AI SECURITY ENGINEERING FIELD GUIDE · 2026

Chapter 09 · Privacy and Data Protection in AI Systems

Standalone learning module for LMS delivery and required reading.

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

CHAPTER 09

Privacy and Data Protection in AI Systems

> AI privacy review starts where source data becomes prompts, embeddings, logs, memory, and derived records.

Chapter 09 · Privacy and Data Protection in AI Systems AI Security Engineering Field Guide 2026

> AI privacy review starts where source data becomes prompts, embeddings, logs, memory, and derived records.
>
> FIELD GUIDE

> **FIELD GUIDE: DATA PROTECTION REVIEW**
>
> Trace sensitive data from source collection through prompt use, retrieval, embedding, model calls, outputs, memory, analytics, and deletion.

This domain covers privacy and data protection for AI systems: data minimization, purpose limitation, sensitive data in prompts, embeddings, memory, logs, training, fine-tuning, provider processing, retention, deletion, consent, and data-subject requests. It matters when AI systems touch personal data, customer data, regulated data, employee data, or confidential business records.

The applied review should follow data after it stops looking like the original record. Prompts, chunks, embeddings, memories, eval sets, analytics rows, support traces, and provider logs can all become derived records that privacy workflows miss.

> **FIELD USE**
>
> Choose one sensitive data element and trace it through prompt use, retrieval, embedding, provider call, memory, logs, eval data, deletion, and customer evidence.

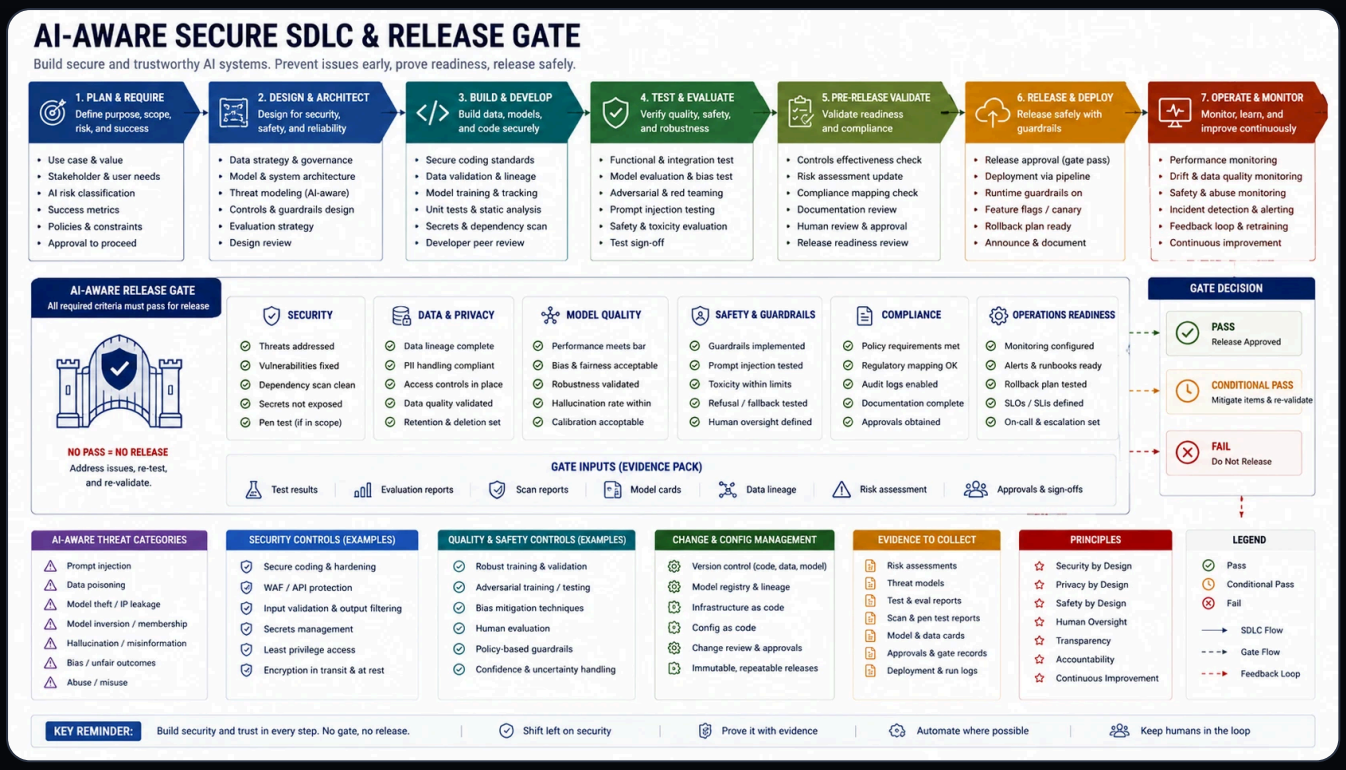

FIGURE 1: FIGURE 10: AI PRIVACY DATA FLOW MAP TRACING SENSITIVE DATA FROM COLLECTION THROUGH PROMPTS, EMBEDDINGS, MEMORY, ANALYTICS, AND DELETION

WHAT BREAKS HERE?

- Prompts and logs store more personal data than the product needs.
- Embeddings and derived records are missing from deletion workflows.
- Provider settings allow retention or training use that conflicts with promises.
- Memory stores sensitive facts without user control.
- Privacy reviews miss retrieval, analytics, eval sets, and support tooling.
- Sensitive data moves into debugging, support, or analytics systems with weaker access controls.

WHAT TO INSPECT

- Data categories entering prompts, retrieval, embeddings, logs, memory, evals, and providers.
- Data-flow diagrams and records of processing.
- Retention, deletion, and data-subject request workflows.
- Provider terms, data processing settings, sub-processors, and regional processing.
- Admin controls, user controls, redaction, masking, and access logs.
- Support access, export paths, analytics events, and evaluation datasets.

WHAT TO ASK

- What personal or sensitive data enters the AI system?
- What derived records does the system create?
- Can data be deleted from prompts, logs, embeddings, memory, evals, and analytics?
- What provider processing promises does the company make to customers?
- Which data is needed for the feature, and which is convenience capture?
- What evidence proves privacy controls ran?
- Who can access prompt, output, memory, and embedding records during support or debugging?

WHAT TO TEST

- Sensitive prompt and output logging.
- Deletion propagation into embeddings, memory, logs, and eval data.
- Provider retention and training-use configuration.
- Cross-border processing and sub-processor records.
- Memory controls and user-visible deletion.
- Unauthorized admin access to prompt and output records.
- Data minimization at prompt assembly, retrieval, and trace capture.

CONTROLS AND GUARDRAILS

- Data minimization before prompt, retrieval, and provider calls.
- Sensitive data classification for prompts, outputs, embeddings, memory, and logs.
- Retention and deletion workflows that include derived AI records.
- Provider processing controls for retention, training use, region, and support access.
- Redaction, masking, access controls, and audit logs for AI traces.
- Privacy review gate for new AI features and material changes.
- Derived-record inventory for prompts, chunks, embeddings, memories, evals, traces, and analytics.

EVIDENCE TO COLLECT

- AI data-flow review.
- Data category inventory.
- Provider processing notes.
- Retention and deletion test results.
- Memory control evidence.
- Access log review.
- Privacy remediation backlog.
- Derived-record deletion and minimization evidence.

OUTPUT ARTIFACTS

- Privacy and data-flow review.
- AI data inventory.
- Deletion propagation checklist.
- Provider processing boundary statement.
- Evidence map.

RELATED SERVICES AND WORKBENCH TOOLS

TYPE	RELATED PATHS
Workbench	Runtime Proxy, AI Control Crosswalk, Trust Scanner
Services	Enterprise AI Security Readiness, AI Product Security Assessment, AI Security Operating Model
Handbook	Data Exposure and Privacy, Model and Provider Risk, Governance Evidence and Customer Trust

ARTIFACT: AI PRIVACY EVIDENCE PACK

Produce data-flow notes, provider settings, retention checks, deletion proof, memory controls, and access evidence.