Cyber Readiness Companion to the AI Vendor Assessment Framework

October 2025

Data & Trusted AI Alliance

Executive Summary

AI adoption is accelerating, yet many corporate projects still fall short. Our members have repeatedly raised this concern, and recent reports by MIT and McKinsey echo the same finding: AI initiatives often fail to deliver expected value. This shared signal points to a clear need for a better way to evaluate AI solutions—one that weighs benefits against risks.

To address this, Alliance members created the AI Vendor Assessment Framework (VAF), a structured, practical tool to help organizations evaluate AI vendors during the procurement process—ultimately leading to the thoughtful and efficient implementation of AI.

Purpose

Published 10.24.2025

Authored by the Data & Trusted AI Alliance

Artificial intelligence has the potential to transform organizations by improving decision-making, increasing efficiency, strengthening resilience, and driving innovation. When implemented responsibly and with the right safeguards, AI can help companies move faster and compete with greater confidence.

At the same time, cybersecurity has become one of the most significant risks in realizing that potential. The same systems that can detect anomalies and secure networks can also be used by adversaries to discover software vulnerabilities, automate exploitation, and conduct adaptive attacks at scale.

As organizations integrate AI into their operations, cyber readiness becomes both a security and a business continuity imperative. The threat of insecure or poorly integrated software is growing rapidly. Each new system expands the organization's threat surface area and multiplies its overall risk exposure. These risks affect not only IT infrastructure but also data supply chains, product ecosystems, and customer-facing systems.

The AI Vendor Assessment Framework (VAF) helps organizations evaluate third-party AI vendors across risk factors such as intellectual property protection and privacy, and value factors such as integration quality and performance. The Cyber Readiness Companion builds on that foundation. It is designed as a practical tool for security leadership and procurement leads, whether part of a centralized procurement function or a product or business line responsible for evaluating and implementing AI systems.

This companion also supports collaboration across disciplines. It encourages product, engineering, legal, privacy, and risk teams to communicate effectively about security, align expectations, and assess vendor readiness in ways that help the organization identify vulnerabilities, mitigate risk, and strengthen resilience.

The guide does not introduce a new domain to the VAF. Instead, it highlights how existing questions already surface security considerations and provides additional context to help organizations protect themselves as they adopt and integrate AI. Future versions of the framework may evolve to include a dedicated Security and Resilience domain informed by this work.

How to use this companion

This guide is designed to help organizations of any size strengthen cybersecurity as they adopt and integrate AI systems. It works best when used alongside the AI Vendor Assessment Framework (VAF) during vendor selection, contracting, and implementation.

Who should use it:

- Security Leaders and CISOs assessing vendor maturity and operational risk.
- Procurement or Business Leads evaluating AI tools for their teams.
- Product and Engineering Teams implementing AI systems in production environments.
- Legal, Privacy, and Risk Officers responsible for compliance and enterprise protection.

How to apply it

Phase	What to Do	Who Leads It
Before Procurement	▸ Identify relevant stakeholders ▸ review the Companion ▸ and align on risk classification and key requirements.	▸ Procurement ▸ Security
During Vendor Review	▸ Use the Cyber Readiness Companion to expand VAF questions on security, provenance, and monitoring. ▸ Document vendor responses and note gaps.	▸ Procurement ▸ Legal ▸ Security
After Contracting	▸ Apply guidance on segmentation, access control, monitoring, and incident response during implementation and review annually.	▸ Product ▸ Engineering ▸ Risk

How long it takes

Integrating the Companion into the vendor review process typically adds 30–60 minutes per vendor, but reduces exposure to significant operational and compliance risks later in deployment.

Why it matters

Each AI integration introduces new dependencies, data flows, and access points. Applying this guide helps teams:

- Identify hidden vulnerabilities early.
- Communicate about security risks using a shared language.
- Strengthen collaboration between business, security, and engineering.
- Align procurement decisions with organizational resilience goals.

Why cyber readiness matters

AI adoption reshapes the enterprise attack surface. Models process sensitive data, connect to critical systems, and learn from dynamic environments.

Traditional security assessments often overlook these dimensions, focusing on networks or applications while missing model provenance, data exposure, and adaptive risk.

A cyber-ready procurement process:

- Aligns business, legal, and technical teams around risk segmentation.
- Ensures vendor integrations respect existing security architecture.
- Builds in visibility, containment, and response mechanisms for AI incidents.
- Reinforces trust, accountability, and resilience across the AI supply chain.

How the AI VAF already supports security

VAF Domain	Security Relevance	Example Insight
Governance & Oversight	Clarifies who is accountable for AI risk management and incident escalation.	Vendor policies indicate whether they maintain internal security and ethics boards or cross-functional governance structures.
Data & Inputs	Evaluates data sourcing and management practices that protect integrity and privacy.	Vendors describe how they vet and secure training and fine-tuning data.
Model Performance & Behavior	Reveals model vulnerabilities through testing, evaluation, and monitoring.	Vendors show how they test for bias, drift, and adversarial manipulation.
Deployment & Monitoring	Assesses operational practices for maintaining model and system security post-deployment.	Vendors provide incident notification and change management processes.

The Cyber Readiness Companion deepens this lens—guiding teams to operationalize these questions through stronger collaboration between procurement, security, and other internal stakeholders.

Using this companion effectively

1 Collaborate Early

AI security is inherently cross-functional. Procurement and security teams should work in tandem with:

- Product and Engineering teams to identify technical dependencies and model integration points.
- Privacy and Compliance teams to align with data governance and regulatory expectations.
- Risk and Legal functions to ensure contracts reflect security standards and incident protocols.
- AI and ML Teams to assess vendor model architecture, data handling, and monitoring capabilities.

Procurement identifies how the AI tool interacts with business processes and sensitive systems.

Security defines baseline requirements—such as access control, data isolation, and monitoring—and reviews vendor responses through that lens. Together with these additional stakeholders, teams can **classify vendors** as low, medium, or high risk based on data exposure, operational impact, and dependency depth.

2 Align Questions & Controls

Use this guide to align VAF responses with internal security and governance frameworks, including:

- NIST AI Risk Management Framework (AI RMF): for mapping organizational and technical safeguards.
- ISO/IEC 42001: for aligning with AI management system standards.
- MITRE ATLAS: for adversarial threat modeling and incident response planning.

Document which VAF responses meet existing enterprise standards and where additional controls or contract language are required to close gaps.

3 Segment and Contain Capability

AI vendors often integrate into production systems or data flows.
Ensure deployment aligns with your organization's segmentation strategy:

- Sandbox high-risk models during testing and limit access to production environments.
- Enforce data boundaries between internal datasets and vendor-provided models.
- Audit API permissions and disable unused endpoints to reduce exposure.

When to escalate

During the vendor evaluation and integration process, teams may encounter questions or gaps that exceed their authority or expertise. Escalation is appropriate whenever a security, data protection, or compliance concern cannot be resolved with the information available.

Escalation is not a sign of failure; it ensures that the right specialists review high-impact or uncertain risks before a decision is made. It should be viewed as a normal part of collaborative due diligence between procurement, product, security, and legal teams.

Escalate immediately if:

- Model provenance is unclear or the vendor cannot describe data sources, training methods, or fine-tuning processes.
- Incident response obligations are missing or the vendor cannot provide escalation contacts, logs, or rollback mechanisms.
- Access controls are weak or environment separation (development, test, production) is not defined.
- Adversarial testing is absent and the vendor has not considered frameworks such as MITRE ATLAS.
- Security assurances conflict with your organization's internal policies or regulatory obligations.

When escalation is necessary:

1. Alert your organization's Security or Risk Officer to review the issue and determine next steps.
2. Engage Legal and Privacy teams if personal or sensitive data may be exposed.
3. Document findings and mitigation actions in the vendor's assessment record to maintain transparency and accountability.

Why this helps:

- It tells readers when in the process escalation may occur (usually during procurement or pre-deployment reviews).
- It normalizes escalation as a standard control, not a bottleneck.
- It reinforces the culture of cross-disciplinary communication the Companion promotes.

Key cyber readiness considerations

❶ Model Provenance & Integrity

Understanding where an AI model comes from and how it was trained is essential to managing security and compliance risk.
Procurement teams should ask vendors to:

- Document model lineage, including training data sources, fine-tuning datasets, and licensing.
- Identify third-party or open-source components and how they are monitored for vulnerabilities.
- Provide attestations or third-party audits verifying model integrity and supply chain traceability.

Why it matters: Provenance provides visibility into the origins and dependencies of a model, critical for incident investigation and accountability.

❷ Vendor Security Posture

AI vendors introduce new dependencies that can become entry points for attackers.
Assess whether the vendor maintains a mature security program by asking:

- Do you apply standard controls such as MFA, encryption, and vulnerability management?
- How are model APIs, endpoints, and hosting environments protected?
- What measures detect and mitigate prompt injection, data poisoning, or model theft?
- Have you aligned your security posture with NIST AI RMF, ISO/IEC 42001, or adversarial testing frameworks such as MITRE ATLAS?

Why it matters: AI systems often run in shared or cloud environments, increasing the risk of data leakage and unauthorized access.

❸ Segmentation & Access Control

Every AI tool should have a clearly defined security boundary.

- Limit data and system access to the minimum necessary for operation.
- Require vendors to support role-based access control and environment separation (development, testing, production).
- Confirm where data is stored and processed, including jurisdictional safeguards.

Why it matters: Isolating AI environments reduces the blast radius of potential compromise or misuse.

Key Cyber Readiness Considerations

4 Incident Response Readiness

AI incidents are distinct: model drift, corrupted training data, or unauthorized fine-tuning can lead to unpredictable behavior.
Ensure vendors can support rapid investigation and containment:

- Include AI-related events in enterprise incident response playbooks.
- Require vendors to provide real-time logs, escalation contacts, and rollback mechanisms.
- Encourage use of MITRE ATLAS or similar frameworks to test and simulate common adversarial attacks (e.g., evasion, poisoning, inference).
- Establish SLAs for reporting and collaboration during incidents.

Why it matters: Shared accountability between vendors and buyers ensures timely response and mitigates reputational and operational harm.

5 Monitoring & Continued Assurance

AI models evolve over time, so must monitoring.

- Require vendors to share telemetry, performance reports, and audit logs.
- Define thresholds for model drift or degradation and triggers for review or retraining.
- Conduct periodic reassessments of vendor risk as systems update.

Why it matters: Continuous monitoring sustains confidence in AI performance and safeguards against emerging threats.

Putting it all together

When paired with the VAF, this companion helps organizations move from vendor assessment to operational assurance.

Recommended Process:

1. Map VAF responses to internal cybersecurity requirements.
2. Collaborate across procurement, legal, security, product, and risk functions.
3. Segment systems and isolate high-risk integrations.
4. Monitor for drift, anomalies, or changes in model provenance.
5. Review vendors regularly to align with evolving business and regulatory expectations.

Looking Ahead

As members continue adopting generative and predictive AI, the **Data & Trusted AI Alliance (D&TA)** will integrate these cyber readiness principles into future iterations of the VAF and related guidance—potentially introducing a *Security and Resilience* domain reflecting best practices across the Alliance.

For now, this companion serves as a practical tool to strengthen the security, accountability, and trustworthiness of AI procurement and deployment today.

Technical appendix

For Security and Technical Teams

This appendix provides a bridge between the VAF's governance-oriented structure and technical threat modeling and assurance frameworks. It enables CISOs, AI security leads, and technical reviewers to translate vendor responses into actionable assessments.

Cross-alignment Framework

Reference Framework	Purpose	Use in VAF Companion Context
NIST AI Risk Management Framework (AI RMF)	Guides organizations to manage AI risk systematically through governance, mapping, measurement, and management functions.	Use to benchmark vendor risk management maturity and align security practices with enterprise-wide AI governance.
ISO/IEC 42001	Establishes management system standards for responsible AI, including security and risk controls.	Use to verify vendors maintain certified AI management systems and operationalize AI-specific security controls.
MITRE Atlas	Catalogs adversarial tactics, techniques, and procedures (TTPs) used to attack AI systems.	Use to test vendor detection and response capabilities against known AI attack vectors such as data poisoning, model inversion, or evasion.

Example Mapping of MITRE Atlas to the VAF

MITRE Atlas Category	Relevant VAF Domain(s)	Example Vendor Assurance Practice
Data Poisoning	Data & Inputs	Vendor demonstrates data validation, anomaly detection, and traceability in training pipelines.
Model Evasion	Model Performance & Behavior	Vendor provides results of adversarial testing or red-teaming exercises.
Model Inversion / Extraction	Deployment & Monitoring	Vendor uses differential privacy or rate-limiting to prevent sensitive data inference.
Supply Chain Compromise	Governance & Oversight	Vendor maintains provenance logs and dependency tracking for model artifacts.
Prompt Injection	Security & Resilience (Companion Focus)	Vendor employs content filtering, sandboxing, or LLM guardrails to detect malicious prompts.

Implementation Tip for Members

Leaders can direct technical teams to:

1. Use MITRE ATLAS to simulate adversarial scenarios during vendor evaluation or red-teaming.
2. Align AI vendor risk criteria with NIST AI RMF “Manage” and “Measure” functions.
3. Integrate ISO/IEC 42001 compliance or alignment statements into vendor contracts or renewals.
4. Report findings back to procurement and risk teams in VAF-aligned language for consistent decision-making.

Glossary of core terms

Term	Definition / Reference
AI Supply Chain	The ecosystem of datasets, models, APIs, and vendors that make up an AI system. Each component can introduce security or integrity risks.
Data Provenance	The record of where data comes from, how it was processed, and how it was used in model training or operation. Enables traceability and accountability.
Model Drift	Gradual changes in model behavior or accuracy due to evolving data or environmental factors, requiring ongoing monitoring and retraining.
Prompt Injection	A manipulation technique where malicious input causes an AI model to behave unexpectedly or reveal sensitive information.
Data Poisoning	A deliberate attack that corrupts training data to alter model behavior or degrade performance.
Model Inversion / Extraction	Attacks that infer or reconstruct sensitive training data or model parameters.
Segmentation	The isolation of systems or datasets to limit access and reduce the potential impact of a breach or misuse.
Telemetry	Data collected from systems and models to monitor performance, detect anomalies, and support investigations.
Incident Response Plan	A structured process for detecting, reporting, and recovering from cybersecurity or AI-related incidents.
Adversarial Testing (Red-Teaming)	Controlled testing to identify vulnerabilities by simulating attacks against AI models or systems.
MITRE ATLAS	A framework cataloging adversarial tactics, techniques, and procedures (TTPs) used to target AI systems. Maintained by MITRE: atlas.mitre.org
NIST AI RMF	The U.S. National Institute of Standards and Technology's AI Risk Management Framework for governing and mitigating AI risks. Reference: NIST AI RMF
ISO/IEC 42001	The international management system standard providing requirements for responsible and secure AI operations. Information: ISO/IEC 42001 Overview

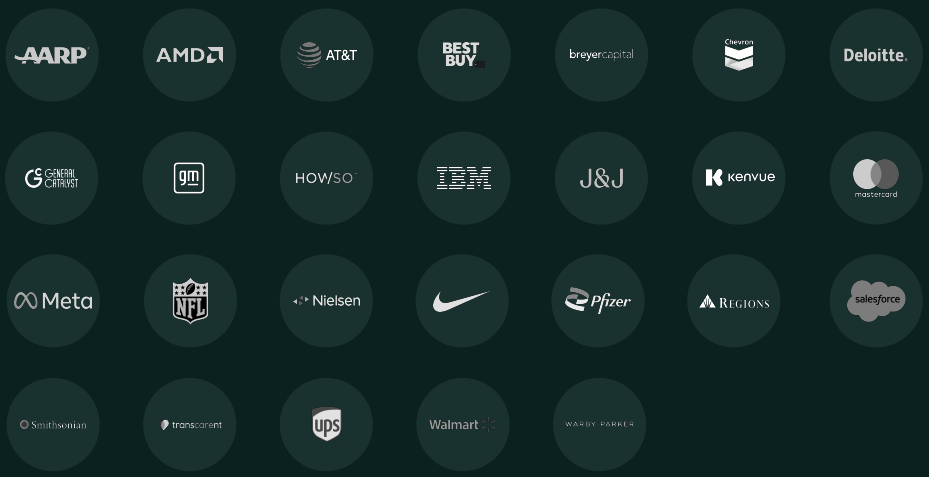
Questions or thoughts:

Please reach out to Camille Stewart Gloster
inquiries@dtaalliance.org

Data & Trusted AI Alliance

D&TA Alliance is a group of industry-leading enterprises, founded by member CEOs.

Created to share and develop business practices for applied and trusted AI—based on the enablers and impediments of real-world AI deployment.



Please note that the views expressed in these materials do not necessarily reflect the views of any D&TA member organization.