NIST AI 600-1 SECURITY RISK MITIGATION IMPLEMENTATION CHECKLIST

The National Institute of Standards and Technology (NIST) AI 600-1, also known as the AI Risk Management Framework (AI RMF), is a framework for understanding and assessing AI security, confidentiality, reliability, and bias risks. Frameworks like the NIST AI RMF offer a structured approach to mitigating these threats. The following checklist details AI risks and mitigation strategies based on the NIST AI RMF controls.

1	Confabulation
☐	Define minimum performance benchmarks.
☐	Review and improve outputs regularly.
☐	Document ML methods, prompt techniques, and risks.
☐	Test across diverse scenarios; benchmark scientifically.
☐	Build output validation into system architecture.
☐	Track and log errors for continual improvement.

2	Dangerous, Violent, or Hateful Content
☐	Gather community input to define unacceptable content.
☐	Build safeguards into system policies and development.
☐	Include SMEs and end-users in testing phases.
☐	Audit training data for harmful material.
☐	Incorporate user feedback to refine content controls.

3	Data Privacy
☐	Align with privacy laws; document training/generated data.
☐	Give users opt-out options and consent controls.
☐	Apply anonymization and privacy-enhancing technologies.
☐	Monitor for PII exposure; test privacy throughout lifecycle.

4	Environmental Impacts
☐	Identify and consult affected ecosystems.
☐	Track energy, water use, emissions across lifecycle.
☐	Publish environmental impact documentation.
☐	Invest in offsets or green energy where applicable.

5	Harmful Bias or Homogenization
☐	Ensure compliance with fairness laws and standards.
☐	Reduce bias via diverse teams, impact assessments, and data curation.
☐	Monitor AI in production; audit and retrain with debiased data.
☐	Use diverse data sources and track limitations transparently.

6	Human-AI Configuration
☐	Control access to AI interfaces and logs.
☐	Log system changes and human interactions securely.
☐	Conduct regular penetration tests on AI-human systems.

7	Information Integrity
☐	Validate outputs and define expected outcomes.
☐	Use automated checks and manual review.
☐	Restrict access to critical AI components.
☐	Detect manipulations with watermarking and authentication.

8	Information Security
☐	Apply multi-layered defenses (tech, admin, physical).
☐	Enforce MFA, least-privilege access, and account controls.
☐	Encrypt data at rest, in transit, and ideally in use.
☐	Continuously scan and patch vulnerabilities.

9	Intellectual Property
☐	Set clear IP policies for AI development and use.
☐	Avoid training on copyrighted content without permission.
☐	Review outputs for IP violations regularly.

10	Obscene, Degrading, and Abusive Content
☐	Define AI acceptable use policies.
☐	Use filters and monitoring tools for offensive content.
☐	Allow users to easily report concerns.

11	Value Chain and Component Integration
☐	Vet all vendors and components.
☐	Set security requirements across the AI value chain.
☐	Implement a vulnerability management process.
☐	Prepare incident response plans and playbooks.

>	Next Steps

Need help getting ahead of AI risks and compliance challenges? Contact Echelon Risk + Cyber to schedule an AI Governance assessment to ensure you're prepared to manage AI risks before they impact your business.