AI SECURITY ENGINEERING FIELD GUIDE · 2026

# Chapter 01 · AI Security Foundations

Standalone learning module for LMS delivery and required reading.

| FORMAT | USE | SCOPE | AUDIENCE |
| --- | --- | --- | --- |
| Standalone PDF | LMS reading module | Single chapter | Learners |

CHAPTER 01

# AI Security Foundations

**INSPECT**

AI inventory, owners, data flows, risk tiers, evidence paths, and release gates.

**PRODUCE**

System inventory, trust-boundary map, owner register, and evidence request list.

> You cannot secure an AI system you cannot name, map, or explain.
>
> FIELD GUIDE

**FIELD GUIDE: FOUNDATION REVIEW**

Start by identifying the AI surface, the trust boundaries, the owners, the evidence path, and the decisions the system can influence.

This domain covers the baseline work behind every AI security review: inventory, system purpose, model and provider use, data movement, trust boundaries, ownership, risk tiering, and evidence readiness. It matters when a team cannot explain which AI features exist, which users they affect, what data they touch, or who owns the control decisions.

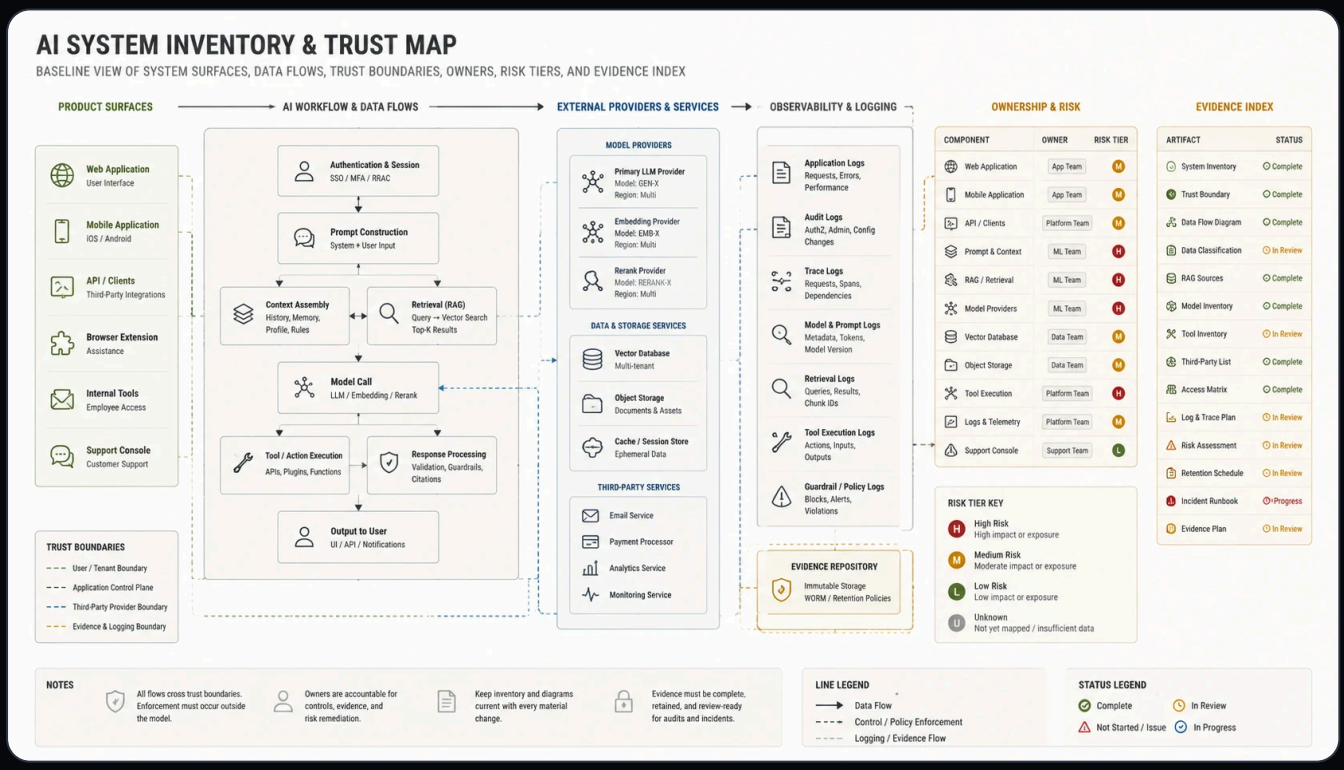


FIGURE 1: FIGURE 2: AI SYSTEM INVENTORY AND TRUST MAP SHOWING MODEL AND PROVIDER CONNECTIONS, DATA FLOWS, OWNERSHIP LAYERS, AND RISK-TIER CRITERIA

### WHAT BREAKS HERE?

- Shadow AI features bypass review, logging, and procurement.
- Teams confuse a model name with a system inventory.
- Ownership is split across product, platform, security, legal, and vendors.
- Risk tiering is based on excitement or revenue, not authority, data, exposure, and reversibility.
- Governance asks for evidence that engineering never designed the system to produce.

### WHAT TO INSPECT

- Product surfaces that call models, agents, embeddings, classifiers, copilots, or AI APIs.
- Data flows into prompts, retrieval systems, tools, logs, eval sets, model providers, and analytics.
- Existing inventories, architecture diagrams, API gateway records, vendor lists, and feature flags.
- Owner records for product, engineering, security, privacy, legal, and support.
- Existing control evidence: logs, approvals, test reports, vendor terms, and release notes.

### WHAT TO ASK

- Which systems use AI directly or indirectly?
- What data can each system read, transform, store, retrieve, or send?
- What user or business decision can the AI output influence?
- Which controls run outside the model?
- Who can approve a launch, exception, rollback, or emergency shutdown?
- What evidence can the team produce today without manual reconstruction?

### WHAT TO TEST

- Compare declared AI inventory against code search, vendor spend, browser surfaces, SDKs, and telemetry.
- Trace one AI request from user action to model call, tool call, retrieved context, output, and log.
- Verify that every high-impact system has an owner, risk tier, and evidence location.
- Check that fallback paths preserve logging, authorization, and approval requirements.
- Sample recent releases for AI security review records and unresolved exceptions.

### CONTROLS AND GUARDRAILS

- AI system inventory with owner, data class, user class, model/provider, authority, and evidence fields.
- Intake gate for new AI features, providers, retrieval sources, and tools.
- Risk-tier rules based on data sensitivity, delegated action, user impact, exposure, and reversibility.
- Architecture review trigger for model changes, retrieval changes, tool scope changes, and provider changes.
- Evidence requirements linked to release gates and exception records.

### EVIDENCE TO COLLECT

- AI system inventory export.
- Data-flow and trust-boundary map.
- Risk-tier rationale.
- Owner and approver register.
- Release review records.
- Exception register and expiry dates.
- Control evidence index.

### OUTPUT ARTIFACTS

- AI system inventory.
- Trust-boundary map.
- AI risk tiering worksheet.
- Evidence request list.
- Review backlog.

### RELATED SERVICES AND WORKBENCH TOOLS

| TYPE | RELATED PATHS |
| --- | --- |
| Workbench | Threat Canvas, Surface Scanner, AI Control Crosswalk |
| Services | AI Product Security Assessment, Enterprise AI Security Readiness, AI Security Operating Model |
| Handbook | AI System Inventory, Architecture and Trust Boundaries, Governance Evidence and Customer Trust |

**ARTIFACT: FOUNDATION EVIDENCE PACK**

Produce an inventory, a trust map, owner records, risk tiers, and an evidence index before debating advanced controls.