

## Day 5



# Generative AI Security

### RACI Models for AI Responsibilities:

A RACI model is a tool that clarifies who does what in a project. It ensures accountability, avoids confusion, and improves cross-functional collaboration.

#### What is a RACI Model?

RACI stands for:

- R – Responsible: The person/team who does the work or executes the task.
- A – Accountable: The person who owns the task and approves it. There is only one Accountable per task.
- C – Consulted: People who provide input, expertise, or advice. Two-way communication.
- I – Informed: People who need to know the outcome or progress. One-way communication.

#### Why RACI is Important in AI Governance:

- AI projects involve many teams: ML engineers, data teams, security, legal, compliance, business owners, executives.
- AI is high-risk → clear responsibility is critical.
- Avoids gaps and overlaps (e.g., who approves risk, who monitors for prompt injection).
- Helps with auditability → regulatory compliance requires clear ownership.
- Ensures smooth collaboration across AI lifecycle stages.

#### RACI Chart for AI Lifecycle Activities:

AI Lifecycle Activity	Responsible (R)	Accountable (A)	Consulted (C)	Informed (I)
Data Sourcing	Data Engineers	Data Owner / DPO	ML Engineers, Legal	Business Owner, CTO
Model Training	ML Engineers	AI Product Owner	Data Team, Security	CISO, CTO
Evaluation & Red Teaming	Red Team AI Specialists	AI Security Lead	ML Engineers, Product Owner	CISO, CTO
Deployment	ML Ops Engineers	AI Product Owner / CTO	Security, Compliance	Business Owner, End Users

AI Lifecycle Activity	Responsible (R)	Accountable (A)	Consulted (C)	Informed (I)
Monitoring	SOC Analysts, ML Ops	AI Security Lead	ML Engineers, Product Owner	CISO, CTO
Incident Handling	Security Team / ML Ops	CISO	AI Security Lead, Legal, Product Owner	CTO, Business Owner
Compliance Reporting	Compliance Legal Team	/ CISO / CTO	Data Team, Product Owner	Executives, Board

#### Examples of Specific Tasks:

Task	Responsible (R)	Accountable (A)	Consulted (C)	Informed (I)
Bias Testing	ML Engineers	AI Product Owner	Data Team, Red Team	CISO, CTO
Prompt Injection Defense	Security Team	AI Security Lead	ML Engineers, Product Owner	CISO, CTO
Model Card Reviews	ML Engineers / Product Owner	AI Governance Board	Compliance, Security	CTO, Executives
Vendor AI Selection	Product Owner / ML Engineers	CTO	Security, Legal, Compliance	CISO, Business Owner

#### Key Points for AI RACI:

1. **One accountable per task** → avoids confusion.
2. **Responsible can be multiple** → several teams can execute work.
3. **Consulted = domain experts** → two-way advice.
4. **Informed = stakeholders who need visibility** → no active participation.
5. **Use RACI across the AI lifecycle** → ensures clarity from data sourcing to monitoring.
6. **Supports regulatory audit** → shows who owns risks, controls, and approvals.

Basically, RACI ensures clear accountability, smooth collaboration, and regulatory readiness.

## **How the AI Governance Board Manages AI-Related Risk:**

An AI Governance Board ensures AI systems are developed, deployed, and monitored safely, focusing on risk management. AI systems have unique risks compared to traditional IT systems, so boards follow structured processes to identify, assess, mitigate, and monitor risks.

### **Types of AI-Specific Risks:**

AI introduces several categories of risks:

1. Security Risks: Approve security controls, threat modeling, and monitoring.
2. Privacy Risks: Ensure compliance with GDPR, DPDP Act, HIPAA; approve data minimization and anonymization measures.
3. Ethical Risks: Review fairness, explainability, and ethical use case guidelines; approve high-risk use cases.
4. Model Drift Risks: Define monitoring, retraining policies, and drift detection mechanisms.
5. Regulatory Risks: Ensure regulatory mapping, approvals, and audit-ready documentation.

### **Building an AI Risk Taxonomy:**

Helps the board prioritize risks and assign responsibility. A risk taxonomy classifies risks for clarity:

Category	Examples	Risk Owner
Security	Prompt injection, API abuse	AI Security Lead
Privacy	PII exposure, consent violations	DPO
Ethical	Bias, discrimination, harmful outputs	Product Owner / Governance Board
Model Drift	Accuracy drop, performance degradation	ML Engineer / MLOps
Regulatory	Non-compliance, audit failure	Compliance / Legal

### **Risk Assessment Methods:**

- 1. Likelihood × Impact Scoring**
  - Assign likelihood of occurrence (1–5)
  - Assign impact if it occurs (1–5)
  - Multiply → risk score
  - Score guides whether risk is low, medium, high, or critical
- 2. Qualitative vs Quantitative Assessment**
  - **Qualitative:** Labels like Low / Medium / High; based on expert judgment
  - **Quantitative:** Numerical scoring, probabilities, and metrics (e.g., % of biased predictions, potential data leakage cost)
  - **Hybrid Approach:** Combine both for precision and governance clarity.

**Reviewing & Approving High-Risk AI Use Cases:** High-risk AI cannot be deployed without board approval.

### **Linking Risks to Controls:**

The board ensures each risk has specific mitigation measures:

Risk	Controls / Mitigations
Data risk	Data validation, anonymization, lineage tracking
Model risk	Bias testing, robustness testing, explainability
Security risk	Access control, encryption, input/output validation
Model drift	Monitoring, retraining, alerting
Privacy risk	Differential privacy, data minimization, consent tracking
Regulatory risk	Documentation, approvals, audits

### **Maintaining an AI Risk Register:**

The risk register is a central log of:

- Identified risks
- Risk owners
- Likelihood & impact scores
- Mitigation measures
- Status and review dates

### **Escalation & Exception Handling Workflows:**

- If a risk cannot be mitigated completely, the board defines:
  - Who can accept the residual risk
  - Approval hierarchy for exceptions
  - Escalation path for critical/high-risk issues

### **Role of Board in Incident-Response Decisions:**

Board Responsibilities:

- Approve incident response plan
- Decide on shutdown, rollback, or retraining
- Escalate to executives if severe
- Ensure post-incident review and lessons learned
- Update policies and risk register to prevent recurrence

Basically,

- AI Governance Board manages security, privacy, ethical, model drift, and regulatory risks.
- Uses a risk taxonomy to categorize and assign ownership.
- Performs risk assessments using likelihood × impact and qualitative/quantitative methods.
- Reviews and approves high-risk AI use cases.
- Links risks to controls like data validation, model hardening, and monitoring.
- Maintains a central AI risk register.
- Defines escalation and exception handling workflows.
- Plays a key role in incident-response decisions and policy updates.

--The End--