Day 13

Generative AI Security

AI Security Policies — Understanding AI Security Policies:

Definition of AI Security Policy:

An AI security policy is a formal written rulebook that defines:

- how AI systems should be used
- how they should be protected
- what is allowed and not allowed
- how risks must be managed
- who is responsible for AI security decisions

It tells employees and teams how to safely build, use, and manage AI systems. Think of it as a security manual specifically for AI, not for normal IT systems.

Difference Between General IT Security Policy vs. AI Security Policy:

General IT Security Policy	**AI Security Policy**
Covers:	Covers:
• servers • networks • databases • endpoints • user access • malware, phishing, ransomware	• AI models (LLMs, ML models) • training and inference data • prompts and embeddings • AI APIs and pipelines • model behavior and misuse • bias, hallucinations, and unsafe outputs
Focus:	Focus:
• traditional cyber risks	• AI-specific risks, not just system access

Purpose of AI Security Policies:

AI security policies exist to: reduce AI risks, ensure compliance, define acceptable use.

Importance of AI-Specific Policies in Generative AI Systems:

AI security policies set clear boundaries for generative AI usage. Generative AI systems are high-risk because:

- they generate new content
- they can leak sensitive data
- they can be manipulated by prompts
- they may hallucinate
- they may be misused by insiders

Without AI-specific policies:

- employees may paste sensitive data into public LLMs
- models may be used for unethical purposes
- regulators may impose penalties

Who Owns and Approves AI Security Policies:

Policy Ownership

Usually owned by:

- CISO / AI Security Lead
- AI Governance Board
- Risk & Compliance Teams

Policy Approval

Approved by:

- Executive leadership
- CIO / CTO
- Legal & Compliance
- Data Protection Officer (for privacy parts)

Ownership ≠ Approval. Ownership maintains the policy; leadership approves it.

Policy Lifecycle: Creation → Approval → Implementation → Review

Basically,

- AI security policy = rulebook for safe AI usage and protection
- It is different from IT security policy because AI has unique risks
- It reduces risks, ensures compliance, and defines acceptable use
- Generative AI must have AI-specific policies
- Owned by security/governance teams and approved by leadership
- Follows a lifecycle: create → approve → implement → review

AI Security Policies — Common Sections of an AI Security Policy:

An AI security policy is usually divided into clear sections so that everyone understands what it covers, why it exists, and who must follow it.

Introduction & Scope

What this section is?

This section explains what the policy applies to and who must follow it.

What it usually includes

- Types of AI systems covered
  - LLMs (ChatGPT-like systems)
  - ML models
  - Generative AI (text, image, code)
- AI components covered
  - models
  - datasets
  - prompts
  - APIs
  - pipelines
- Teams covered
  - developers
  - ML engineers
  - data scientists
  - security teams
  - vendors and contractors

Purpose and Objectives

What this section is

Explains why the policy exists and what it wants to achieve.

Common objectives

- Secure AI development
- Safe AI deployment
- Prevent misuse and attacks
- Protect sensitive data
- Ensure compliance with laws and standards
- Promote responsible AI use

Example objectives

- Reduce prompt injection and data leakage
- Ensure models are tested before deployment
- Align with EU AI Act, ISO 42001, NIST AI RMF

Roles & Responsibilities: Defines who does what in AI security.

Typical roles:

Developers / ML Engineers

- build and train models
- follow secure coding and ML practices
- perform basic testing

Data Teams

- ensure data quality
- manage data lineage
- protect PII and sensitive data

Security Teams

- perform threat modeling

Compliance & Legal Teams

- check regulatory requirements
- review AI use cases
- support audits

Leadership / Governance Board

- approve AI use cases
- approve policies
- accept or reject risks

- conduct security testing
- monitor AI systems
- respond to incidents

Risk Management & Compliance References: Links the policy to recognized frameworks and standards.

Common references

- NIST CSF
- NIST AI RMF
- ISO 42001
- ISO 27001
- GDPR / DPDP / EU AI Act
- Internal security standards

What it explains

- how AI risks are assessed
- how risks are prioritized
- how controls are selected
- how compliance is tracked

Definitions and Terminology: Explains important AI and security terms used in the policy.

Examples of definitions

- AI system
- Generative AI
- Prompt injection
- Model drift
- Training data
- Inference
- High-risk AI
- Personally Identifiable Information (PII)

Basically,

An AI security policy usually has these main sections:

1. **Introduction & Scope** → what and who the policy applies to
2. **Purpose & Objectives** → why the policy exists
3. **Roles & Responsibilities** → who does what
4. **Risk & Compliance References** → which standards and laws it follows
5. **Definitions** → clear meanings of AI and security terms

These sections make the policy clear, enforceable, and audit-ready.

--The End--