AZURE DRAGON CYBER

Training | Mentoring | Cybersecurity Consulting



WHITE PAPER | OCTOBER 2026

Building Trust Into the Model

Cybersecurity, Data Privacy, and Compliance Challenges for AI-Driven Startups

Mark Adams

Founder, Azure Dragon Cyber | Cybersecurity, GRC, and AI Security Governance

azuredragoncyber.com

Contents

Executive Summary
1. Why AI Startups Face a Different Risk Profile
2. The Regulatory Landscape (October 2026)
3. Fourteen Core Data Privacy and Protection Challenges
4. The AI Threat Landscape at a Glance
5. Reference Control Architecture for AI Data Protection
6. Framework Alignment
7. Stage-Based Roadmap
8. Key Compliance Dates
9. Questions Every Founder and Board Should Be Able to Answer
10. Conclusion
About the Author
Sources

Citations appear as numbered footnotes with full source URLs; a consolidated source list follows the About the Author section.

Executive Summary

Startups that put artificial intelligence at the center of their product carry a risk profile that differs in kind, not just degree, from traditional SaaS companies. Customer data no longer just sits in a database behind an application. It is copied into training sets, chunked into vector stores, sent to third-party model providers, mixed into prompts, reflected in outputs, and, increasingly, acted upon by autonomous agents with access to real systems. Each of those flows creates a privacy obligation, a security exposure, and an audit question that a founder must be able to answer.

The external pressure is rising at the same time. As of October 2026:

- Twenty U.S. states have comprehensive privacy laws in force, and four more (Alabama, Louisiana, Oklahoma, and Vermont) have enacted laws that take effect later (Byte Back[1]).
- California's updated CCPA regulations took effect January 1, 2026, requiring privacy risk assessments before new high-risk processing, with automated decisionmaking technology (ADMT) obligations starting January 1, 2027 and cybersecurity audit certifications phasing in from April 1, 2028 (California Privacy Protection Agency[2]).
- The EU AI Act's transparency rules began applying in August 2026, while high-risk obligations were pushed to December 2, 2027 (stand-alone systems) and August 2, 2028 (systems embedded in products) under the Digital Omnibus on AI (European Commission[3]).
- IBM's most recent breach study found that 13% of organizations reported breaches involving AI models or applications, and 97% of those lacked proper AI access controls (IBM Newsroom[4]).

The central argument of this paper is simple: for an AI startup, privacy and security are product features and sales enablers, not overhead. Enterprise buyers, investors, and regulators now ask the same questions: What data trains or grounds your model? Who can see it? Can you delete it? Can you prove it? Startups that build answers into their architecture early move through security reviews faster, raise capital with fewer diligence surprises, and avoid the most expensive remedy in AI privacy enforcement: being ordered to delete a model built on improperly obtained data.

This paper outlines the regulatory landscape, the fourteen most consequential data privacy and protection challenges for AI startups, a reference control architecture, framework alignment, and a stage-based roadmap a founding team can execute without a large security department.

[1]Byte Back. https://www.bytebacklaw.com/2026/06/u-s-state-privacy-law-landscape-expands-to-24-states-what-the-latest-legislative-wave-means-for-businesses/
[2]California Privacy Protection Agency. https://cppa.ca.gov/announcements/2025/20250923.html
[3]European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[4]IBM Newsroom. https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls

© 2026 Azure Dragon Cyber | azuredragoncyber.com | Informational only; not legal advice.

1. Why AI Startups Face a Different Risk Profile

Traditional SaaS security focuses on protecting data at rest and in transit and controlling who can access the application. AI-centric products add five structural differences.

Difference	What changes	Why it matters for privacy
Data becomes model behavior	Personal data can be memorized in model weights or embeddings	Deletion, access, and correction rights become technically harder to honor
Natural language is the interface	Any text input, including retrieved documents, can carry instructions	Prompt injection can turn the model into an exfiltration channel
The supply chain is the product	Foundation models, APIs, vector databases, and orchestration frameworks are usually third-party	Customer data flows to sub-processors the customer may never have approved
Outputs are probabilistic	The same request can produce different answers	Testing, explainability, and "accuracy" claims are harder to substantiate
Agents take action	AI can call tools, query systems, and change records	Authorization errors become operational and financial incidents, not just disclosure events

Startups also operate under constraints that magnify these risks: small teams, speed-to-market pressure, heavy reliance on cloud and AI platform defaults, founders acting as de facto security and privacy owners, and investor-driven pivots that change data use faster than policies and contracts are updated.

2. The Regulatory Landscape (October 2026)

United States: Federal

There is still no comprehensive federal privacy or AI statute. The Federal Trade Commission remains the primary federal enforcer through Section 5 of the FTC Act, which prohibits unfair and deceptive practices (FTC[5]). Current FTC priorities relevant to AI startups include:

- **AI capability and data-practice claims ("AI washing").** In August 2026 the FTC finalized orders against Cox Media Group and two other firms over claims about an "Active Listening" AI-powered marketing service (FTC[6]). Practitioner analysis indicates the current FTC has shifted emphasis toward unsubstantiated AI claims rather than AI bias cases (Privacy + Security Academy[7]).
- **Model and data deletion remedies.** The FTC has prohibited companies from training AI models on ill-gotten data and required strong data retention and minimization (FTC[8]). This "algorithmic disgorgement" remedy, which requires deleting models derived from improperly collected data, was applied in the Weight Watchers/Kurbo and Amazon Alexa matters (Recording Law[9]).
- **Children's data.** The amended COPPA Rule's compliance deadline of April 22, 2026 has passed, bringing expanded parental-consent, data-retention, and written information security program obligations (Privacy + Security Academy[10]).

Federal preemption is a signal, not a safe harbor. Executive Order 14365 (December 11, 2025) directed the Attorney General to create an AI Litigation Task Force to challenge state AI laws (White House[11]), and DOJ established it on January 9, 2026 (U.S. Department of Justice[12]). However, because Congress has not passed a preemptive federal AI law, existing state AI laws are unlikely to be displaced by executive action alone, and organizations should continue to comply (Reuters[13]).

United States: States

Comprehensive privacy laws. Twenty states have comprehensive privacy laws in effect and 24 have enacted them (Byte Back[14]). Most share a common core: notice, access, deletion, correction,

[5]FTC. https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security/privacy-security-enforcement
[6]FTC. https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security/privacy-security-enforcement
[7]Privacy + Security Academy. https://www.privacysecurityacademy.com/wp-content/uploads/2026/04/FTC-Privacy-Enforcement-in-2026-and-Beyond.pdf
[8]FTC. https://www.ftc.gov/system/files/ftc_gov/pdf/ai-accomplishments-1.17.25.pdf
[9]Recording Law. https://www.recordinglaw.com/us-laws/data-privacy-laws/ai-and-data-privacy/
[10]Privacy + Security Academy. https://www.privacysecurityacademy.com/wp-content/uploads/2026/04/FTC-Privacy-Enforcement-in-2026-and-Beyond.pdf
[11]White House. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/
[12]U.S. Department of Justice. https://www.justice.gov/ag/media/1422986/dl?inline
[13]Reuters. https://www.reuters.com/legal/legalindustry/who-governs-ai-federal-governments-challenge-state-regulation-what-organizations--pracin-2026-08-12/
[14]Byte Back. https://www.bytebacklaw.com/2026/06/u-s-state-privacy-law-landscape-expands-to-24-states-what-the-latest-legislative-wave-means-for-businesses/

portability, opt-outs for targeted advertising, sale, and profiling, plus heightened treatment of sensitive data. For an AI startup, the "profiling" and "sensitive data" provisions are the most likely to be triggered.

California (CCPA/CPRA regulations). California is the most demanding state for AI companies:

- **Risk assessments.** Required before initiating processing that presents significant risk, including processing sensitive personal information, using ADMT for significant decisions, or using personal information to train ADMT (Legal 500[15]). For processing that began before January 1, 2026, assessments must be completed by December 31, 2027 (Morrison Foerster[16]). Assessments must be updated at least every three years and within 45 days of a material change (Alston & Bird[17]).
- **ADMT.** By January 1, 2027, businesses using ADMT for significant decisions in areas such as employment, financial services, housing, education, or healthcare need pre-use notices, opt-out mechanisms, explainability processes, and human review options (FTI Consulting[18]).
- **Cybersecurity audits.** Required for businesses that derive 50% or more of revenue from selling or sharing personal information, or that exceed the revenue threshold (about $26.6 million) and processed personal information of 250,000+ consumers or sensitive information of 50,000+ consumers (Latham & Watkins[19]). Certifications are due April 1, 2028, 2029, or 2030 depending on revenue (California Privacy Protection Agency[20]), and the audit must be performed by a qualified, objective, independent professional (Alston & Bird[21]).

Colorado. Colorado's original AI Act (SB 24-205) was repealed and replaced on May 14, 2026 by SB 26-189, which takes a narrower, notice-based approach to AI and ADMT used in consequential decisions and is effective January 1, 2027 (McDermott Will & Schulte[22]). The Colorado Attorney General published draft rules on August 11, 2026 for the new law and for Colorado's Chatbot Safety Act, signed May 29, 2026 (Davis Polk[23]). The lesson for startups is that AI state law is volatile; build to durable principles (notice, human review, documentation) rather than to any one statute's text.

European Union

GDPR applied to AI. The European Data Protection Board's Opinion 28/2024 is the most important guidance for AI companies handling EU personal data. It concludes that AI models trained on personal data cannot in all cases be considered anonymous; anonymity requires that both extraction of training data and retrieval of personal data through queries be insignificant (EDPB[24]). Legitimate

[15] Legal 500. https://www.legal500.com/intelligence/united-states/privacy/new-ccpa-regulations-receive-final-approval
[16] Morrison Foerster. https://www.mofo.com/resources/insights/251007-ccpa-regulations-on-cybersecurity-risk-assessments
[17] Alston & Bird. https://www.alston.com/en/insights/publications/2025/10/ccpa-cybersecurity-audits-admt-risk-assessments
[18] FTI Consulting. https://www.fticonsulting.com/insights/articles/ccpa-final-regulations-what-weve-learned-what-comes-next
[19] Latham & Watkins. https://www.lw.com/en/insights/navigating-new-obligations-under-the-ccpa-updated-regulations
[20] California Privacy Protection Agency. https://cppa.ca.gov/announcements/2025/20250923.html
[21] Alston & Bird. https://www.alston.com/en/insights/publications/2025/10/ccpa-cybersecurity-audits-admt-risk-assessments
[22] McDermott Will & Schulte. https://www.mcdermottlaw.com/insights/colorado-ai-law-in-flux-comprehensive-replacement-bill-signed-after-federal-court-blocks-predecessors-enforcement/
[23] Davis Polk. https://www.davispolk.com/insights/client-update/colorado-repeals-and-replaces-landmark-ai-act-publishes-draft-rules-public
[24] EDPB. https://www.edpb.europa.eu/system/files/documents/2024-12/edpb_opinion_202428_ai-models_en.pdf

interest can serve as a legal basis for AI development and deployment, but only after a three-step test of legitimacy, necessity, and balancing (EDPB[25]). Critically, a model developed with unlawfully processed personal data may make its later deployment unlawful unless the model has been properly anonymized (EDPB[26]).

Enforcement is real but contested. Italy's Garante fined OpenAI €15 million in December 2024 for training on personal data without an adequate legal basis, transparency failures, and inadequate age verification (Reuters[27]); a Rome court cancelled that fine in March 2026 (Reuters[28]). Even when fines are overturned, the investigation cost, service suspension risk, and remediation orders are significant burdens for a startup.

EU AI Act. The AI Act entered into force August 1, 2024, and the Digital Omnibus amendments entered into force July 27, 2026 (European Commission[29]). Key points for startups:

- Transparency obligations, such as disclosing that people are interacting with AI and disclosing deepfakes, apply from August 2, 2026 for systems placed on the market after that date (Mayer Brown[30]).
- Machine-readable marking of AI-generated content is required by December 2, 2026 for systems already on the market before August 2, 2026 (European Parliament[31]).
- High-risk use cases such as CV screening and credit scoring face risk management, data quality, logging, documentation, human oversight, and cybersecurity requirements starting December 2, 2027 (European Commission[32]).
- Fines reach €35 million or 7% of worldwide turnover for prohibited practices and €15 million or 3% for most other obligations (AI Act Service Desk[33]). For SMEs, including startups, the cap is the lower of the two amounts (IBM[34]), which helps but remains existential for an early-stage company.

[25] EDPB. https://www.edpb.europa.eu/news/edpb-opinion-on-ai-models-gdpr-principles-support-responsible-ai_en

[26] EDPB. https://www.edpb.europa.eu/news/edpb-opinion-on-ai-models-gdpr-principles-support-responsible-ai_en

[27] Reuters. https://www.reuters.com/technology/italy-fines-openai-15-million-euros-over-privacy-rules-breach-2024-12-20/

[28] Reuters. https://www.reuters.com/technology/italian-court-scraps-15-million-euro-privacy-watchdog-fine-chatgpt-maker-openai-2026-03-19/

[29] European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[30] Mayer Brown. https://www.mayerbrown.com/en/insights/publications/2026/07/eu-ai-act-news-digital-omnibus-on-ai-new-guidance-on-risk-classification-gpai-and-transparency-obligations

[31] European Parliament. https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-omnibus-on-ai

[32] European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[33] AI Act Service Desk. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-99

[34] IBM. https://www.ibm.com/think/topics/eu-ai-act

3. Fourteen Core Data Privacy and Protection Challenges

Challenge 1: Not knowing where personal data lives in the AI pipeline

Most startups can describe their production database. Few can describe every place personal data travels: raw ingestion buckets, labeling tools, fine-tuning datasets, evaluation sets, embeddings in a vector database, prompt and completion logs, observability tools, model-provider retention, and developer notebooks. IBM found that breaches involving shadow data (data in unmanaged sources) took longer to identify and contain and cost more on average (IBM[35]).

> **What good looks like:** A living AI data inventory and data-flow diagram covering each model, its training and grounding sources, data categories, legal basis, retention, location, and sub-processors. Treat embeddings, prompts, and logs as personal data unless proven otherwise.

Challenge 2: Lawful basis and purpose limitation for training

The most common startup mistake is assuming that data collected to deliver a service can also be used to train or improve models. Under GDPR this requires a defensible legal basis and a documented legitimate-interest assessment where relied upon (EDPB[36]). Under U.S. law, changing data use without clear notice risks a deception claim. Under California rules, using personal information to train ADMT for significant decisions can trigger a risk assessment (Legal 500[37]).

> **What good looks like:** Separate "service delivery" from "model improvement" purposes in notices, contracts, and technical controls. Default enterprise customer data to no-training unless the customer opts in in writing. Record the legal basis per dataset.

Challenge 3: Honoring data subject rights against models

Deletion, access, and correction are straightforward for database rows and hard for model weights. If a person's data was used in fine-tuning, deleting the source record does not remove its influence. The EDPB's position that models are not automatically anonymous (EDPB[38]) means regulators may expect a real answer.

> **What good looks like:** Prefer retrieval-augmented generation (RAG) over fine-tuning on personal data, since deleting a document from a retrieval index is tractable. Keep training-set lineage so affected models can be identified. Define a retraining or suppression procedure and document it before the first request arrives.

[35] IBM. https://www.ibm.com/think/insights/hidden-risk-shadow-data-ai-higher-costs
[36] EDPB. https://www.edpb.europa.eu/system/files/documents/2024-12/edpb_opinion_202428_ai-models_en.pdf
[37] Legal 500. https://www.legal500.com/intelligence/united-states/privacy/new-ccpa-regulations-receive-final-approval
[38] EDPB. https://www.edpb.europa.eu/news/edpb-opinion-on-ai-models-gdpr-principles-support-responsible-ai_en

Challenge 4: The model-deletion risk ("algorithmic disgorgement")

For an AI startup, the model is often the company's main asset. U.S. regulators have required deletion of models derived from improperly collected data (Recording Law[39]), and EU guidance links unlawful development to the lawfulness of deployment (EDPB[40]). Gathering web data, buying datasets of unclear origin, or repurposing customer data without consent can put the core asset at risk.

> **What good looks like:** Data provenance records for every training source (origin, license, consent basis, collection date), a dataset intake review, and contractual warranties from data vendors. This also strengthens investor diligence.

Challenge 5: Sensitive information disclosure through prompts and outputs

The OWASP Top 10 for LLM Applications (2025) ranks Prompt Injection first and Sensitive Information Disclosure second (OWASP Gen AI Security Project[41]). Indirect prompt injection is especially dangerous: instructions hidden in a web page, file, or email that the model processes can cause it to disclose sensitive data or misuse connected functions (OWASP[42]).

> **What good looks like:** Treat the model as an untrusted component. Never place secrets or authorization logic in system prompts. Apply input and output filtering for PII and secrets, segregate untrusted content, and conduct adversarial testing that treats the model as a potentially compromised user (OWASP[43]).

Challenge 6: Authorization failures in retrieval (RAG) and multi-tenant isolation

A RAG system that searches one shared index and relies on the model to "only show the right documents" will eventually leak data across users or tenants. OWASP's 2025 list added Vector and Embedding Weaknesses, including cross-tenant leakage and embedding inversion (Gravitee[44]).

> **What good looks like:** Identity-aware retrieval in which the user's identity and entitlements are enforced at query time by deterministic code, not by the model. Use per-tenant indexes or enforced metadata filters, enforce source-system permissions, and test isolation as part of every release.

Challenge 7: Third-party model providers and the AI supply chain

Most AI startups depend on one or more foundation-model APIs, plus hosting, vector database, observability, and evaluation vendors. Each is a sub-processor that may log prompts, retain data, or

[39] Recording Law. https://www.recordinglaw.com/us-laws/data-privacy-laws/ai-and-data-privacy/
[40] EDPB. https://www.edpb.europa.eu/news/edpb-opinion-on-ai-models-gdpr-principles-support-responsible-ai_en
[41] OWASP Gen AI Security Project. https://genai.owasp.org/llm-top-10/?cat=253
[42] OWASP. https://genai.owasp.org/llmrisk/llm01-prompt-injection/
[43] OWASP. https://genai.owasp.org/llmrisk/llm01-prompt-injection/
[44] Gravitee. https://www.gravitee.io/blog/owasp-top-10-for-llm-applications-2025-a-practical-guide

© 2026 Azure Dragon Cyber | azuredragoncyber.com | Informational only; not legal advice.

use it to improve services unless contractually prevented. IBM reported that 60% of studied organizations experienced direct data compromise from AI supply chain and model attacks (IBM[45]).

> **What good looks like:** A sub-processor register that includes AI providers; enterprise or zero-data-retention terms where available; data processing agreements that prohibit training on customer data; region pinning; and review of model and package provenance (OWASP LLM03 Supply Chain).

Challenge 8: Agentic AI and excessive agency

When AI agents can send emails, change records, run code, or call APIs, an authorization mistake becomes an action, not just a disclosure. OWASP recommends giving applications their own tokens, handling functions in code rather than delegating to the model, restricting privileges, and requiring human approval for high-risk actions (OWASP[46]).

> **What good looks like:** Treat each agent as a privileged non-human identity with a scoped workload identity, an allowlist of tools, short-lived user-delegated tokens, policy enforcement outside the model, approval gates for consequential writes, complete audit logs, rate limits, and a tested kill switch. Protocols that connect models to tools are integration layers, not authorization boundaries.

Challenge 9: Shadow AI inside the startup

Employees and contractors at startups adopt new AI tools constantly, often pasting code, customer data, or roadmaps into unapproved services. IBM found that 20% of organizations suffered a breach involving shadow AI, that high levels of shadow AI added about $670,000 to average breach costs, and that only 37% had policies to manage or detect it (IBM Newsroom[47]). Shadow AI breaches compromised customer PII in 65% of cases, compared with 53% overall (IBM Newsroom[48]).

> **What good looks like:** A short acceptable-use policy, a list of approved AI tools with enterprise terms, SSO enforcement, browser or endpoint DLP for high-risk data, and a fast approval path so employees are not tempted to work around controls.

Challenge 10: Cross-border transfers and data residency

AI inference and training frequently happen in a different country than the data subject, sometimes chosen automatically by the provider. EU customers will ask for transfer mechanisms and EU hosting; regulated U.S. customers may require domestic processing. Foreign-adversary data restrictions are also an emerging enforcement area for U.S. data flows (Privacy + Security Academy[49]).

---

[45] IBM. https://www.ibm.com/think/insights/data-matters/cost-of-a-data-breach

[46] OWASP. https://genai.owasp.org/llmrisk/llm01-prompt-injection/

[47] IBM Newsroom. https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls

[48] IBM Newsroom. https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls

[49] Privacy + Security Academy. https://www.privacysecurityacademy.com/wp-content/uploads/2026/04/FTC-Privacy-Enforcement-in-2026-and-Beyond.pdf

© 2026 Azure Dragon Cyber | azuredragoncyber.com | Informational only; not legal advice.

> **What good looks like:** Documented processing locations for every model endpoint, regional deployment options, and transfer mechanisms in standard contracts.

Challenge 11: Transparency, automated decisions, and explainability

If an AI product influences hiring, lending, insurance, housing, education, or healthcare decisions, startups face a convergence of obligations: California ADMT notices, opt-outs, and human review by 2027 (FTI Consulting[50]), Colorado's notice-based regime in 2027 (McDermott Will & Schulte[51]), and EU high-risk requirements in late 2027 (European Commission[52]). Many startups are "providers" of these systems even when their customers are the ones making decisions.

> **What good looks like:** Classify each AI use case by decision impact early. For consequential use cases, build human review, reason codes, and notice templates into the product so customers can meet their own obligations.

Challenge 12: Children's and sensitive data

Health, biometric, financial, precise location, and children's data carry heightened obligations under most state laws, COPPA, and GDPR. Consumer-facing AI tools, such as chatbots, tutors, and wellness apps, can collect these categories without intending to. The Garante's OpenAI decision included findings on inadequate age verification (Reuters[53]).

> **What good looks like:** Decide explicitly whether the product will process sensitive or minors' data, implement age assurance where applicable, and apply stricter retention, access, and training exclusions to these categories.

Challenge 13: Overstated security, privacy, and AI claims

Marketing pressure leads startups to claim "your data is never stored," "bank-grade security," or "unbiased AI." These statements become enforceable promises. The FTC continues to pursue deceptive AI-capability and data-practice claims (FTC[54]), and its Section 5 authority centers on whether companies live up to what they tell consumers (FTC[55]).

> **What good looks like:** A security and privacy review step for marketing copy, trust-center content, and sales questionnaire answers, with each claim traceable to a control and evidence.

Challenge 14: AI-specific incident response and customer assurance

Incident response plans written for ransomware and credential theft rarely cover prompt-injection exfiltration, model poisoning, a model provider's breach, a rogue agent action, or harmful generated

[50] FTI Consulting. https://www.fticonsulting.com/insights/articles/ccpa-final-regulations-what-weve-learned-what-comes-next
[51] McDermott Will & Schulte. https://www.mcdermottlaw.com/insights/colorado-ai-law-in-flux-comprehensive-replacement-bill-signed-after-federal-court-blocks-predecessors-enforcement/
[52] European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[53] Reuters. https://www.reuters.com/technology/italy-fines-openai-15-million-euros-over-privacy-rules-breach-2024-12-20/
[54] FTC. https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security/privacy-security-enforcement
[55] FTC. https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security/privacy-security-enforcement

output. Meanwhile, enterprise buyers increasingly require SOC 2 reports, ISO certifications, AI-specific questionnaires, and contractual commitments before signing.

> **What good looks like:** AI scenarios added to the incident response plan and tabletop exercises; logging sufficient to reconstruct what the model saw, retrieved, and did; breach-notification decision trees across jurisdictions; and a trust package (security overview, AI data-use statement, sub-processor list, and audit reports) ready for sales.

4. The AI Threat Landscape at a Glance

The OWASP Top 10 for LLM Applications (2025) provides a common vocabulary for AI application risk (OWASP Gen AI Security Project[56]). The privacy-relevant view:

OWASP 2025 Risk	Privacy and data protection impact	Priority controls
LLM01 Prompt Injection	Exfiltration of personal data via direct or indirect instructions	Untrusted-content segregation, least privilege, output validation, red teaming
LLM02 Sensitive Information Disclosure	PII, secrets, or confidential data in outputs or logs	Data minimization, PII filtering, no secrets in prompts, retention limits
LLM03 Supply Chain	Compromised models, packages, or providers expose data	Vendor due diligence, model provenance, dependency scanning
LLM04 Data and Model Poisoning	Manipulated training or retrieval data corrupts outputs	Dataset intake controls, integrity checks, source allowlists
LLM05 Improper Output Handling	Model output executed or rendered unsafely downstream	Treat output as untrusted input; encode and validate
LLM06 Excessive Agency	Agents take unauthorized actions on data or systems	Scoped identities, allowlisted tools, approval gates
LLM07 System Prompt Leakage	Exposure of business logic or embedded credentials	Keep secrets and authorization out of prompts
LLM08 Vector and Embedding Weaknesses	Cross-tenant leakage, embedding inversion	Per-tenant isolation, identity-aware retrieval
LLM09 Misinformation	Inaccurate outputs about individuals or decisions	Grounding, citations, human review for consequential uses
LLM10 Unbounded Consumption	Cost and availability attacks; model extraction	Rate limits, quotas, abuse monitoring

[56] OWASP Gen AI Security Project. https://genai.owasp.org/llm-top-10/?cat=253

5. Reference Control Architecture for AI Data Protection

The following design principles convert the challenges above into an architecture a small engineering team can implement.

1. **Identity first.** Authenticate users through the customer's identity provider where possible. Carry the user's identity through every AI request so that retrieval, tool calls, and APIs make authorization decisions on behalf of a specific user, not a shared service account.
2. **Authorization outside the model.** The model proposes; deterministic policy decides. Entitlement checks, consent checks, and tenant boundaries are enforced in code at the API gateway, retrieval layer, and tool layer.
3. **Minimize before you send.** Strip, mask, or tokenize personal data before it reaches third-party models when the task does not require it. Retrieve only the fields needed to answer.
4. **Segregate by tenant and sensitivity.** Separate indexes, encryption keys, and storage for each enterprise tenant where feasible; apply stricter handling to sensitive categories.
5. **Control egress.** Allowlist the model endpoints and external tools the system may call. Block unexpected outbound destinations from AI workloads.
6. **Protect secrets.** Store API keys and tokens in a secrets manager; issue short-lived, scoped credentials to agents.
7. **Log for reconstruction, retain for necessity.** Log prompts, retrieved sources, tool calls, and decisions with enough detail to investigate incidents and answer access requests, while applying retention limits and access controls to the logs themselves.
8. **Test adversarially and continuously.** Include prompt injection, cross-tenant retrieval, and data-extraction tests in CI/CD and before major releases.
9. **Keep humans in the loop where it matters.** Require human approval for consequential actions and for decisions with legal or similarly significant effects.
10. **Design for deletion.** Maintain lineage from source records to indexes and training sets so deletion and correction can be executed and evidenced.

6. Framework Alignment

Startups do not need to adopt every framework. They need one organizing framework, one AI-specific lens, and the attestation their customers actually request.

Need	Recommended framework	How to use it
Overall security program structure	NIST Cybersecurity Framework 2.0	Organize the program around Govern, Identify, Protect, Detect, Respond, Recover
AI-specific cybersecurity	NIST Cyber AI Profile (IR 8596, draft)	Apply its Secure, Defend, and Thwart focus areas to AI systems (NIST CSRC[57])
AI risk beyond security (fairness, validity, accountability)	NIST AI Risk Management Framework	Use Govern, Map, Measure, Manage for AI use-case risk assessment
AI application threats	OWASP Top 10 for LLM Applications	Threat modeling and testing checklist
Customer attestation (U.S.)	SOC 2 (Security, Confidentiality, Privacy criteria)	Map AI controls to existing Trust Services Criteria
Customer attestation (global)	ISO/IEC 27001, with ISO/IEC 27701 for privacy	Certifiable information security and privacy management systems
AI management system assurance	ISO/IEC 42001	Certifiable AI management system for buyers who request it

NIST released the preliminary draft Cyber AI Profile in December 2025, and it is designed to be used with CSF 2.0 and the AI RMF (KPMG[58]). Although voluntary, it may become a benchmark for regulators and others assessing cybersecurity diligence (KPMG[59]). Startups that align now will have less rework later.

A practical note: no official, regulator-issued mapping exists between every AI framework and every control catalog. Cross-walks are practitioner judgments and should be documented as such.

[57] NIST CSRC. https://csrc.nist.gov/pubs/ir/8596/iprd

[58] KPMG. https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2026/cybersecurity-nist-draft-cybersecurity-framework-for-ai-reg-alert.pdf

[59] KPMG. https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2026/cybersecurity-nist-draft-cybersecurity-framework-for-ai-reg-alert.pdf

7. Stage-Based Roadmap

Security and privacy investment should match company stage, data sensitivity, and customer expectations. The roadmap below assumes a B2B AI product handling customer personal data.

Stage 1: Pre-seed and Seed (Foundations, 0–90 days)

- Name an accountable owner for security and privacy (often a founder, supported by a fractional CISO or vCISO).
- Build the AI data inventory and data-flow diagram, including model providers and logs.
- Publish an accurate privacy notice that separates service delivery from model improvement.
- Execute data processing agreements with AI providers; select no-training and minimal-retention options.
- Enforce SSO and MFA, least-privilege cloud access, and secrets management.
- Adopt an AI acceptable-use policy and approved-tools list for the team.
- Implement tenant isolation and identity-aware retrieval from the first version.
- Record training-data provenance for every dataset.

Stage 2: Series A (Repeatable program, 3–12 months)

- Align to NIST CSF 2.0 and begin a SOC 2 Type I, then Type II.
- Run AI use-case risk assessments using the NIST AI RMF; complete California risk assessments where applicable.
- Add prompt injection and data-extraction testing to the release pipeline; commission an AI-focused penetration test.
- Stand up a data subject request process that covers indexes, logs, and training lineage.
- Add AI scenarios to the incident response plan and run a tabletop exercise.
- Publish a trust center with a sub-processor list and an AI data-use statement.
- Review all marketing and sales security claims for substantiation.

Stage 3: Series B and Beyond (Assurance at scale, 12–24 months)

- Pursue ISO/IEC 27001 and, where buyers request it, ISO/IEC 42001.
- Prepare ADMT notices, opt-outs, and human review where the product supports significant decisions (California and Colorado, January 1, 2027).
- Classify products under the EU AI Act and prepare for high-risk obligations ahead of December 2, 2027 where applicable.
- Assess whether California cybersecurity audit thresholds apply and plan the independent audit cycle.

- Establish KPIs and KRIs (for example, percentage of AI systems inventoried, open high-risk AI findings, time to fulfill deletion requests, and sub-processors without current reviews) reported to the board.

8. Key Compliance Dates

Dates reflect publicly available information as of October 2026. AI and privacy law is changing rapidly; confirm current requirements with qualified counsel before relying on them.

Date	Requirement	Applies to
January 1, 2026	California CCPA regulation updates effective; risk assessments required before new high-risk processing (CPPA[60])	Businesses subject to CCPA
April 22, 2026	Amended COPPA Rule compliance deadline (Privacy + Security Academy[61])	Services directed to or knowingly collecting from children under 13
August 2, 2026	EU AI Act transparency obligations for new systems (Mayer Brown[62])	AI systems offered in the EU
December 2, 2026	EU AI-generated content marking for systems on the market before August 2, 2026 (European Parliament[63])	Generative AI providers
January 1, 2027	California ADMT requirements (CPPA[64]); Colorado SB 26-189 effective (McDermott[65])	AI used for significant or consequential decisions
December 2, 2027	EU AI Act Annex III high-risk obligations (European Commission[66])	High-risk AI providers and deployers
December 31, 2027	California risk assessments due for high-risk processing that began before 2026 (MoFo[67])	CCPA businesses
April 1, 2028	First California cybersecurity audit certifications (revenue over $100M) and risk assessment submissions (CPPA[68])	Qualifying CCPA businesses
August 2, 2028	EU AI Act high-risk rules for AI embedded in regulated products (European Commission[69])	Product manufacturers and AI component providers

[60] CPPA. https://cppa.ca.gov/announcements/2025/20250923.html

[61] Privacy + Security Academy. https://www.privacysecurityacademy.com/wp-content/uploads/2026/04/FTC-Privacy-Enforcement-in-2026-and-Beyond.pdf

[62] Mayer Brown. https://www.mayerbrown.com/en/insights/publications/2026/07/eu-ai-act-news-digital-omnibus-on-ai-new-guidance-on-risk-classification-gpai-and-transparency-obligations

[63] European Parliament. https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-omnibus-on-ai

[64] CPPA. https://cppa.ca.gov/announcements/2025/20250923.html

[65] McDermott. https://www.mcdermottlaw.com/insights/colorado-ai-law-in-flux-comprehensive-replacement-bill-signed-after-federal-court-blocks-predecessors-enforcement/

[66] European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[67] MoFo. https://www.mofo.com/resources/insights/251007-ccpa-regulations-on-cybersecurity-risk-assessments

[68] CPPA. https://cppa.ca.gov/announcements/2025/20250923.html

[69] European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

© 2026 Azure Dragon Cyber | azuredragoncyber.com | Informational only; not legal advice.

9. Questions Every Founder and Board Should Be Able to Answer

1. Can we list every place customer personal data exists in our AI pipeline, including embeddings, logs, and model providers?
2. Do we train or fine-tune on customer data, and is that use documented, disclosed, and contractually permitted?
3. Could a regulator or customer force us to delete a model, and do we have provenance records that would let us defend it?
4. Is authorization enforced in code at retrieval and tool-call time, or do we rely on the model to behave?
5. Have we tested whether one tenant's user can retrieve another tenant's data through the AI interface?
6. What can our AI agents do without human approval, and can we stop them immediately?
7. Which AI tools are our own employees using, and what data have they shared with them?
8. Do any of our use cases support hiring, lending, housing, education, insurance, or healthcare decisions?
9. Is every security and AI claim on our website and in our sales materials backed by evidence?
10. If our model provider were breached tomorrow, do we know what data of ours they hold and how we would notify customers?

© 2026 Azure Dragon Cyber | azuredragoncyber.com | Informational only; not legal advice.

10. Conclusion

AI startups compete on speed, but in data privacy and protection, speed without structure creates debt that compounds: unknown data flows, unverifiable training sources, and access models that rely on the model to behave. That debt comes due during an enterprise security review, an investor's diligence, a regulator's inquiry, or a breach.

The startups that succeed will treat trust as part of the product. They will know their data, separate service delivery from model improvement, enforce authorization outside the model, govern their agents as privileged identities, and keep evidence ready. None of this requires a large security organization. It requires early decisions, clear ownership, and a roadmap matched to the company's stage. The regulatory calendar through 2028 gives founders a window to build these capabilities deliberately rather than reactively. The time to use it is now.

© 2026 Azure Dragon Cyber | azuredragoncyber.com | Informational only; not legal advice.

About the Author

Mark Adams is a cybersecurity, GRC, and technology risk leader with more than 30 years in IT, including service as a CISO. He has helped U.S. and European startups, particularly SaaS businesses, address cybersecurity and regulatory compliance challenges, and advises on security architecture, identity governance, AI security governance, data protection, third-party risk, and incident response.

Sources

- AI Act Service Desk, European Commission. Article 99: Penalties. https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-99
- Alston & Bird. California Expands the Cybersecurity & Privacy Impact of the CCPA. https://www.alston.com/en/insights/publications/2025/10/ccpa-cybersecurity-audits-admt-risk-assessments
- Byte Back. U.S. State Privacy Law Landscape Expands to 24 States. https://www.bytebacklaw.com/2026/06/u-s-state-privacy-law-landscape-expands-to-24-states-what-the-latest-legislative-wave-means-for-businesses/
- California Privacy Protection Agency. California Finalizes Regulations to Strengthen Consumers' Privacy. https://cppa.ca.gov/announcements/2025/20250923.html
- Davis Polk. Colorado repeals and replaces landmark AI Act. https://www.davispolk.com/insights/client-update/colorado-repeals-and-replaces-landmark-ai-act-publishes-draft-rules-public
- European Commission. AI Act: Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
- European Data Protection Board. Opinion 28/2024 on AI models. https://www.edpb.europa.eu/system/files/documents/2024-12/edpb_opinion_202428_ai-models_en.pdf
- European Data Protection Board. EDPB opinion on AI models: GDPR principles support responsible AI. https://www.edpb.europa.eu/news/edpb-opinion-on-ai-models-gdpr-principles-support-responsible-ai_en
- European Parliament. Digital Omnibus on AI, Legislative Train Schedule. https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-omnibus-on-ai
- Federal Trade Commission. Privacy and Security Enforcement. https://www.ftc.gov/news-events/topics/protecting-consumer-privacy-security/privacy-security-enforcement
- Federal Trade Commission. The FTC is on the Front Lines of AI Innovation & Regulation. https://www.ftc.gov/system/files/ftc_gov/pdf/ai-accomplishments-1.17.25.pdf
- FTI Consulting. CCPA Final Regulations: What We've Learned. https://www.fticonsulting.com/insights/articles/ccpa-final-regulations-what-weve-learned-what-comes-next
- Gravitee. OWASP Top 10 for LLM Applications (2025): A Practical Guide. https://www.gravitee.io/blog/owasp-top-10-for-llm-applications-2025-a-practical-guide
- IBM. What is the EU AI Act? https://www.ibm.com/think/topics/eu-ai-act
- IBM. What data leaders need to know from the Cost of a Data Breach Report 2025. https://www.ibm.com/think/insights/data-matters/cost-of-a-data-breach
- IBM. Hidden risk of shadow data and shadow AI. https://www.ibm.com/think/insights/hidden-risk-shadow-data-ai-higher-costs
- IBM Newsroom. 13% of Organizations Reported Breaches of AI Models or Applications. https://newsroom.ibm.com/2025-07-30-ibm-report-13-of-organizations-reported-breaches-of-ai-models-or-applications,-97-of-which-reported-lacking-proper-ai-access-controls
- KPMG. Cybersecurity: NIST Draft Cybersecurity Framework for AI. https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2026/cybersecurity-nist-draft-cybersecurity-framework-for-ai-reg-alert.pdf
- Latham & Watkins. Navigating New Obligations Under the CCPA Updated Regulations. https://www.lw.com/en/insights/navigating-new-obligations-under-the-ccpa-updated-regulations

- Legal 500. New CCPA Regulations Receive Final Approval. https://www.legal500.com/intelligence/united-states/privacy/new-ccpa-regulations-receive-final-approval
- Mayer Brown. EU AI Act News: Digital Omnibus on AI. https://www.mayerbrown.com/en/insights/publications/2026/07/eu-ai-act-news-digital-omnibus-on-ai-new-guidance-on-risk-classification-gpai-and-transparency-obligations
- McDermott Will & Schulte. Colorado AI law in flux. https://www.mcdermottlaw.com/insights/colorado-ai-law-in-flux-comprehensive-replacement-bill-signed-after-federal-court-blocks-predecessors-enforcement/
- Morrison Foerster. CCPA Regulations on Cybersecurity, Risk Assessments. https://www.mofo.com/resources/insights/251007-ccpa-regulations-on-cybersecurity-risk-assessments
- NIST CSRC. IR 8596, Cybersecurity Framework Profile for Artificial Intelligence. https://csrc.nist.gov/pubs/ir/8596/iprd
- OWASP Gen AI Security Project. 2025 Top 10 Risk & Mitigations for LLMs and Gen AI Apps. https://genai.owasp.org/llm-top-10/?cat=253
- OWASP Gen AI Security Project. LLM01:2025 Prompt Injection. https://genai.owasp.org/llmrisk/llm01-prompt-injection/
- Privacy + Security Academy. FTC Privacy Enforcement in 2026 and Beyond. https://www.privacysecurityacademy.com/wp-content/uploads/2026/04/FTC-Privacy-Enforcement-in-2026-and-Beyond.pdf
- Recording Law. AI and Data Privacy: Legal Requirements (2026). https://www.recordinglaw.com/us-laws/data-privacy-laws/ai-and-data-privacy/
- Reuters. Italy fines OpenAI over ChatGPT privacy rules breach. https://www.reuters.com/technology/italy-fines-openai-15-million-euros-over-privacy-rules-breach-2024-12-20/
- Reuters. Italian court scraps 15-million-euro privacy watchdog fine on OpenAI. https://www.reuters.com/technology/italian-court-scraps-15-million-euro-privacy-watchdog-fine-chatgpt-maker-openai-2026-03-19/
- Reuters. Who governs AI? The federal government's challenge to state regulation. https://www.reuters.com/legal/legalindustry/who-governs-ai-federal-governments-challenge-state-regulation-what-organizations--pracin-2026-08-12/
- U.S. Department of Justice. Artificial Intelligence Litigation Task Force memorandum. https://www.justice.gov/ag/media/1422986/dl?inline
- White House. Ensuring a National Policy Framework for Artificial Intelligence. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

This white paper is provided for general informational purposes and does not constitute legal advice. Organizations should consult qualified legal counsel regarding their specific obligations.