NIST AI Risk Management Framework (AI RMF 1.0)

Managing Risks for Trustworthy AI

The NIST AI Risk Management Framework (AI RMF) provides a flexible, structured, and measurable process to address the unique risks associated with Artificial Intelligence systems. Unlike static compliance checklists, the AI RMF emphasizes a lifecycle approach, recognizing that AI risks evolve from design and development through to deployment and operation.

The framework is anchored by four core functions, **GOVERN**, **MAP**, **MEASURE**, and **MANAGE**, that help organizations cultivate a culture of risk management. By implementing these functions, organizations strive to build "Trustworthy AI" systems that are valid, safe, secure, resilient, accountable, transparent, explainable, privacy-enhanced, and fair.

Risk-Based Approach:

- **GOVERN:** Cultivates a culture of risk management and outlines processes, documents, and organizational schemes. It is a cross-cutting function that infuses risk management throughout the organization.
- **MAP:** Establishes context to frame risks, identifying interdependencies and visibility gaps. It ensures organizations understand the context of use and potential impacts before deployment.
- **MEASURE:** Employs quantitative and qualitative tools to analyze, assess, and monitor AI risk. It involves rigorous testing and verification to ensure systems perform as intended.
- **MANAGE:** Prioritizes and acts upon risks, allocating resources to respond to and recover from incidents. It ensures that residual risks are documented and kept within

Characteristics of Trustworthy AI:

- **Secure & Resilient:** AI systems must withstand adversarial attacks (e.g., data poisoning, evasion) and maintain function under stress.
- **Privacy-Enhanced:** Systems must safeguard anonymity, confidentiality, and control to protect the data used to train and operate models.
- **Accountable & Transparent:** Organizations must ensure traceability of AI decisions and maintain access to appropriate levels of information about system operation.

14. NIST AI Risk Management Framework (AI RMF 1.0)

Why It Matters:

Adopting the NIST AI RMF is critical for organizations to deploy AI confidently and responsibly. As AI systems become integrated into critical infrastructure and decision-making, the potential for "black box" failures increases. Following the framework not only mitigates reputational and technical risks but also signals to partners and regulators that the organization is taking a proactive, safety-first approach to AI innovation.

API Security Requirements:

- **Adversarial Defense:** APIs must be protected against machine learning attacks, including model evasion, extraction, and data poisoning.
- **Data Integrity & Privacy:** Enforce strict access controls and encryption on APIs to prevent the leakage of sensitive training data or model intellectual property.
- **Inventory & Governance:** Maintain a comprehensive inventory of all AI system components, including the APIs that serve and connect them, to prevent "Shadow AI."
- **Continuous Monitoring:** Implement real-time logging of API inputs and outputs to detect anomalies and ensure system accountability.

Relevant Sections Referencing APIs:

- **Section 3.3** (Secure and Resilient): Explicitly highlights risks regarding the "exfiltration of models, training data, or other intellectual property through AI system endpoints."
- **GOVERN 1.6:** Mandates that "mechanisms are in place to inventory AI systems," which necessitates continuous observability and mapping of the APIs and agentic interfaces that facilitate their operation.
- **Appendix B:** Notes that existing guidance often fails to account for the "complex attack surface" and "security abuses enabled by AI systems," which frequently target the API layer.

How Salt Security Helps:

Salt Security provides the verifiable technical controls required to align with the NIST AI RMF. By leveraging the **Agentic Security Graph**, Salt establishes a centralized system of record that supports the MAP and GOVERN functions through continuous observability of all AI-to-API and Model Context Protocol (MCP) interactions. To operationalize MANAGE, posture governance enforces strict access controls, eliminating unmanaged risk before deployment. Furthermore, Salt satisfies the *Secure & Resilient and Privacy-Enhanced* characteristics by delivering deep data flow traceability and behavioral threat protection. This blocks logic-based adversarial threats—such as data poisoning or model extraction—providing immutable evidence that technical risks are continuously measured and mitigated.