ALINEDS

ALINEDS GUIDE · AI GOVERNANCE

Secure, Compliant AI for Government

A practical guide to the NIST AI Risk Management Framework — what it is, its four core functions, the added risks of generative AI, and the questions every public agency should ask before adopting AI.

What the NIST AI RMF is

The NIST AI Risk Management Framework (AI RMF 1.0) was released by the National Institute of Standards and Technology on January 26, 2023. It is voluntary, sector-agnostic, and designed to help organizations build trustworthiness into the design, development, use, and evaluation of AI systems. NIST pairs it with a companion Playbook of suggested actions, and in July 2024 published a Generative AI Profile (NIST AI 600-1) addressing the unique risks of generative AI. NIST has signaled the framework is being revised as U.S. AI policy evolves.

Why it matters for the public sector

Government agencies operate under a higher bar: decisions must be accountable, explainable, and defensible to the public, auditors, and oversight bodies. The AI RMF gives agencies a common vocabulary and a repeatable process for adopting AI responsibly — which also makes AI initiatives easier to procure, govern, and justify.

The characteristics of trustworthy AI

The framework defines trustworthy AI as valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with harmful bias managed.

The four core functions

- **Govern** — establish the culture, policies, roles, and accountability for managing AI risk across its lifecycle. Govern is the foundation the other three functions run on.
- **Map** — understand the context: what the AI system is for, who it affects, and where the risks and impacts lie.
- **Measure** — assess, analyze, and track those risks using appropriate methods and metrics, including testing and evaluation.

- **Manage** — act on the risks: prioritize, treat, and monitor them, with plans to respond and recover when issues arise.

The added risks of generative AI

The Generative AI Profile highlights risks amplified or unique with generative systems — including confabulation ("hallucination"), data leakage and privacy exposure, harmful or biased outputs, and provenance and authenticity concerns. Agencies adopting generative AI should treat these explicitly in their Map and Measure work.

Questions to ask before adopting AI

- Does the system keep our data private and out of public model training?
- Can it show its sources and reasoning?
- Is a human in the loop for consequential decisions?
- How is bias identified and managed?
- How will we monitor it in production, and how do we turn it off if it misbehaves?

> **How ALINEDS aligns to the AI RMF.** ALINEDS designs public-sector AI to align with the NIST AI RMF: governed, private data; retrieval-augmented generation that answers from your approved sources with citations; human-in-the-loop review; and governance and monitoring throughout the lifecycle. We treat the AI RMF as an alignment target that shapes how we build — not a certification. Our in-production system, Hermes, applies this pattern today.