

## Day 6



# Generative AI Security

### **Meeting Structure & Decision-Making Workflows in AI Governance:**

AI Governance Boards manage high-risk AI projects. To do this effectively, they follow structured meetings, decision-making models, and documented workflows.

#### **Setting Meeting Cadence:**

The board should meet based on AI project needs, risk levels, and organizational policy. Typical frequencies: weekly, monthly, Pre AI milestone.

#### **Typical Agenda Items:**

Every meeting usually covers:

1. AI Use case Proposals
2. Security test results
3. Compliance gaps
4. Risk Assessment Updates
5. Incidents & Near Misses
6. Action Items

#### **Decision-Making Models:**

Use a hybrid model: consensus for high-risk ethical/regulatory issues, voting or owner authority for operational approvals. Different boards use different models depending on risk, urgency, and policy:

1. Consensus based
2. Majority Voting
3. Risk owner Authority

**Standard Workflows for Approvals:** A structured approval workflow ensures nothing is skipped.

#### **Documenting Decisions in Traceable Logs:**

- All board decisions should be logged for traceability:
  - Meeting date, attendees, decisions made
  - Approvals, rejections, and conditions
  - Assigned owners for actions
- Logs support regulatory audits and internal accountability

**Stakeholder Sign-Off Processes:** Stakeholders provide formal sign-offs to ensure accountability and compliance.

Basically,

- AI Governance Board meetings are structured by cadence: weekly, monthly, or per milestone.
- Agenda items include AI proposals, security tests, compliance gaps, risk updates, and incidents.
- Decision-making models: consensus-based, majority voting, or risk-owner authority.
- Approval workflows: data → model training → deployment → post-deployment monitoring.
- All decisions are documented in traceable logs.
- Stakeholders provide formal sign-offs to ensure accountability and compliance.

## **Documenting Governance Decisions in AI Security:**

Documenting decisions is a core part of AI governance. It ensures that every decision about AI use, deployment, or risk mitigation is traceable, auditable, and accountable. Good documentation supports compliance, monitoring, audits, and post-incident investigations.

### **Templates for Documenting Governance Decisions:**

Using standardized templates ensures consistency and completeness. A typical governance decision template includes:

1. Decision ID / Reference Number
2. Date of Decision
3. Decision Maker(s) (board members, risk owner, compliance officer)
4. AI Project / Model Name
5. Type of Decision (approve, reject, escalate, exception)
6. Rationale / Justification
7. Risk Level (low, medium, high, critical)
8. Responsible Parties (who executes the decision)
9. Controls / Mitigations (linked to the decision)
10. Expiry / Review Date
11. Sign-offs (stakeholders who approved the decision)
12. Reference to Frameworks (NIST CSF, ISO 42001, EU AI Act)
13. Version Number (for updates)

**Version Control for Governance Artifacts:** Use version control to track updates and maintain historical records.

### **Decision Logs:**

Decision logs are central repositories for all governance decisions. Maintains traceable evidence that can be audited at any time. They should include:

1. **What Was Approved / Rejected**
  - a. e.g., AI use case, model deployment, risk acceptance
2. **Risk Level**
  - a. Low, Medium, High, Critical
  - b. Helps auditors and stakeholders quickly understand the decision's impact
3. **Rationale / Justification**
  - a. Why the decision was made

- b. Link to risk assessment, ethical considerations, or regulatory review
- 4. **Responsible Parties**
  - a. Who executes the decision (ML Engineers, Security Lead, Product Owner)
- 5. **Expiry / Review Date**
  - a. When the decision should be reviewed or renewed

#### **Compliance Evidence Generation:**

Documented decisions serve as proof for regulators and auditors:

- Demonstrates adherence to NIST CSF, ISO 42001, EU AI Act
- Provides evidence that risks were assessed, mitigations applied, and approvals given
- Supports ethical and safe AI deployment
- Can be used to respond to compliance inquiries or external audits

**Mapping Decisions to Frameworks:** Makes audits simpler and proves alignment with best practices.

- Governance boards often map decisions to recognized standards:
  - **NIST CSF:** Identify → Protect → Detect → Respond → Recover
  - **ISO 42001:** Risk management, governance, lifecycle oversight
- Each decision can be tagged to:
  - Which control/requirement it satisfies
  - Which compliance standard it aligns with

**Maintaining Audit Trails:** Maintain audit trails to support regulators and internal review.

#### **How Documentation Supports Post-Incident Investigations?**

Documentation is critical for post-incident investigations and improving governance processes.

Basically,

- Document every AI governance decision using templates: decision, risk, rationale, responsible parties, review date.
- Use version control to track updates and maintain historical records.
- Maintain decision logs as a central repository for approvals, risk levels, and mitigation links.
- Map decisions to frameworks (NIST, ISO 42001) for compliance.
- Maintain audit trails to support regulators and internal review.
- Documentation is critical for post-incident investigations and improving governance processes.

#### **Building Communication Loops Between AI & Security Teams:**

Communication between AI teams (ML Engineers, MLOps, Product Owners) and Security teams (CISO, Security Architects, SOC) is critical for safe AI development. Gaps in communication often lead to security incidents, privacy violations, and operational failures.

#### **Why Communication Gaps Cause AI Security Failures**

- Security requirements not understood by ML teams
- New threats not communicated
- Architecture changes not shared
- Isolated workflows (silos)

### **Building Continuous Communication Channels:**

Effective communication ensures both AI functionality and security are maintained. Key channels:

1. Security Requirements -> ML teams
2. Architecture Updates -> Security Teams
3. New Threats -> Engineering Teams

### **Use of Governance Tools:**

Use tools: ticketing, dashboards, model lifecycle tracking.

### **Establishing Mandatory Checkpoints:**

Checkpoints ensure governance oversight and reduce errors.

1. Design Review
2. Pre Deployment Approval
3. Post Deployment Monitoring Reviews

### **Sharing Insights Between AI Red Team and Blue Team:**

- Sharing findings:
  - Red team insights inform security hardening
  - Blue team can update ML engineers for defensive coding
- Creates continuous improvement loop, preventing repeated attacks

### **How Communication Helps Avoid Silos During AI Development**

- Reduces risk of isolated decisions
- Ensures security, compliance, and ethical considerations are integrated into AI lifecycle
- Promotes collaborative culture between AI engineers, product owners, and security
- Improves incident response speed because everyone knows roles and reporting channels

Basically,

- Communication gaps between AI and security teams cause failures and vulnerabilities.
- Continuous channels: security → ML, architecture → security, new threats → engineering.
- Use tools: ticketing, dashboards, model lifecycle tracking.
- Mandatory checkpoints: design review, pre-deployment approval, post-deployment monitoring.
- Red and Blue teams share insights for continuous improvement.
- Strong communication prevents silos and ensures AI is safe, compliant, and resilient.

--The End--