Apptega

GUIDE

NIST AI RMF 100-1

Compliance Guide

For organizations that design, develop, deploy, procure, or use AI systems and need a practical path to implement the NIST AI Risk Management Framework (AI RMF 1.0)

Current as of August 20, 2025

Apptega

Apptega

1. Introduction

This guide helps organizations become compliant with the NIST AI Risk Management Framework (AI RMF 100-1) by converting its four functions—GOVERN, MAP, MEASURE, MANAGE—into concrete actions, owners, parameters, acceptance criteria, and evidence. It is designed for teams that build, operate, or procure AI (classical ML, deep learning, and generative AI).

1A. Beginner Quick-Start (First 30—90 Days)

Days 1—15 — Program stand-up

- Appoint: Executive AI Risk Owner, Responsible AI Lead, Model/Use-Case Owner(s), Data Owner, TEVV Lead, Security/Privacy Lead, Legal/Ethics Counsel, Third-Party/Procurement Lead.
- Define the AI authorization boundary and draft data/model/workflow diagrams.
- Publish an AI Policy and Risk Appetite statement; create Use-Case Intake and Review Gates.
- Open an AI Registry (AIBOM entries per model/use case).
- Draft AI RMF Profile (v0) for the first use case.

Days 16—45: Implement & tailor

- Complete MAP for each prioritized use case (context, data rights, affected parties, harms, constraints).
- Stand up MEASURE baselines: metrics (performance, safety, robustness, privacy, fairness, explainability), test datasets, bias slices, drift signals, red-team plan (for GenAI).
- Define MANAGE controls: guardrails, kill-switch, rollout/rollback, monitoring plan, incident taxonomy.

Days 46—90: Validate & evidence

- Execute TEVV and red-team exercises; fix gaps.
- Populate the Evidence Register.
- Conduct a tabletop (e.g., prompt injection → harmful output; privacy leak).
- Update AI RMF Profile (v1); record residual risks and owners in the Remediation Register.

2. Scope & Alignment

Scope. All people, processes, models, datasets, prompts, code, pipelines, inference services, integrations, and third parties that create, receive, maintain, transmit, or can materially affect AI system behavior or its impacts on individuals, organizations, or society. Include labeling, training, evaluation, deployment, monitoring, and retirement.

Key definitions

- **AI system:** An engineered system generating outputs (predictions, recommendations, content, or decisions).
- **TEVV:** Testing, Evaluation, Verification, and Validation across the lifecycle.
- **AI RMF Profile:** Tailored context, outcomes, constraints, and priorities for a use case; updated at gates and material changes.
- **Trustworthiness characteristics:** validity/reliability; safety; security/resilience; accountability/transparency; explainability/interpretability; privacy-enhancement; fairness (harmful bias managed).

Governance & roles (minimum set)

Executive Risk Owner; Responsible AI Lead; Model/Use-Case Owner; TEVV Lead; Security/Privacy Lead; Legal/Ethics; Procurement/TPRM Lead.

3. Standards & Practices (Mapped to AI RMF Functions)

For each family: Intent • Minimums • Implement (Procedural/Technical/Contractual) • Evidence • Acceptance Criteria • Common Failures • Internal QA Plan • Documentation tie-in.

GOVERN (cross-cutting)

Intent: Establish accountable, transparent, and ethical AI risk management.

Minimums: AI policy; roles & segregation (builders vs independent evaluators); risk appetite; intake; review gates; registry; documentation standards.

Implement — Procedural:

- **Use-Case Intake Checklist (pre-work):** objective; stakeholders/affected parties; legal/contract drivers; data rights basis; risk tier (matrix below); monitoring owner; rollback owner.
- **Pre-Launch Readiness Gate (must have):** current AI RMF Profile; data rights/provenance attestation; TEVV report(s) vs thresholds; **independent evaluator** sign-off (per tier); guardrails configured & tested; monitoring plan + rollback plan proven in a drill; completed **registry/AIBOM** entry; residual risks documented and accepted by the Executive AI Risk Owner.
- **Post-Launch Gate:** monitoring signals online; alert thresholds verified; first review date set; rollback artifacts validated.
- **Risk Tiering Matrix (3×3 Likelihood×Impact):**

Tier	Example drivers	Required independence	Monitoring cadence	Drills
Low	Internal users, reversible, low harm	Internal peer review	Monthly metrics	Semiannual incident drill
Medium	External users/sensitive data, moderate impact	Independent team (same org)	Biweekly metrics + bias slices	Quarterly drill
High	Safety-critical/rights-impacting/broad reach	Independent org/unit + executive sign-off	Weekly metrics + red-team quarterly	Quarterly drill + annual full rollback

Implement — Technical: Access controls & audit trails; **AI Registry / AIBOM** entry for each model (minimum fields below); reproducible training; secure MLOps; model & dataset versioning.

- **AIBOM minimum fields:** use-case ID; model/version; datasets/versions (train/val/test); prompts/policies; guardrails; evaluation reports/metrics; known failure modes; deployment dates; rollback artifacts; owners; supplier/dependency versions; license/usage limits.

Implement — Contractual: Supplier agreements with responsibilities, data usage limits, evaluation/monitoring rights, incident notice timelines, transparency on updates, and flow-downs.

- **Evidence:** AI policy; role matrix; risk appetite; intake records; gate minutes; registry entries; supplier responsibility matrices.
- **Acceptance Criteria:** Every AI use case has an owner, a **current profile**, documented gates, and a monitoring plan; **independent evaluation** is recorded for medium/high tier.
- **Common Failures:** No gatekeeping; unclear ownership; no registry; reliance on vendor claims without evaluation rights.
- **Internal QA Plan:** Quarterly audit of 3 recent launches for gate adherence, registry completeness, and documentation quality.
- **Documentation tie-in:** AI Policy; Risk Appetite; Review Gate SOP; Model Registry Standard; Supplier Responsibility Matrix.

MAP

Intent: Establish context, rights, affected parties, foreseeable harms, and constraints for each AI system.

Minimums: Use-case profile; operating environment; legal/contract drivers; **data inventory & provenance**; foreseeable harms; success/exit criteria.

Implement — Procedural: Use-case profiling (goals, context, stakeholders, environment); Data Rights Decision Rules for each dataset (lawful basis/consent or license; allowed purposes—training vs inference; retention; geography; re-sharing; "no-train"/derivative limits; DSAR/record request handling path where applicable); Affected Parties & Harm Scenarios (direct users, bystanders, inference subjects, downstream decision targets; harms across safety, privacy, fairness, security, misinformation, accessibility) with a mitigation owner for each harm.

Implement — Technical: Data lineage and quality checks; sandboxing; synthetic data flags/labels; environment constraints (domain, cultural, accessibility); dataset version control.

Implement — Contractual: Data license and use constraints; model/content licensing; downstream flow-downs.

- **Evidence:** Approved profiles; data maps; rights/consent attestations; harm analysis; environment constraints.
- **Acceptance Criteria:** Profile approved; rights validated; harms documented with mitigation owners; success/exit criteria defined.
- **Common Failures:** Vague objectives; missing data rights; harms not documented; constraints not operationalized.
- **Internal QA Plan:** Review 2 use-case profiles/quarter for completeness and clarity.
- **Documentation tie-in:** AI RMF Profile; Data Inventory & Lineage Standard; Harm Analysis SOP.

MEASURE

Intent: Quantify behavior and risks; validate assumptions; measure uncertainty, drift, and bias.

Minimums: Metric suite for performance, safety, robustness, privacy, fairness, explainability; TEVV plans; independent evaluators; reproducible results.

Implement — Procedural: Metric definitions & thresholds; TEVV schedules (pre-launch, periodic, event-driven); independence rules by tier; versioned evaluation reports; red-team plan (esp. for GenAI and High-tier).

Implement — Technical: Hold-out and realistic datasets; **bias & subgroup analyses**; OOD checks; uncertainty estimation; canary/online evals; **GenAI Safety Battery** (prompt injection, jailbreaks, harmful content classes, PII leakage, refusal/override behavior, guardrail efficacy); privacy tests (e.g., membership inference).

Implement — Contractual: Evaluation rights; access to logs/telemetry; model update disclosures and compatible eval hooks.

- **Metric Suite & Thresholds (default targets—tailor per use case/tier):**
  - **Performance:** task-appropriate (e.g., AUROC ≥ 0.90 or macro-F1 ≥ 0.85 Medium; AUROC ≥ 0.92 or macro-F1 ≥ 0.90 High).
  - **Fairness:** parity gap ≤ 5 percentage points across defined slices or documented justification + mitigation plan.
  - **Robustness:** accuracy drop under agreed perturbations ≤ 10% (Medium) / ≤ 5% (High).
  - **Privacy:** no confirmed membership-inference leakage at α=0.05; zero PII echo in targeted generative spot-checks.
  - **Explainability:** method declared (e.g., SHAP, counterfactuals) and reproducible on 5 representative cases.
  - **Uncertainty/Calibration:** ECE ≤ 0.05 (High); confidence gating routes low-confidence cases to human review.
- **Evidence:** TEVV plans; metric dashboards; bias/robustness/privacy reports; red-team reports; uncertainty & drift analyses.
- **Acceptance Criteria:** Metrics meet thresholds with limitations documented; known failure modes & residual risks accepted by the Executive AI Risk Owner; evaluations reproducible by an independent party.
- **Common Failures:** Accuracy-only focus; missing bias slices; no robustness/privacy testing; non-reproducible evals.
- **Internal QA Plan:** Independently reproduce one evaluation/quarter; spot-check three bias slices across different attributes.
- **Documentation tie-in:** TEVV Plan; Metrics Catalog; Red-Team Playbook; Evaluation Reports.

MANAGE

Intent: Treat risks; implement controls; deploy safely; monitor; respond to incidents and change.

Minimums: Risk treatment decisions; guardrails; change management; monitoring; incident response with AI-specific taxonomy.

Implement — Procedural: Risk treatment register; Human-in-the-Loop design rules (when review is required, what evidence reviewers see, how overrides are logged); AI Incident Taxonomy (privacy leak, harmful/unsafe output, safety violation, model theft/tamper, policy drift, abuse/attack) with severities, detection signals, owner, notification commitments; rollout/rollback & kill-switch procedures

Implement — Technical: Input validation; output filtering; rate/abuse controls; PII scrubbing where applicable; policy/usage enforcement; uncertainty-aware UX; provenance/watermark checks where supported; automated drift detection; retraining triggers; versioned configs and rollback artifacts

Implement — Contractual: SLA/OLA for model quality/safety; incident notice and remediation timelines; transparency on updates; provenance commitments where available.

- **Guardrails "must be on" pre-launch:** input validation; output filtering; abuse/rate limits; PII scrubbing (as applicable); uncertainty cues & deferral; kill-switch/rollback; provenance/watermark check (if supported). Verified in the pre-launch gate.
- **Evidence:** Risk treatment decisions; guardrail configs & tests; monitoring dashboards; incident logs & AARs; rollback test results; change approvals.
- **Acceptance Criteria:** Controls active & tested; clear rollback path; monitoring detects defined conditions; incidents handled within SLA; material changes re-evaluated before release.
- **Common Failures:** Launch without guardrails; no rollback plan; no monitoring for drift/abuse; undefined incident classes.
- **Internal QA Plan:** Pre-launch control verification checklist and simulated rollback; quarterly incident drill per tier.
- **Documentation tie-in:** Risk Treatment Register; Deployment & Rollback SOP; Monitoring Runbooks; Incident Response Plan.

4. AI Authorization Boundary & Data / Model / Workflow Mapping

Define components, trust zones, datasets, model artifacts (checkpoints, embeddings), pipelines (training, fine-tuning, inference), human workflows, and third-party integrations. Map data/prompt ingress and output egress, safety-/privacy-critical components, and change authorities. Maintain diagrams under change control.

5. Documentation Set — Canonical Narratives & AI RMF Profile

Create cohesive narratives proving conformance to GOVERN, MAP, MEASURE, MANAGE: overview & mission; roles & accountability; data provenance; trustworthiness priorities; TEVV approach; risk treatment & guardrails; deployment & monitoring; incident & change handling; dependency management.

Maintain an AI RMF Profile per use case (context, priorities, outcomes, constraints, acceptance criteria, selected measurements) and keep it current at lifecycle gates and material changes.

Apptega

6. Applicability & Acceptance Criteria Mapping (to AI RMF Functions)

Maintain a record of each selected outcome with applicability, implementation summary, parameters, measurable acceptance criteria, inheritance (if any), and evidence link.

Excerpt:

Function	Outcome Area	Applicable	Implementation Summary	Parameters	Acceptance Criteria	Inheritance	Evidence
GOVERN	Ownership & Gates	Y	Policy, role matrix, gate checklist	Gate set, risk tiers	Gate minutes; registry entry	Some vendor	Policy/Minutes
MAP	Context & Harms	Y	Use-case profile; data rights; harm list	Risk appetite	Profile approved; rights validated	None	Profiles
MEASURE	Bias & Robustness	Y	Subgroup metrics; adversarial probes	Thresholds	Metrics ≥ thresholds; residual risk accepted	None	Eval Reports
MANAGE	Monitoring & IR	Y	Alerts; rollback; IR playbooks	SLA	MTTR ≤ SLA; drills complete	Some vendor	Runbooks/Logs

7. Program Parameters (Organization-Defined Settings for AI)

- **Governance:** Review gates (design, pre-launch, post-launch); minimum documentation at each gate; independence level by risk tier.
- **Identity & Change:** Authorized roles for training/fine-tuning; prompt/policy edits; guardrail changes; audit trail retention.
- **Data:** Allowed sources; consent/license checks; retention; synthetic data disclosure; dataset versioning.
- **Security & Privacy:** Threat model; red-team frequency (High: quarterly; Medium: semiannual); secrets handling; PII handling; privacy testing cadence.
- **Trustworthiness Targets (defaults):** thresholds listed in MEASURE; parity gap ≤ 5 pp unless justified; ECE ≤ 0.05 for High tier.
- **Monitoring SLOs:** drift/bias alert MTTD ≤ 24h (High) / ≤ 72h (Medium); incident MTTR ≤ 24h (High) / ≤ 72h (Medium).
- **Re-evaluation triggers:** any material change to data, model, prompts/policies, guardrails, dependencies; adverse incident; threshold breach.
- **Third-Party:** Evaluation rights; incident notice ≤ 72h; update transparency; provenance/watermark stance; flow-downs to subcontractors.

Apptega

8. Evidence Register

Artifact	Function(s)	Location/Path	Owner	Format	Retention
AI RMF Profiles (per use case)	GOV/MAP/ MES/MAN	AI/Profiles/	RAI Lead	PDF/MD	Current + 1 yr
Policy, Roles, Gates & Minutes	GOVERN	AI/Governance/	RAI Lead	PDF	3 yrs
AI Registry / AIBOM Entries	GOVERN	AI/Registry/	Model Owner	CSV/MD	Current + 1 yr
Data Inventory & Lineage	MAP	AI/Data/	Data Owner	CSV/PDF	3 yrs
Data Rights & Provenance Attestations	MAP	AI/Data/Rights/	Legal/Data	PDF	3 yrs
Harm Analysis & Affected Parties	MAP	AI/Risk/	Model Owner	PDF	3 yrs
TEVV Plans & Metrics Catalog	MEASURE	AI/TEVV/	TEVV Lead	PDF/MD	3 yrs
Evaluation/Bias/Robustness/ Privacy Reports	MEASURE	AI/Evals/	TEVV Lead	PDF/CSV	3 yrs
Red-Team & Safety Test Reports	MEASURE/ MANAGE	AI/RedTeam/	Sec/TEVV	PDF	3 yrs
Guardrail Configs & Approvals	MANAGE	AI/Controls/	Model Owner	JSON/PDF	2 yrs
Monitoring Dashboards & Alerts	MANAGE	AI/Mon/	SRE/Model Ops	PNG/CSV	1–2 yrs
Incident Logs & AARs	MANAGE	IR/AI/	IR Lead	PDF	3 yrs
Pre-Launch Gate Checklist & Minutes	GOV/ MANAGE	AI/Governance/ Gates/	RAI Lead	PDF	3 yrs
Rollback Test Evidence	MANAGE	Change/AI/Rollback/	Ops	PDF/CSV	2 yrs
Supplier Responsibility Matrices	GOV/ MANAGE	TPRM/AI/	Procurement	PDF/XLSX	Active + 1 yr
Change Tickets & Approvals	MANAGE	Change/AI/	Ops	PDF/CSV	2 yrs

9. Continuous Monitoring

- **Daily:** Ingestion/log health; safety/abuse alerts; guardrail violations; unexpected output classes; drift signals; failed eval jobs.
- **Weekly:** Bias slice deltas; prompt/attack telemetry review (GenAI); secrets/cert checks; access change deltas.
- **Monthly:** Threshold review; retraining candidate identification; dependency updates (models/SDKs); supplier notices.
- **Quarterly:** Full evaluation rerun for High-tier use cases; red-team exercise; management KPI review.
- **Automation:** Auto-create tickets for drift, bias exceedances, abuse spikes, missing logs, expired exceptions, or unreviewed alerts; track to closure.

10. Remediation & Risk Acceptance Workflow

- Identify → Log (owner, severity, due date, milestones) → Treat (control changes, retraining, data fixes, UI/UX changes, vendor updates) → Verify with evidence → Report to governance.
- Risk acceptance only by the Executive AI Risk Owner, with explicit expiry, conditions, and follow-up measurements. Keep residual risks visible in the AI RMF Profile and governance reviews.

11. External Providers & Model/Service Dependencies

Inventory foundation models, hosted APIs, labeling vendors, data brokers, cloud/MLOps, and safety tooling. For each: shared responsibilities, evaluation rights, update transparency, logging access, incident notice timelines, provenance/watermarking stance, data usage limits and retention. Keep on-file evaluations of provider safety/robustness claims and ensure contractual flow-downs.

12. Training & Awareness (Role-Based)

Tracks for general workforce, business owners, developers/DS/ML engineers, TEVV evaluators, red-teamers, SRE/Model Ops, legal/ethics, and leadership. Onboarding before access; annual refresh; targeted refreshers after incidents. Metrics: completion ≥98% and declining repeat findings.

13. Change Management for AI Lifecycle Impact

Gate changes to datasets, models, prompts/policies, guardrails, and dependencies. Re-run material TEVV; update profile and registry; obtain gate approval before production; maintain rollback artifacts and versioned configs.

14. Evidence Sampling Plans (Internal QA)

- **Governance:** Sample 3 launches for gate compliance and **registry/AIBOM** completeness.
- **MAP:** Inspect 2 profiles for data rights/provenance and harm analysis completeness.
- **MEASURE:** Reproduce 1 evaluation end-to-end; review 3 bias slices and 1 robustness/abuse test.
- **MANAGE:** Simulate 1 rollback and 1 AI-specific incident per quarter for High-tier use cases; verify monitoring detects seeded issues.
- **Third-Party:** Review 2 supplier responsibility matrices and their most recent safety update.

15. Common Pitfalls

Accuracy-only mindset; no bias/fairness or robustness testing; unclear ownership; launching without guardrails/rollback; no monitoring for drift/jailbreak/abuse; missing data rights; evaluators not independent; results not reproducible.

16. Quick Reference Summary

Area	Core Artifacts	Examples
GOVERN	Policy; roles; gates; registry	Gate minutes; AIBOM entries
MAP	Profiles; data lineage; rights; harms	Consent/licenses; affected parties
MEASURE	TEVV plan; metrics; reports	Bias slices; robustness; privacy tests; red-team
MANAGE	Controls; monitoring; IR	Guardrails; rollback; alert SLAs
Third-Party	Responsibility matrices	Evaluation rights; incident notice; transparency

17. Self-Assessment & Leadership Attestation

Use status: **Compliant (C) / Partially Compliant (PC) / Not Compliant (NC) / Not Applicable (N/A).**

Tracker (illustrative):

Function/Outcome	Status	Rationale (for N/A or gaps)	Evidence Link	Owner	Action ID
GOVERN – Ownership & Gates	C	All gates complete	AI/Governance/ Minutes_2025Q3.pdf	RAI Lead	—
MEASURE – Bias & Robustness	PC	Two slices below threshold	AI/Evals/2025Q3.pdf	TEVV Lead	ACT-2025-019
MANAGE – Monitoring & IR	C	Alerts & drill successful	AI/Mon/Runbook.pdf	Ops Lead	—

Leadership checks (explicit):

- ☐ Registry/AIBOM entry complete
- ☐ Independent evaluation completed (per tier)
- ☐ Pre-launch gate passed; residual risks accepted by Executive AI Risk Owner
- ☐ Monitoring + rollback tested and operational

18. References & Resources

NIST AI RMF 1.0 (NIST AI 100-1)
https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

NIST AI RMF Landing Page
https://www.nist.gov/itl/ai-risk-management-framework

NIST AI RMF Playbook
https://airc.nist.gov/airmf-resources/playbook/

ISO/IEC 42001:2023 (AI Management System)
https://www.iso.org/standard/81230.html

ISO/IEC 23894:2023 (AI Risk Management)
https://www.iso.org/standard/77304.html

NIST SP 800-218 (Secure Software Development Framework)
https://csrc.nist.gov/publications/detail/sp/800-218/final

NIST Privacy Framework
https://www.nist.gov/privacy-framework

OWASP Top 10 for LLM Applications
https://owasp.org/www-project-top-10-for-large-language-model-applications/

ENISA AI Threat Landscape
https://www.enisa.europa.eu/publications

U.S. Executive Order 14110 on Safe, Secure, and Trustworthy AI
https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Apptega

Apptega Product Features

16+ Security Frameworks

One-Click Reporting

Automated Alerts & Notifications

API & Application Connectors

Automated Framework Crosswalking

Real-Time Compliance Scoring

Restricted Auditor View

Single Sign-On Connectivity

Policy & Plan Templates

Automated Risk Assessments

Document Repository for Artifacts

Multi-Tenant Environment

Apptega

About Apptega

A perennial G2 leader across various cybersecurity categories, Apptega is the cybersecurity and compliance platform purpose-built for security providers who are growing lucrative security practices, creating stickier customer relationships, and winning more business from competitors.

With Apptega's end-to-end platform, IT providers can:

1. Build continuous compliance offerings that set them apart, retain customers, and grow their ARR.
2. Build and manage world-class cybersecurity and compliance programs for their clients.
3. Increase the capacity and efficiency of their existing team so they can service 2–3× more customers.
4. Assess, measure, and manage their own security programs against a standard set of industry-leading frameworks.

To learn more, visit apptega.com

Visit apptega.com