Velocis Technologies

AI Security Guardrails & Governance

A practical, identity-first, and security-operations-led approach to responsible enterprise AI adoption

Protect innovation without exposing sensitive data, intellectual property or enterprise systems.

Velocis combines governance, identity, Zero-Trust application control, endpoint security, continuous monitoring, human-risk management and incident response into an operational AI guardrail program.

Prepared by: Velocis Technologies LLC

Version: 1.0 | July 2026

Contact: info@velocistech.com | +1 903-436-4044 | www.velocistech.com

Executive Summary

Securing enterprise AI adoption without blocking innovation

Organizations are rapidly adopting generative AI, copilots, and autonomous digital workers across sales, finance, HR, legal, engineering, customer service, and operations. The same capabilities that boost productivity can also create new avenues for sensitive data disclosure, intellectual property leakage, unapproved “shadow AI,” identity abuse, insecure integrations, and high-speed automated actions.

Velocis provides an operational AI security and governance program that integrates policy, risk management, identity controls, application allowlisting, endpoint detection and response, vulnerability remediation, security awareness, centralized monitoring, and digital forensics. The goal is not to ban AI. It is to establish a secure path for approved use cases, block or constrain unapproved activity, and provide verifiable evidence that controls are operating as intended.

The program aligns the four NIST AI Risk Management Framework functions — Govern, Map, Measure, and Manage — with ISO/IEC 42001’s management-system approach. The NIST AI RMF is voluntary and designed to help organizations incorporate trustworthiness into the design, development, use, and evaluation of AI systems. ISO/IEC 42001 specifies requirements for establishing, operating, and continually improving an Artificial Intelligence Management System (AIMS). [1][4]

Key outcome: A controlled AI environment in which every sanctioned AI use case has a business owner, a risk tier, approved data boundaries, managed identities, technical guardrails, monitoring, response playbooks and auditable evidence.

What this document provides

- A practical Velocis control model for sensitive data, intellectual property, shadow AI, and digital-worker identities.
- A technology-by-technology explanation of how the current Velocis stack supports AI security.
- A clear statement of where complementary controls, such as enterprise DLP, CASB/SWG, or an AI gateway, may be needed.
- Detailed mappings to NIST AI RMF 1.0, the NIST Generative AI Profile, and ISO/IEC 42001.
- A phased implementation roadmap, measurable outcomes, and a managed-services operating model.

1. Velocis AI Guardrail Model

From departmental AI adoption to protected business outcomes

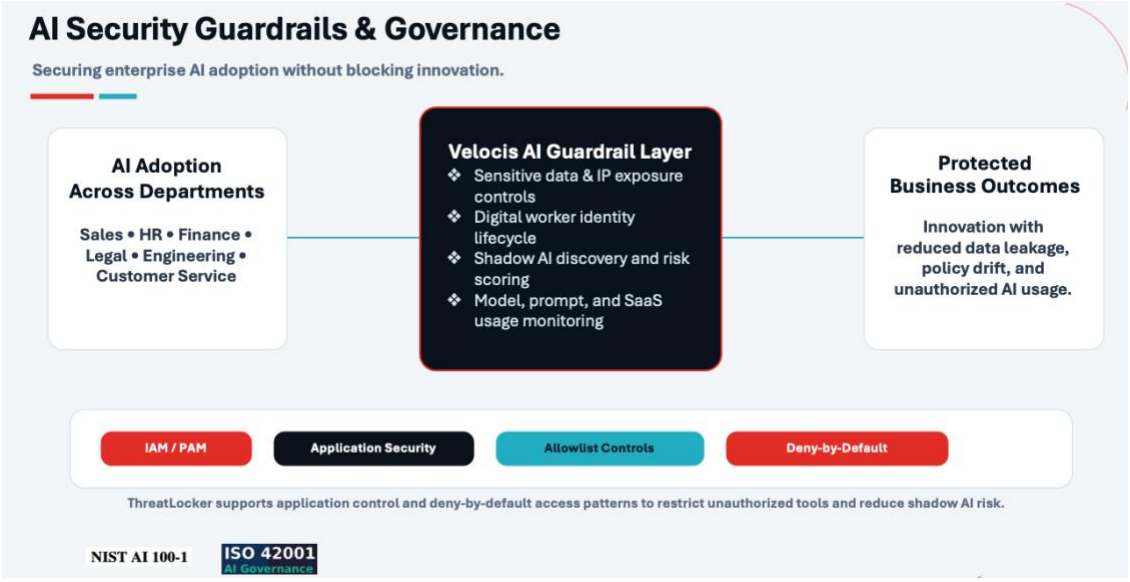

Figure 1. Velocis AI Security Guardrails & Governance model.

The Velocis guardrail layer sits between business adoption and business outcomes. It establishes rules for what can be used, which identities may use it, what data may be processed, how AI tools and agents can interact with enterprise systems, and how every material event is logged and reviewed.

Design principles

Principle	How Velocis applies it
Secure by default	Unapproved AI software, extensions, scripts and agents are blocked or restricted until explicitly approved.
Identity first	Humans, service accounts, bots and AI agents receive unique identities, accountable owners and least-privilege access.
Data centric	Controls follow the sensitivity of the information, not simply the application being used.
Human accountable	High-impact decisions and privileged actions retain human review, approval, and override.
Continuously monitored	AI activity, identity events, endpoint behavior, and control failures feed the Q-SOC for detection and response.
Evidence driven	Policies, approvals, test results, access reviews, incidents and corrective actions are retained as audit-ready evidence.

2. Target Control Architecture

How Velocis brings governance, identity, data, endpoint, and SOC controls together

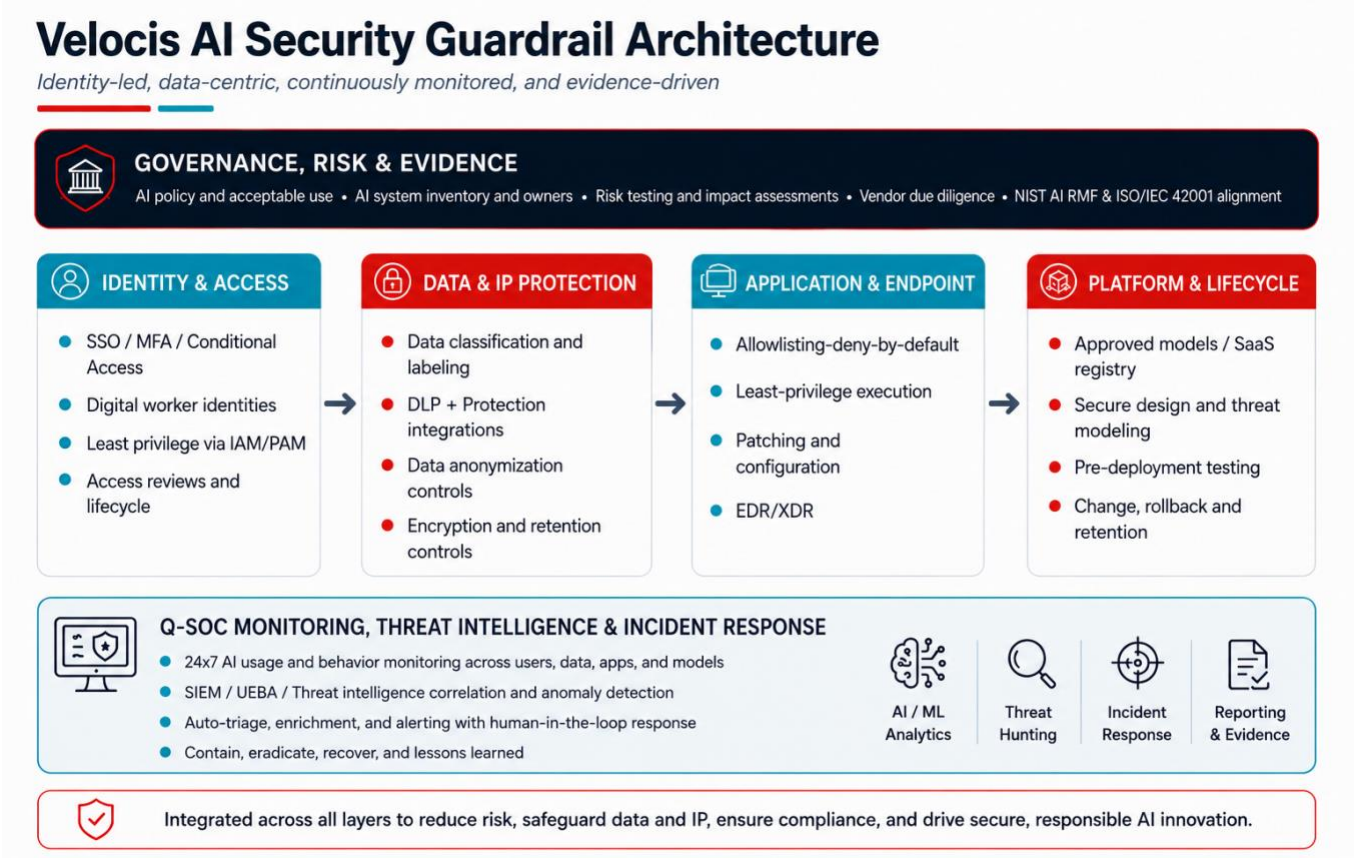

Figure 2. Reference architecture for Velocis AI security guardrails.

The architecture is intentionally vendor-neutral. Velocis can use the client's existing identity, cloud, data-protection, and AI-platform controls and integrate them with the Velocis-managed stack. This "no rip-and-replace" approach reduces deployment time while preserving a consistent governance and monitoring model.

Control boundary, Application Security, EDR, and identity controls can stop unapproved software and constrain access, but they do not replace prompt-level inspection, model safety evaluation, or enterprise DLP. When these capabilities are required, Velocis integrates the client's native cloud/AI controls or a purpose-built security gateway and forwards the resulting telemetry to Q-SOC.

3. Current Velocis Technology Stack

How existing tools and expertise support AI security

Capability	AI security role	Evidence produced
Velocis AI Governance / vCISO	AI policy, governance charter, risk appetite, use-case inventory, risk tiering, impact assessments, vendor reviews and executive reporting.	Governance records, approved-use register, risk register, impact assessments, management-review minutes.
Gurucul NG-SIEM / UEBA + Velocis Q-SOC	Centralize identity, endpoint, cloud, SaaS and AI-platform logs; correlate events; detect anomalous user/entity behavior; score risk; manage cases and investigations. Gurucul describes UEBA as behavior analytics, anomaly detection, and dynamic risk scoring. [7]	Log-source inventory, use cases, alerts, cases, analyst notes, SLA reports, monthly risk dashboards.
ThreatLocker	Application allowlisting and deny-by-default execution for unapproved AI clients, scripts and extensions; Ringfencing to constrain file, process and network interaction; elevation and storage/network controls where licensed. ThreatLocker states that only approved software runs and that Ringfencing can restrict file, program, and internet interaction. [5][6]	Application inventory, allow/deny policies, approval history, blocked-execution events, elevation records, and policy exceptions.
CrowdStrike Falcon	Endpoint detection and response, process/network telemetry, threat hunting, containment, and identity threat detection, where licensed. CrowdStrike positions its platform around endpoint, identity and exposure visibility and response. [8]	EDR detections, host-isolation actions, identity alerts, investigation timelines, and remediation records.
Automox	Automated cross-platform patching, configuration enforcement, and vulnerability remediation, including scripted mitigations when a patch is unavailable. [9]	Patch compliance, vulnerability-remediation status, configuration results, and exception tracking.
Okta / Microsoft Entra / Oracle IAM / AWS IAM	SSO, MFA, conditional access, RBAC/ABAC, privileged access, lifecycle workflows, access reviews and workload/non-human identity governance. Microsoft notes that workload identities require dedicated lifecycle and conditional-access controls. [11]	Identity inventory, owners, access assignments, access reviews, PIM/JIT records, token and service-principal policies.
OutThink Human Risk Management	Role-based and adaptive awareness, phishing simulation, targeted nudges and behavior-focused risk reduction. [10]	Training completion, risk scores, targeted campaigns, policy acknowledgments, and behavior trends.
Velocis ASM / DRP / Threat Intelligence	Discover externally visible AI-related assets, exposed credentials, impersonation, data leaks, brand abuse and third-party risk signals across surface, deep and dark web sources.	Exposure register, findings, takedown/support actions, threat briefs, and remediation tracking.
Velocis DFIR	Investigate suspected AI data leakage, account abuse, malicious automation, prompt-injection impact, malware, insider activity and third-party compromise; preserve evidence and support recovery.	Chain of custody, forensic timeline, indicators, root-cause report, containment actions, and lessons learned.

Capability	AI security role	Evidence produced
GRC / compliance automation	Maintain control library, evidence, policy lifecycle, audit findings, corrective actions, and cross-framework mappings.	NIST profile, ISO/IEC 42001 readiness matrix, Statement of Applicability support, evidence index, and audit trail.

Technology availability depends on client licensing, platform architecture, and selected service scope.

4. AI Security Control Domains

Operational controls for the most common enterprise AI risks

4.1 Governance, policy, and AI inventory

Velocis begins by establishing a single source of truth for AI. Every AI service, model, copilot, agent, API integration, and material use case is registered with an accountable business owner, a technical owner, data classification, risk tier, lifecycle status, and an approval decision. This directly reduces shadow AI and makes subsequent monitoring meaningful.

- Enterprise AI acceptable-use policy and prohibited-use rules.
- AI governance charter, decision rights, and RACI (Responsible, Accountable, Consulted, Informed) matrix.
- Approved AI catalog and exception process.
- Risk-tiering methodology based on data sensitivity, autonomy, business impact, and external exposure.
- AI impact assessment covering security, privacy, legal, operational, safety, and reputational consequences.
- Third-party due diligence and contractual requirements for retention, data use, subprocessors, security, and incident notification.

4.2 Sensitive data and intellectual-property protection

The core rule is that AI systems may access only the data necessary for an approved purpose. Velocis translates existing data classification and retention rules into AI-specific handling requirements and technical enforcement mechanisms.

- Define prohibited data classes for public or consumer AI services, including regulated records, credentials, source code, trade secrets, unreleased financial information, and client-confidential content.
- Require approved enterprise tenants, contractual data protections, and configured retention settings for sanctioned services.
- Integrate DLP, CASB/SWG, API gateway or cloud-native controls to inspect, label, redact, tokenize or block sensitive content before it reaches an external model.
- For retrieval-augmented generation (RAG), enforce source-level authorization so the AI response cannot reveal content the requesting identity could not access directly.
- Keep secrets, tokens and API keys outside prompts and code; use managed secret stores, short-lived credentials and rotation.
- Log data-policy violations and forward them to Gurucul/Q-SOC for correlation, case management, and response.

4.3 Identity management for digital workers

AI agents and other digital workers are treated as first-class, artificial identities rather than shared service accounts. Each agent must have a unique identity, a named owner, an approved purpose, permitted tools, scoped

data access, an expiration/retirement date, and a complete audit trail. Okta and Microsoft both describe AI agents or workload identities as non-human identities requiring managed credentials, fine-grained permissions, lifecycle controls, and auditability. [11][12]

- Register every agent, bot, service principal, and automation account in the identity inventory.
- Separate human authentication from agent authorization; never allow an agent to inherit unrestricted user access.
- Use least privilege, role- or attribute-based access, privileged identity management, and just-in-time elevation.
- Prefer short-lived OAuth/OIDC tokens, managed identities, and signed workload identities over static credentials.
- Require human approval for high-impact actions such as deleting records, transferring funds, changing security controls, or isolating production systems.
- Conduct periodic access reviews and automatically disable orphaned, inactive, or expired digital workers.

4.4 Shadow AI prevention and deny-by-default control

Velocis combines governance with enforceable technical controls. ThreatLocker can inventory applications, enforce deny-by-default execution, and restrict how approved applications interact with files, other processes, and the internet. These capabilities can block unapproved AI desktop clients, command-line tools, scripts, and browser extensions while allowing approved business tools. [5][6]

- Create a sanctioned AI application and extension allowlist by user group, role, and endpoint class.
- Block unapproved executables, scripts, libraries, and browser extensions by default.
- Use Ringfencing to prevent approved applications from reading sensitive directories, launching risky child processes, or communicating with unapproved destinations.
- Use elevation control to prevent users or agents from obtaining broad administrative rights.
- Apply storage and network controls where licensed, and integrate secure web gateway/DNS/CASB controls for browser-based AI services.
- Route temporary business exceptions through documented approval, time-bound policy, and post-use review.

4.5 Secure AI lifecycle and third-party assurance

- Architecture and threat-model review before production use.
- Data provenance, quality, licensing, and retention review for training, fine-tuning, and RAG sources.
- Security testing for prompt injection, data exfiltration, excessive agency, insecure tool use, authorization bypass, and unsafe output handling.
- Vendor assessment covering model/provider responsibilities, subprocessors, training on customer data, logging, residency, breach notification, and service continuity.
- Change management for model versions, prompts, tools, permissions, and data sources, including rollback criteria.
- Retirement process that revokes identities, secrets, connectors, data access, and retention exceptions.

4.6 Continuous monitoring and incident response

The Q-SOC operationalizes AI governance. Controls are valuable only when failures, exceptions, and suspicious patterns are visible and acted upon. Velocis builds AI-specific monitoring use cases and response playbooks using telemetry from identity providers, endpoints, Application Security Tools (ThreatLocker), EDR, cloud platforms, AI services, DLP/CASB, APIs, and application logs.

Monitoring use case	Illustrative indicators
Sensitive-data leakage	Repeated blocked uploads, large prompt payloads, source-code transfer, regulated-data patterns or unusual outbound volume.
Shadow AI	Execution of unapproved AI clients, extensions, scripts or command-line tools; access to unsanctioned AI domains.
Digital-worker abuse	Agent access from unusual locations, new privileges, abnormal tool calls, high-volume actions, or attempts to bypass approval.
Prompt injection/tool abuse	Agent follows instructions embedded in untrusted content, invokes unexpected tools, or accesses unrelated data.
Credential compromise	API-key exposure, impossible travel, token replay, anomalous service-principal sign-ins or suspicious privilege changes.
AI supply-chain issue	Compromised plugin, connector, model dependency, vendor incident, or malicious update.

5. Example AI Access Decision Model

Policy logic that can be translated into technical controls

Scenario	Decision	Minimum controls
Approved enterprise AI; public/internal data	Allow	SSO/MFA; approved tenant; logging; standard retention; user awareness.
Approved enterprise AI; confidential data	Allow with controls	DLP/redaction; approved data sources; role-based access; no provider training where contractually supported; enhanced logging.
External consumer AI; confidential or regulated data	Deny	Block upload/use; alert Q-SOC; user coaching; incident review if data was transferred.
Unapproved AI desktop app, extension, or script	Deny by default	ThreatLocker block; exception workflow; inventory and risk review.
AI agent with read-only low-risk access	Allow with scoped identity	Unique non-human identity; short-lived token; source-level authorization; owner and expiry.
AI agent with privileged or destructive actions	Conditional	PIM/JIT; transaction limits; human approval; full audit; kill switch and rollback.
Experimental model or vendor pilot	Isolated pilot	Non-production data; sandbox; limited users; security test; time-bound approval; exit criteria.

6. NIST AI RMF Alignment

Mapping Velocis services to Govern, Map, Measure, and Manage

NIST AI RMF 1.0 organizes AI risk management into four functions. GOVERN is cross-cutting, while MAP, MEASURE and MANAGE are applied to specific AI systems and lifecycle stages. The NIST Generative AI Profile extends the framework with suggested actions for risks unique to or amplified by generative AI and emphasizes governance, content provenance, pre-deployment testing, and incident disclosure. [1][2][3]

Function	Primary objective	Velocis implementation activities	Technology/service enablers	Representative evidence
GOVERN	Policies, roles, accountability, and risk culture	AI governance charter; acceptable-use policy; ownership/RACI; risk appetite; legal/regulatory obligations; training; third-party governance.	Velocis vCISO/GRC; OutThink; IAM/PAM; vendor-risk reviews; executive reporting.	AI policy suite; AI inventory; risk register; RACI; training records; vendor assessments; management-review minutes.
MAP	Context, intended purpose, actors, data, and impacts	Document use case, stakeholders, autonomy, data flows, affected parties, dependencies, threat model, risk tier, and potential harms/benefits.	Architecture workshops; ASM/DRP; data-flow mapping; identity inventory; impact and privacy/security assessments.	Use-case profile; system/data-flow diagram; AI impact assessment; threat model; risk classification; dependency register.
MEASURE	Testing, evaluation, verification, and validation	Define metrics; test security, privacy and control effectiveness; assess data leakage, prompt injection, access controls, model behavior, resilience and monitoring coverage.	ThreatLocker validation; CrowdStrike telemetry; Automox posture; penetration testing/red teaming; Gurucul analytics; access reviews.	Test plans/results; control validation; vulnerability status; access-review results; model/vendor test evidence; KPI/KRI dashboard.
MANAGE	Prioritize, treat, monitor, and communicate risk	Select treatment; implement safeguards; accept/transfer/avoid risk; monitor residual risk; respond to incidents; retire unsafe systems; communicate material events.	Q-SOC; SIEM/UEBA; DFIR; incident playbooks; containment; corrective actions; GRC workflow.	Risk-treatment plan; alerts/cases; incident reports; corrective actions; exception approvals; residual-risk acceptance; retirement evidence.

Cybersecurity-focused NIST Generative AI Profile priorities

Priority risk area	Velocis control approach
Information security	Threat modeling, secure architecture, least privilege, EDR, allowlisting, vulnerability remediation, prompt-injection and agent-tool testing.
Data privacy and IP	Data classification, DLP/redaction, approved data sources, retention controls, RAG authorization and vendor contract review.
Value-chain and component integration	Third-party/model/plugin inventory, dependency assessment, software/model provenance, vendor incident obligations, and change control.
Human-AI configuration	Clear user guidance, role-based training, human approval for high-impact actions, override and escalation paths.
Confabulation and output risk	Use-case-specific validation, source attribution where appropriate, quality checks, human review, and defined unacceptable-output criteria.
Incident disclosure	AI incident taxonomy, evidence preservation, internal escalation, customer/regulator communication criteria, and post-incident lessons learned.

Recommended NIST profile output: A current-state and target-state profile, prioritized risk-treatment plan, system-specific test plan, defined metrics, incident-disclosure criteria, and an evidence index linking each outcome to an owner and control source.

7. ISO/IEC 42001 Alignment

Building an Artificial Intelligence Management System (AIMS)

ISO/IEC 42001 is a management system standard for organizations that develop, provide, or use AI systems. It applies the Plan-Do-Check-Act model to policies, objectives, processes, risk management, performance evaluation, and continual improvement. The mapping below is an implementation crosswalk; formal certification requires a licensed copy of the standard and an independent accredited certification body. [4]

ISO/IEC 42001 clause	Management-system intent	Velocis implementation	Representative evidence
4 — Context of the organization	Define AIMS scope, internal/external issues, interested parties, AI activities, and process boundaries.	AI landscape assessment; regulatory and stakeholder requirements; scope statement; interface with existing ISMS/privacy/risk programs.	AIMS scope; interested-party register; AI inventory; process map; obligations register.
5 — Leadership	Establish policy, accountability, resources, and executive commitment.	AI policy; governance committee; accountable executive; system owners; decision rights; risk acceptance authority.	Approved AI policy; governance charter; RACI; meeting minutes; resource approvals.
6 — Planning	Address risks and opportunities; establish AI objectives and plans.	AI risk methodology; impact assessments; risk treatment; control selection; measurable objectives and roadmap.	Risk register; impact assessments; treatment plan; objectives; control applicability matrix.
7 — Support	Provide resources, competence, awareness, communication, and controlled documentation.	Role-based training; technical competency; document lifecycle; communication plan; evidence repository.	Training and competency records; communications; document-control log; evidence index.
8 — Operation	Plan and control AI lifecycle activities, data, use, suppliers, changes, and risk treatment.	Secure AI lifecycle; identity and access; data controls; ThreatLocker/EDR; vendor assurance; testing; change/incident management.	Design approvals; access records; test results; supplier reviews; change records; incident playbooks and cases.
9 — Performance evaluation	Monitor, measure, analyze, audit, and perform management review.	KPI/KRI dashboard; Q-SOC reporting; control testing; internal audit; access reviews; management review.	Monthly metrics; audit reports; access-review results; management-review outputs; findings log.
10 — Improvement	Correct nonconformities and continually improve the AIMS.	Root-cause analysis; corrective action; lessons learned; policy/control updates; maturity roadmap.	Corrective-action records; closure evidence; post-incident reviews; updated risk and control documents.

Representative Annex A control themes

Annex A theme	Velocis implementation focus
A.2 Policies related to AI	AI policy suite, acceptable use, secure development, and data-handling standards.
A.3 Internal organization	Governance committee, roles, segregation of duties, ownership, and escalation.
A.4 Resources for AI systems	Inventory of data, models, compute, tools, identities, skills, and supporting infrastructure.
A.5 Assessing impacts	Security/privacy/operational impact assessments, affected-party analysis, and residual-risk acceptance.
A.6 AI system lifecycle	Requirements, design, testing, deployment, operation, change, monitoring, and retirement controls.
A.7 Data for AI systems	Data quality, provenance, classification, access, retention, protection, and authorized use.
A.8 Information for interested parties	User instructions, limitations, transparency, incident communication, and stakeholder notices.
A.9 Use of AI systems	Approved-use rules, human oversight, competence, monitoring, and misuse prevention.
A.10 Third-party and customer relationships	Supplier due diligence, contracts, shared responsibilities, customer obligations, and ongoing monitoring.

ISO/IEC 42001 readiness output: An AIMS scope and process map, policy and objective set, risk and impact methodology, control applicability matrix, documented operational controls, KPI/KRI dashboard, internal-audit plan, management-review package and corrective-action register.

8. Combined Framework Evidence Model

One control environment, multiple assurance outcomes

Velocis avoids maintaining separate programs for every framework. A single control and evidence model can support NIST AI RMF profiling, ISO/IEC 42001 readiness, privacy/security obligations, and client assurance requests. Each control is linked to an owner, system scope, implementation status, evidence source, test frequency, exceptions, and corrective actions.

Evidence domain	Core artifacts
Governance	AI policy, charter, RACI, approved-use catalog, risk appetite, exception process.
Inventory and context	AI systems, models, providers, data sources, connectors, identities, owners, dependencies and affected parties.
Risk and impact	Risk register, AI impact assessment, privacy/security assessment, threat model, vendor assessment and residual-risk acceptance.
Technical controls	IAM policies, ThreatLocker policies, EDR configuration, DLP rules, network controls, vulnerability and patch status.
Testing and validation	Pre-deployment test plan, prompt-injection and authorization tests, red-team results, access reviews and control-effectiveness tests.
Operations	SIEM detections, Q-SOC cases, incident playbooks, escalation records, incident reports, and corrective actions.
Performance and improvement	KPIs/KRIs, internal audit findings, management review, lessons learned, maturity plan, and closure evidence.

Suggested governance cadence

Cadence	Activity
Continuous / daily	Q-SOC alerting, identity and endpoint telemetry, blocked AI activity, high-risk data events and incident handling.
Weekly	New AI requests, exceptions, urgent vendor changes, critical vulnerabilities, and open incident actions.
Monthly	AI inventory reconciliation, KPI/KRI dashboard, access anomalies, control exceptions, supplier events, and risk-treatment status.
Quarterly	Access reviews, policy/standard review, use-case revalidation, tabletop or technical test, executive steering committee.
Annually	AIMS scope and risk methodology review, internal audit, management review, framework profile update, and maturity roadmap.

9. Implementation Roadmap

From discovery to managed AI security

Figure 3. Illustrative implementation roadmap; duration varies with scope, licensing, and environment complexity.

Phase 1 — Discover and establish scope

AI discovery workshops; shadow AI assessment; inventory of AI systems, data flows and digital workers; regulatory and stakeholder analysis; current-state maturity assessment; immediate high-risk containment.

Phase 2 — Design governance and controls

Policy suite; governance committee and RACI; risk tiers; reference architecture; control matrix; NIST current/target profile; ISO/IEC 42001 readiness plan; logging and evidence requirements.

Phase 3 — Implement and integrate

IAM/PAM configuration; ThreatLocker allowlisting/ringfencing; EDR and patching; approved AI access; DLP/CASB or gateway integration; SIEM ingestion and detections; user training.

Phase 4 — Validate and operationalize

Technical testing; prompt-injection and authorization tests; access review; tabletop exercise; control-effectiveness assessment; incident playbooks; management acceptance of residual risks.

Phase 5 — Operate and improve

24×7 Q-SOC monitoring; monthly metrics; incident response; risk and exception management; vendor monitoring; internal audit support; management review and continual improvement.

10. Metrics and Success Criteria

Demonstrating control effectiveness and risk reduction

Metric domain	Illustrative measures
Coverage	Percentage of AI systems and agents inventoried; percentage with named owner, risk tier and approval status.
Identity	Percentage of digital workers with unique identity and owner; privileged agents under PIM/JIT; orphaned identities; access-review completion.
Data protection	Sensitive-data policy violations; blocked uploads; approved data-source coverage; retention exceptions; DLP false-positive/negative trends.
Shadow AI	Unauthorized AI applications/extensions blocked; unsanctioned domains observed; exception volume; repeat-user behavior.
Security posture	Critical vulnerabilities within SLA; endpoint and identity coverage; high-risk configuration findings; unsupported AI components.
Detection and response	Mean time to detect, triage, contain, and close AI-related incidents; playbook adherence; recurrence rate.
Human risk	Training completion, role-specific behavior scores, policy acknowledgment, and reduction in repeated risky behavior.
Governance	Open high risks, overdue treatment actions, vendor reviews completed, audit findings, corrective-action closure, and management-review cadence.

Management objective: Metrics should show whether risk is decreasing and controls are effective, not simply whether tools are deployed. Velocis establishes baselines, thresholds, owners, and escalation criteria for each KPI/KRI.

11. Velocis Service Deliverables

Modular services that can be adopted independently or as a managed program

Service module	Primary deliverables
AI Governance Foundation	Current-state assessment; AI inventory; policy suite; governance charter/RACI; risk-tier methodology; NIST AI RMF profile; ISO/IEC 42001 readiness gap assessment.
AI Security Architecture & Controls	Target architecture; digital-worker identity design; data/IP controls; ThreatLocker policies; EDR/patching integration; AI logging and Q-SOC use cases.
Secure AI Use-Case Review	Impact assessment; threat model; vendor review; data-flow analysis; testing plan; go/no-go recommendation; residual-risk acceptance package.
AI Red Team / Control Validation	Prompt injection, authorization, agent/tool abuse, data leakage, and resilience testing; findings and remediation validation.
Managed AI Security with Q-SOC	24×7 monitoring; alert triage; incident handling; monthly AI risk report; risk/exception tracking; threat intelligence and continuous control improvement.
ISO/IEC 42001 Readiness Support	AIMS scope, documentation, control applicability, evidence repository, internal-audit preparation, management-review support, and corrective-action tracking.
AI Incident Response & DFIR	Rapid investigation, evidence preservation, containment, root-cause analysis, regulatory/client support, and post-incident remediation.

Typical policy and procedure package

- Enterprise AI Acceptable Use Policy
- AI Governance Charter and RACI
- AI System and Digital Worker Inventory Standard
- AI Risk and Impact Assessment Procedure
- Sensitive Data and Prompt Handling Standard
- Digital Worker Identity and Access Standard
- Secure AI Development / Integration Standard
- AI Third-Party Risk Standard
- AI Logging, Monitoring, and Evidence Standard
- AI Incident Response and Disclosure Playbook
- AI Change, Exception, and Retirement Procedure

12. Conclusion

A secure path to enterprise AI adoption

AI security is not a single product. It is an operating model that connects business ownership, data governance, identity, endpoint control, cloud and AI platform safeguards, continuous monitoring and incident response. Velocis already operates the core technologies and disciplines needed to build this model: Q-SOC and SIEM/UEBA, IAM/PAM, ThreatLocker, CrowdStrike, Automox, OutThink, ASM/DRP, threat intelligence, GRC and DFIR.

By combining these capabilities with a structured NIST AI RMF profile and an ISO/IEC 42001-aligned management system, organizations can move from reactive AI restrictions to a controlled, auditable and innovation-friendly program. The result is reduced data leakage, stronger protection of intellectual property, accountable digital workers, fewer shadow AI pathways and faster response when issues occur.

Recommended first step: Conduct a focused AI Security & Governance Discovery Assessment to identify existing AI use, high-risk data flows, digital workers, shadow AI, current controls and the shortest path to an approved target state.

References

Official framework and technology sources

[1] NIST AI Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, NIST AI 100-1, January 2023.

[2] Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. National Institute of Standards and Technology, NIST AI 600-1, July 2024.

[3] NIST AI RMF Playbook. NIST AI Resource Center.

[4] ISO/IEC 42001:2023 — Artificial intelligence management systems. International Organization for Standardization, published December 2023.

[5] ThreatLocker Allowlisting. ThreatLocker official capability page.

[6] ThreatLocker Ringfencing. ThreatLocker official capability page.

[7] Gurucul User and Entity Behavior Analytics (UEBA). Gurucul official product page.

[8] CrowdStrike Falcon platform — endpoint, identity and exposure management. CrowdStrike official product information.

[9] Automox vulnerability remediation and autonomous endpoint management. Automox official platform information.

[10] OutThink AI-native Human Risk Management. OutThink official platform information.

[11] Microsoft Entra Conditional Access for workload identities. Microsoft Learn.

[12] Okta AI Agent Governance and Security. Okta official identity guidance.

Important notice

This document provides a practical approach to security and governance and does not reproduce the full text of ISO/IEC 42001. It is not legal advice and does not guarantee certification or regulatory compliance. Formal ISO/IEC 42001 certification requires an organization-specific implementation, a licensed copy of the standard, and assessment by an accredited certification body. Technology capabilities depend on licensing, configuration, and the client environment.