# Day 9


## NIST AI Risk Management Framework (AI RMF):

1. **Core Function: GOVERN**

It answers one key question:

**"How should an organization manage AI risks in a structured, responsible way?"**

It establishes policies, roles, responsibilities, and decision-making rules that guide safe AI usage. Think of GOVERN as the "brain" of AI risk management.

**Establishing Governance Policies:**

Governance policies define exact rules for how AI should be developed, tested, deployed, and monitored. Policies include: AI development standards, Security standards, Risk management rules, Ethical guidelines, Compliance policies.

Purpose:

- Everyone follows the same rules
- No unsafe or uncontrolled AI usage
- AI development becomes predictable and safe

**Leadership Responsibilities:**

Leaders ensure AI is used responsibly and safely. Leaders set the direction and enforce responsible AI behaviour.

Leaders include:

- CISO
- CIO/CTO
- AI Governance Board
- Compliance Officers
- Business Heads

Their key responsibilities:

a. Approve AI policies
b. Allocate resources
c. Oversee high-risk AI systems
d. Ensure cross-team coordination
e. Uphold accountability

**Model Lifecycle Governance:**

Lifecycle governance ensures the entire AI lifecycle is controlled, from the beginning to the end. This includes:

1. Data acquisition governance
2. Model training governance
3. Evaluation governance

4. Deployment governance
5. Monitoring governance
6. Retirement governance

Purpose of lifecycle governance:
✓ No “shadow AI”
✓ No undocumented model
✓ No risky deployment
✓ Every step is reviewed & controlled

**Roles & Accountability Definitions:**

This ensures every person knows exactly what they must do and what they are responsible for. Often organizations use RACI matrices to assign roles.

Common roles in AI governance:

a. Developers / ML engineers
b. Security teams
c. Compliance / Legal
d. Data teams
e. Executives
f. AI product owners
g. AI governance Board

**Integrating AI Governance with Cybersecurity Governance**

AI governance must not work separately from cybersecurity governance—they must work together.

Cybersecurity protects:

- networks
- systems
- data

AI governance protects:

- models
- datasets
- AI decisions
- AI behaviors

Why integration is important:

✓ AI attacks (like prompt injection, model extraction) behave like cyber attacks
✓ Security teams understand threats
✓ AI teams understand models
✓ Together they can defend effectively

Integration happens in several ways:

a. **Shared risk management pipelines:** AI risks added into cybersecurity risk registers.
b. **Shared incident response:** AI incidents treated as security incidents.
c. **Shared monitoring tools:** Security logs + AI logs integrated into SIEM/SOAR.

d. **Common access control systems:** Identity and Access Management (IAM) applies to model endpoints.
e. **Unified policies**
    - secure coding + secure model training
    - vulnerability management extended to AI vulnerabilities

Basically,

GOVERN = the foundation of the AI Risk Framework. It sets the rules, roles, responsibilities, and approval system for safe AI use.

You learn:

- How to create AI policies
- What leaders must do
- How to control the entire AI lifecycle
- Who is responsible for what
- How AI governance connects to cybersecurity
- GOVERN = "Create rules + assign responsibilities + oversee AI safely."

## 2. Core Function: MAP

MAP = Understand your AI system in context, its risks, and where gaps exist. While GOVERN sets the rules, MAP tells you what exactly you're managing and where attention is needed. Think of it as creating a blueprint for AI risk.

**Understanding AI System Purpose and Context:**

Before managing AI risk, you must know what the system is designed to do:

- **Purpose:** What problem does it solve?
    - Example: LLM for customer support, AI for credit scoring, vision model for quality inspection
- **Context:** Where and how will it be used?
    - Internal vs external
    - Human-in-the-loop vs fully automated
    - Frequency and scale of use

Why it matters:

- Understanding context helps identify potential harm if the system fails or is misused.

**Defining Intended Use & Misuse:** Every AI system has a proper use case and potential misuses:

- **Intended use:**
    - Model performs tasks it was designed for
    - Example: LLM generating support emails

- **Potential misuse:**
    - Deliberate or accidental use outside design
    - Example: Generating phishing emails, producing biased outputs, model extraction attempts

By clearly defining intended use and misuse:

- You can design safeguards
- Identify risk scenarios
- Reduce regulatory and ethical exposure

**Stakeholder Analysis:**

MAP requires identifying all parties impacted by or responsible for AI systems:

- Internal stakeholders:
    - Developers, ML engineers, product owners, security teams, compliance/legal
- Executives: CISO, CTO, business owners
- End-users: Employees, customers, clients
- External stakeholders: Vendors, auditors, regulators

Analysis includes:

- Who relies on the AI system?
- Who can influence the system's behaviour?
- Who is accountable for decisions and outcomes?

**Mapping Risks and Impacts**

Here you map where things could go wrong and what the consequences would be:

- **Types of AI risks:**
    - Security: prompt injection, model theft, poisoning
    - Privacy: PII exposure, data leaks
    - Ethical: bias, discrimination, unsafe content
    - Operational: performance drops, drift, misclassifications
    - Regulatory: non-compliance with AI Act, GDPR, DPDP
- **Impact mapping:**
    - Identify who or what is affected
    - Assess severity: minor, moderate, critical
    - Determine probability: low, medium, high

Purpose:

- Helps prioritize mitigation efforts
- Supports risk-informed decision-making
- Provides input for the RISK REGISTER

**Identifying Trustworthiness Gaps:** Trustworthiness gaps = areas where the AI system may fail to meet reliability, safety, ethics, or compliance expectations.

Check for gaps in:

- Fairness: biased outputs or discrimination
- Robustness: vulnerability to attacks or adversarial inputs
- Explainability: lack of documentation or interpretability
- Privacy: data handling, training, and inference
- Accountability: unclear responsibilities or decision ownership
- Compliance: gaps vs regulations like EU AI Act

Purpose:

- Gaps indicate where protections, monitoring, or redesigns are needed
- Allows proactive risk reduction

Basically,

**MAP = Understand your AI system + stakeholders + risks.**

You learn to:

a. Define purpose and context
b. Identify intended use and possible misuse
c. Analyze stakeholders
d. Map risks and potential impacts
e. Spot trustworthiness gaps
f. MAP = “Know your AI system, who it affects, what can go wrong, and where it is weak.”

--The End--