100-1 · Adopted in NIST SP 800-218A and federal procurement

Voluntary framework widely adopted as the de facto U.S. AI governance standard across regulated industries.

REF	REQUIREMENT	MCP IMPLICATION
GOVERN 1	AI risk management policies and procedures	Policies must address risks specific to the organization's AI deployment. MCP servers require explicit policy coverage: deployment approval, authentication requirements, tool access standards, and logging mandates. Generic AI policies that predate MCP architecture are insufficient.
MAP 1	Context and risk identification	Organizations must identify and classify AI risks in context. MCP servers must be inventoried and their risk context documented: what systems they connect to, what tools they expose, which agents have access, and what actions they enable. Undiscovered MCP servers cannot be mapped.
MAP 3	Third-party and supply chain risk	Risks from third-party AI components must be assessed. Vendor-deployed MCP servers — embedded in AI products and integrations — are third-party components with direct access to enterprise systems. They must be evaluated under the same supply chain risk process as other vendor-provided components.
MEASURE 2	Risk metrics and monitoring	Metrics must be defined and monitored for AI risk. For MCP-enabled systems, this includes MCP server inventory completeness, authentication coverage, tool permission compliance rate, and anomalous agent behavior detection. Without instrumentation, these metrics cannot be produced.
MANAGE 1	Risk response and treatment	Identified risks must have defined response plans. MCP risk response requires incident response procedures specific to agentic systems: agent abuse detection, MCP server isolation capabilities, tool invocation forensics, and post-incident audit trail reconstruction.

FRAMEWORK: OWASP

What OWASP requires at the MCP layer.

OWASP provides the practitioner-level risk standards that security teams, auditors, and pen testers use as evaluation benchmarks. Two OWASP lists are directly relevant to MCP security: the LLM Top 10, which covers the AI reasoning layer, and the API Security Top 10, which covers the action layer where agents interact with enterprise systems. MCP servers span both.

OWASP LLM + API Security Top 10

OWASP LLM AI Security Top 10 (2025) · OWASP API Security Top 10 (2023)

Practitioner standards used as security evaluation benchmarks across industries. Regularly referenced in penetration testing scopes and vendor assessments.

REF	REQUIREMENT	MCP IMPLICATION
LLM01	Prompt injection	Attackers manipulate agent inputs to cause unintended MCP tool invocations. MCP servers that do not validate inputs before executing tool calls are directly exposed to this risk. Compliance requires input validation at the tool handler level, not just at the model layer.
LLM06	Sensitive information disclosure	AI systems may expose sensitive data through their outputs or tool behaviors. MCP tools that return unrestricted database records, file contents, or API responses create disclosure risk. Data scope controls at the tool level are required to address this OWASP category.
LLM08	Excessive agency	AI agents with overly broad capabilities can take unintended consequential actions. MCP servers that expose more tools than agents require create excessive agency risk. Tool-level least-privilege policies are the primary control.
API1	Broken object level authorization	APIs that do not enforce authorization at the object level allow agents to access data beyond their intended scope. MCP tools that proxy API calls without enforcing the requesting agent's authorization context expose this risk directly.
API3	Broken object property level authorization	APIs that return excessive data fields expose sensitive properties. MCP tools that return complete database records or API responses without field-level filtering create object property authorization failures at scale when combined with AI agent data processing.
API8	Security misconfiguration	Misconfigured API deployments create exploitable security gaps. MCP servers running without TLS, without authentication, or with default configurations are security misconfigurations by OWASP definition — and are currently the norm rather than the exception in MCP deployments.

THE COMPLIANCE GAP

What most organizations cannot demonstrate today.

The requirements across EU AI Act, NIST AI RMF, and OWASP are specific and consistent. They converge on the same set of controls: inventory, logging, access control, posture assessment, and behavioral monitoring. The gap is not in understanding what is required. The gap is in having the technical controls in place to demonstrate compliance.

01	**No demonstrable MCP server inventory** Every framework requires that organizations know what AI systems they operate. A complete, current, and auditable MCP server inventory is the baseline evidence requirement. Most organizations do not have one — particularly one that includes vendor-deployed and developer-created servers outside the security team's direct oversight.
02	**No MCP-layer audit trail** EU AI Act Article 12, NIST MEASURE 2, and the OWASP security testing standard all require logging sufficient to reconstruct what an AI system did and why. MCP tool invocations — the actual actions agents take — are typically unlogged. Compliance cannot be demonstrated without this instrumentation.
03	**No access control evidence** Every framework requires that access to sensitive systems be restricted to what is necessary. For MCP servers, this means tool-level access controls enforced per agent identity. Most MCP servers expose all tools to all connecting agents with no per-identity enforcement. This is a compliance failure under every applicable framework.
04	**No behavioral monitoring or human oversight capability** EU AI Act Article 14 and NIST MANAGE 1 require that organizations be able to detect anomalous AI behavior and intervene. Runtime monitoring of agent behavior across MCP server interactions is required to satisfy this. Organizations with no MCP-layer monitoring cannot demonstrate this capability.
05	**No third-party MCP server assessment** NIST MAP 3 and EU AI Act supply chain requirements mandate that third-party AI components be assessed. Vendor-deployed MCP servers are third-party components. Most organizations have never assessed them — and many do not know they exist.

HOW SALT CLOSES THE GAP

Compliance evidence built into the platform.

Salt's Agentic Security Platform produces the technical evidence that each framework requires — continuously, not on a point-in-time basis.

Framework	Requirement	Salt capability	Module
EU AI Act	Art. 9 — Risk management	Continuous MCP posture assessment with risk scoring	Posture
EU AI Act	Art. 12 — Logging	Structured tool invocation logs with agent identity	Runtime
EU AI Act	Art. 13 — Transparency	Full audit trail from agent decision to tool execution	Runtime
EU AI Act	Art. 14 — Human oversight	Behavioral anomaly alerts with intervention capability	Runtime
EU AI Act	Art. 17 — Quality mgmt	MCP server lifecycle inventory with governance workflows	Discovery
NIST AI RMF	GOVERN 1 — Policies	Policy enforcement engine for tool access and permissions	Posture
NIST AI RMF	MAP 1 — Risk identification	Automated MCP discovery and risk classification	Discovery
NIST AI RMF	MAP 3 — Third-party risk	Vendor MCP server discovery and posture assessment	Discovery
NIST AI RMF	MEASURE 2 — Metrics	Compliance dashboards and posture trend reporting	Platform
NIST AI RMF	MANAGE 1 — Risk response	Incident detection and MCP-specific response workflows	Runtime
OWASP LLM	LLM01 — Prompt injection	Behavioral anomaly detection on tool invocation patterns	Runtime
OWASP LLM	LLM08 — Excessive agency	Tool-level least-privilege policy engine per agent	Posture
OWASP API	API1 — Object auth	Authorization scope controls on MCP data retrieval tools	Posture
OWASP API	API8 — Misconfiguration	MCP server configuration posture checks and remediation	Posture