BUYER'S GUIDE · 2026

A COMPANION TO THE STATE OF AI DATA SECURITY

The AI Security Buyer's Guide

Make enterprise knowledge safe for AI. What to require from AI data security — and how to evaluate it — before you choose a vendor.

WHO IT'S FOR

- CISOS
- DATA LEADERS
- AI / ML LEADERS

Use it to set requirements, write your RFP, and score vendors against a consistent standard.

INSIDE THIS GUIDE

- **5-domain** requirement framework
- Must-have **criteria checklists**
- **RFP questions** to ask vendors
- A fillable **vendor scorecard**

HOW TO USE THIS GUIDE

Define what "good" looks like before you talk to vendors.

Most AI security evaluations start with a demo. This one starts with your requirements — so you compare vendors against your standard, not their pitch.

Our research found that security and compliance reviews are now the single biggest bottleneck moving AI from prototype to production: 83% of leaders said reviews had already delayed their AI projects. The fix isn't fewer controls — it's the right ones, defined early.

This guide breaks AI data security into **five requirement domains**. For each, you get the criteria to require, the questions to ask, and the watch-outs that separate real data-layer protection from perimeter controls with a new label.

Work through it in order. Mark each criterion as a must-have or should-have for your context, capture vendor answers to the RFP questions, then carry your notes into the scorecard on page 9.

> ### The one principle
>
> Enterprises need safe access to their knowledge. That means the data must stay protected *before* AI uses it — if protection doesn't travel with the data into training, retrieval, and model output, you're only guarding the door it already walked through.

HOW EACH DOMAIN IS LAID OUT

- ☑ **What to require** — the must-have and should-have criteria.
- ☑ **Ask your vendor** — RFP questions that expose the gaps.
- ☑ **Watch-outs** — answers that should give you pause.

PAIR IT WITH THE DATA

Every criterion here maps to a finding in The State of AI Data Security — read them together.

THE FRAMEWORK

Five domains to evaluate any AI data security solution.

Together they trace the path your data takes into AI — from discovery, through protection and governance, to operating at scale and proving compliance.

01	Discover & Classify	Know what's sensitive, and which AI systems can reach it, before anything ships.
02	Protect at the Data Level	Protection that travels with the data through training, RAG, and inference.
03	Govern & Control Access	One policy, role- and context-aware, enforced across every AI surface and agent.
04	Operate at AI Speed	Embedded in developer workflows, fast at high throughput, light to deploy.
05	Prove & Comply	Continuous evidence mapped to your regulations — and a quantum-safe roadmap.

DOMAIN 01

Discover & Classify

01

You can't protect — or prove you protect — what you can't see. Require complete visibility into sensitive data and which AI systems can reach it.

WHAT TO REQUIRE

MUST-HAVE

- Automated discovery of **structured and unstructured** sensitive data across cloud, on-prem, SaaS, and data lakes.
- Classification by sensitivity **and regulatory category** — PII, PHI, PCI, secrets, IP.
- Coverage of **AI-specific stores**: vector databases, RAG indexes, fine-tuning sets.
- A live map of **which pipelines and agents** can reach which regulated data.
- **Continuous re-scan** as new sources connect — not a point-in-time snapshot.

ASK YOUR VENDOR

Q1 How do you discover sensitive data in unstructured stores and vector databases?

Q2 Can you show which AI pipelines and agents currently have access to regulated data?

Q3 How frequently is classification refreshed as sources change?

WATCH-OUTS

Structured-data-only coverage · point-in-time scans · no visibility into RAG or vector stores.

WHY IT MATTERS · 17% NAMED SECURING SENSITIVE DATA FOR TRAINING A TOP BOTTLENECK — IT STARTS WITH KNOWING WHERE THAT DATA IS.

DOMAIN 02

Protect at the Data Level

02

This is the domain most "AI security" tools skip. Require protection that is bound to the data itself — so it stays protected everywhere the model takes it.

WHAT TO REQUIRE

MUST-HAVE

- Protection that **travels with the data** — tokenization or format-preserving encryption, not just at-rest or perimeter controls.
- **Field-level granularity** that preserves referential integrity, so protected data stays usable for analytics and AI.
- **Vaultless tokenization** that holds up at high throughput — no central vault bottleneck.
- Protection that **persists through training, RAG retrieval, and model output**.
- Re-identification only by **policy**, fully logged and reversible under control.

ASK YOUR VENDOR

Q1 Does protection persist when data enters a prompt, a vector store, or a model's output?

Q2 Is tokenization vaultless? What is the throughput and latency at our scale?

Q3 Can protected fields preserve format and referential integrity for downstream AI?

WATCH-OUTS

"We rely on access control" · protection stripped before model use · a vault that caps throughput.

WHY IT MATTERS · ONLY 9% USE VAULTLESS TOKENIZATION AT HIGH THROUGHPUT — THE GAP BEHIND THE #1 AGENTIC RISK: DATA LEAKAGE.

DOMAIN 03

Govern & Control Access

03

As agents gain autonomy, governance has to follow the action — not stop at the login. Require one policy enforced across every AI surface.

WHAT TO REQUIRE

MUST-HAVE

- **Policy-based access by role and context**, applied consistently to apps, pipelines, and agents.
- **Dynamic masking and least privilege** resolved at query and retrieval time.
- **Centralized policy, distributed enforcement** — author once, enforce everywhere.
- Controls scoped to **agent autonomy** — what an agent may retrieve, return, and act on.
- A clear, assignable **owner** for AI data governance baked into the workflow.

ASK YOUR VENDOR

Q1 Is it one policy across every AI surface, or a separate policy per tool?

Q2 How do you govern what an autonomous agent can retrieve, return, and execute?

Q3 Can access decisions factor in both role and runtime context?

WATCH-OUTS

Per-tool policies that drift · no agent-action controls · all-or-nothing static access.

WHY IT MATTERS · 69% OF AGENTS ALREADY ACT WITH MODERATE-TO-HIGH AUTONOMY — GOVERNANCE HAS TO REACH THE ACTION.

DOMAIN 04

Operate at AI Speed

04

Security that slows the pipeline gets bypassed. Require controls that live where developers already work and keep up at production throughput.

WHAT TO REQUIRE

SHOULD-HAVE

- **Embeds in developer workflows** — SDK, container, or API — not a separate gate to clear.
- **High throughput, low latency** at scale, with published benchmarks you can validate.
- **Centrally managed, workflow-embedded** — the model leaders prefer over appliances.
- Broad coverage of your stack — **Snowflake, Databricks, cloud, Kubernetes**.
- **Minimal re-architecture** — protection added without rebuilding pipelines.

ASK YOUR VENDOR

Q1 What is the latency overhead per operation at our target throughput?

Q2 How is it deployed into developer workflows — SDK, sidecar, container?

Q3 Which data platforms and AI frameworks are natively integrated?

WATCH-OUTS

Appliance bottlenecks · heavy pipeline rework · manual approvals sitting in the data path.

WHY IT MATTERS · LEADERS FAVOR A CENTRALLY MANAGED PLATFORM (29%) AND SDK/CONTAINER DELIVERY (23%) OVER APPLIANCES.

DOMAIN 05

Prove & Comply

05

The outcome leaders want most is less manual audit work. Require evidence that generates itself — and a strategy that survives the next crypto shift.

WHAT TO REQUIRE

MUST-HAVE

- **Continuous monitoring** of how AI touches data, with a full audit trail of access and re-identification.
- Evidence mapped to **GDPR, HIPAA, PCI-DSS, CCPA, DORA, and the EU AI Act**.
- Reporting that **cuts manual audit effort** — not just more logs to sift.
- **Data lineage** tying every protected field to its policy and owner.
- A **quantum-safe / crypto-agile** roadmap to future-proof today's protection.

ASK YOUR VENDOR

Q1 Which frameworks can you generate audit evidence for, out of the box?

Q2 How specifically do you reduce our manual audit and compliance effort?

Q3 What is your crypto-agility and post-quantum cryptography roadmap?

WATCH-OUTS

Logging without lineage · evidence gathered by hand · no answer on post-quantum.

WHY IT MATTERS · 47% WANT LOWER AUDIT COST AS THEIR #1 OUTCOME; 85% PRIORITIZE QUANTUM-SAFE ENCRYPTION.

VENDOR SCORECARD

Score each vendor against your standard.

Rate every domain 1–5 based on the criteria and the answers you collected. Weight to your context, then total. Run the same sheet for each vendor.

REQUIREMENT DOMAIN	SUGGESTED WEIGHT	SCORE (1–5)
01 · Discover & Classify Visibility into sensitive data & AI access	20%	○ ○ ○ ○ ○
02 · Protect at the Data Level Protection bound to the data itself	30%	○ ○ ○ ○ ○
03 · Govern & Control Access One policy across every AI surface & agent	20%	○ ○ ○ ○ ○
04 · Operate at AI Speed Embedded, fast, light to deploy	15%	○ ○ ○ ○ ○
05 · Prove & Comply Self-generating evidence & quantum-safe path	15%	○ ○ ○ ○ ○

SCORING GUIDE

5 meets every must-have, evidenced · **3** partial / roadmap · **1** perimeter controls relabeled.

THE DECIDING QUESTION

If Domain 02 scores low, the rest can't compensate — perimeter strength never offsets unprotected data.

WHAT THIS MEANS FOR YOUR ROLE

Same data problem, three mandates.

Security, data, and AI leaders are buying toward the same outcome from different angles. Here's the through-line — and where the scorecard earns its keep for each.

ROLE	WHAT THEY NEED	WHY IT MATTERS	WHY PROTEGRITY
CISO SECURITY & RISK	Secure AI agents and models against data exposure.	Security teams are accountable for zero data exposure as agents gain autonomy to act.	Protegrity stops leaks **before** AI acts — protection bound to the data, not the perimeter.
CDAO DATA & ANALYTICS	Make data available for AI without losing control of it.	Teams need trusted data utility — data that stays useful once it's protected.	Protegrity keeps data **useful and protected** — format-preserving, referentially intact.
AI Leader AI & PLATFORM	Govern agentic data consumption at scale.	Agents need safe access to enterprise knowledge — with limits on what they can use.	Protegrity **controls what agents can use** — one policy across every AI surface.

The common thread: enterprises need safe access to knowledge. Protegrity protects the data before AI uses it — so every role can say yes.

HOW PROTEGRITY MAPS TO THE FRAMEWORK

Built for the domain most tools skip.

Use this guide with any vendor. Here is where Protegrity fits — protection bound to the data, governed by policy, proven by evidence.

01 · DISCOVER Find & classify sensitive data before AI.	Automated discovery and classification across stores, pipelines, and AI sources.
02 · PROTECT Tokenize at the source.	Vaultless tokenization and format-preserving encryption that travel into training, RAG, and inference.
03 · GOVERN One policy, every surface.	Policy-based, role- and context-aware access control enforced consistently across AI workflows.
04 · OPERATE & 05 · PROVE Embedded & auditable.	Workflow-embedded delivery at high throughput, with continuous monitoring and compliance evidence.

Bring this scorecard to a working session with our team.

BOOK A DEMO →