NIST AI-600-1

Generative AI Risk Management Profile for Ethical Integration, Deployment, and Governance

---

1. Executive Context

NIST AI-600-1 — *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* — was released on **July 26, 2024,** as a **cross-sectoral profile of the NIST AI RMF 1.0** to address unique risks presented by generative AI systems. It was developed pursuant to U.S. Executive Order 14110 on *Safe, Secure, and Trustworthy Artificial Intelligence* and functions as a companion resource to the foundational AI RMF. The profile is voluntary and meant to help organizations identify and manage generative AI risks in ways aligned with their goals, legal requirements, and best practices.

---

2. Scope and Intent

This profile applies specifically to generative AI systems, including but not limited to large-scale models that produce content such as text, images, audio, or other media. It does not replace AI RMF 1.0 but **extends it by interpreting AI RMF functions (Govern, Map, Measure, Manage) in the context of GenAI risks**, including content integrity, hallucination phenomena, data privacy, and intellectual property considerations, as well as upstream/downstream information risks.

Intended uses include:

- Contextualizing AI RMF principles for GenAI risk patterns
- Helping organizations assess and prioritize GenAI risk controls
- Suggesting actions to align generative AI governance practices with organizational risk tolerance

---

3. Alignment to Ethical AI Integration Strategy

Strategically, NIST AI-600-1 **bridges high-level risk governance and operational controls** for generative AI. Key implications:

- Encourages leadership to treat generative AI risk as part of enterprise risk portfolios
- Advances ethical commitments (transparency, fairness, reliability) into practical risk categories

- Promotes documentation of GenAI risk decisions as governance artifacts that support downstream auditability

This alignment ensures that ethical intent does not remain abstract but becomes embedded in risk frameworks tailored to GenAI's distinctive harm vectors.

---

4. Alignment to Deployment and Lifecycle Controls

The narrative in AI-600-1 maps generative AI risk considerations to lifecycle stages:

- **Govern:** Establish enterprise policies and oversight specific to GenAI risks
- **Map:** Define context of use, system boundaries, and stakeholder impacts for GenAI
- **Measure:** Assess risks such as hallucinations, misinformation, IP leakage, and dual-use vulnerabilities
- **Manage:** Identify controls to mitigate those risks and monitor effectiveness across operations

Suggested actions are mapped to these functions, enabling organizations to integrate GenAI risk controls at critical gates in design, deployment, and operation.

---

5. Governance, Oversight, and Accountability

Governance implications of AI-600-1 include:

- Clear assignment of roles for GenAI risk identification and response
- Policies documenting risk assessment criteria and mitigation thresholds
- Traceable evidence of risk monitoring, incident response, and iterative improvement

Such structures mirror those in AI RMF and support internal and external assurance of GenAI governance.

---

6. Risk Management and Ethical Safeguards

NIST AI-600-1 focuses on risks that are particularly salient for generative AI, including:

- **Hallucinations/confabulations:** False or misleading outputs
- **Content misuse:** Harmful or unsafe content generation
- **Privacy and IP exposure:** Output-related data issues
- **Security vulnerabilities:** Exploitable GenAI behaviors
- **Value chain risks:** Third-party model dependencies

Ethical safeguards are suggested through **risk measurement and treatment actions** such as evaluation protocols, provenance tracking, operational boundaries, and mitigation controls. These are not mandatory, but provide implementers with a structured approach to risk hygiene.

---

7. Strategic Implications for Organizations

Adopting NIST AI-600-1 enables organizations to:

- Elevate generative AI risk management into enterprise risk governance
- Translate abstract trustworthiness goals into practical, documented actions
- Support defensible oversight of GenAI initiatives across business units
- Align generative AI risk posture with broader risk appetite frameworks

Importantly, the profile complements broader standards and helps organizations demonstrate *evidence-based governance practices* even in the face of regulatory or stakeholder scrutiny.

---

8. Relationship to Other Instruments

NIST AI-600-1 functions as a **technology-specific refinement** within the AI governance ecosystem:

- **AI RMF 1.0:** Provides the foundational risk management structure; AI-600-1 tailors it to GenAI risk landscapes
- **ISO/IEC 42001:** Offers management system controls that can encompass GenAI governance
- **ISO/IEC 23894:** Supports risk categorization that can be enriched by GenAI-specific insights
- **ISO 8000:** Ensures data and information quality as a defense against GenAI misinformation and bias
- **ISO/IEC 27001/27701:** Provides security and privacy baselines that mitigate GenAI misuse risks

This layered model ensures that GenAI risk governance does not live in isolation but is part of a holistic AI governance system.

---

9. Why NIST AI-600-1 Matters

Generative AI introduces risks distinct from those of traditional algorithmic systems, including failures in content authenticity, hallucinations, and dual-use vulnerabilities. NIST AI-600-1 matters because it helps organizations systematically:

- Recognize these risks
- Integrate them into existing risk governance structures
- Propose actionable risk management steps tailored to GenAI

As such, it strengthens governance beyond general frameworks and improves institutional readiness for a faster-evolving AI landscape.