AN EXECUTIVE BRIEFING

The AI IT Security Audit™

Finding the exposure the green dashboard was never built to see.

Stephen R. Jordan

Founder, SRJ Consulting & Services LLC

Every dashboard is green.

And you still cannot prove what your AI is doing.

WHERE MOST CISOS ARE

AI is already inside the security perimeter.

Most mid-market and enterprise organizations already run AI across at least three business functions. Some is sanctioned. Most is not. Vendor AI features quietly activate inside Microsoft, Google, Salesforce, and every embedded platform the business already licenses.

The question is no longer whether AI is in the environment. It is whether the security program can prove what it is doing, and answer four questions when the regulator, insurer, enterprise customer, or audit committee asks.

Four questions the CISO cannot answer today:

01 What AI systems are actually running, sanctioned or not?

02 Which non-human identities can act in the environment?

03 How quickly can any of it be contained if it fails?

04 Where is the evidence a regulator would accept?

Three forces have made the AI security question personal:

3

The regulator now examines AI governance posture directly. SEC cybersecurity disclosure rules already touch AI when the exposure is material.

The enterprise customer has added an AI section to the security questionnaire. Ten to twenty-five specific questions that either land the deal or stall it.

The insurance carrier now asks AI exposure questions at renewal, and adds AI exclusions when the CISO cannot demonstrate containment capability.

None of these are going away. The timeline belongs to whoever moves first.

PILLAR 1

The Green Dashboard Fallacy

Every existing security dashboard shows green because it was built before AI agents entered the environment at scale. The tools are not lying. They are watching a control surface that no longer describes what is running.

The greatest AI security risk facing your organization is not that you are undefended. It is that leadership believes the existing program is already watching.

- The dashboard is a detection failure only in the mind of the reader. It is an assumption failure in the field.
- AI agents, MCP servers, embedded vendor features, and autonomous tools entered the environment faster than the operating model absorbed them.
- Naming the fallacy is the first move. Producing the evidence to close it is the audit.

THE EXPOSURE INSIDE THE GAP

The Four-Layer AI Environment

The AI in your business is not one thing. It is four, and only one of them is on the inventory the security program actually maintains.

WHAT LEADERSHIP SEES

Sanctioned AI Approved tools on the inventory, on the P&L, contractually reviewed.

WHAT IS ACTUALLY RUNNING

Sanctioned On the inventory. Roughly one tenth of the actual footprint.

Shadow Individual employees using AI without security approval or contract.

Embedded Vendor features quietly activated inside platforms already in use.

Agentic Autonomous agents and MCP servers acting on their own privileges.

PILLAR 2

Non-Human, Privileged, Auditable

> AI agents must be audited as privileged,
> non-human actors inside the control environment.

Once an agent can retrieve data, invoke tools, write records, generate code, or trigger a workflow, it is no longer a passive software feature. It is an operational actor with an identity, a privilege scope, and a blast radius.

Your access review was designed for human identities with enumerable behavior. It was not designed for identities whose behavior is shaped by inputs the organization does not control. That is not a competence gap in your IAM team. It is a foundational assumption gap in the model itself.

WHAT EVERY AUDIT NEEDS

Three instruments carry the method.

Every finding maps to one of three instruments. Each converts an anecdote into evidence.

01 **The Visibility Triangle**

Sorts every gap into visible, suspected, or undetectable. The zone determines the response, remediate, confirm, or bring into view.

02 **The Six-Domain Operating View**

Structures the work across governance, security operations, architecture, application security, third-party risk, and data protection.

03 **The Defensible AI Security Baseline**

A dated, scored standard across seven areas with a named owner for each, which turns a one-time audit into a program.

PILLAR 3

Confidence vs Evidence

What survives the meeting is not what the CISO believes. It is what the CISO can hand over.

CONFIDENCE

- A senior explanation of what should be in place.
- A dashboard that reports green.
- A vendor deck stating best practice.
- A verbal assurance to the audit committee.
- A responsible AI principles page.

EVIDENCE

- A dated exposure map across six domains, signed by the CISO.
- An identity inventory enumerating every non-human actor.
- A tested Red Button procedure with a timed shutdown record.
- A scored baseline with a named owner for each area.
- A four-page board pack the audit committee reads without translation.

S E L F - D I A G N O S T I C

Three signs your AI security is not defensible

01 **No one can enumerate the non-human identities.**

Ask IAM to list every service account, agent API key, MCP credential, and retrieval token that can act in the AI environment. If the answer takes longer than a week, the inventory does not exist.

02 **The AI incident response plan routes to "other".**

A prompt injection, model exfiltration, or agent misuse event does not fit the ransomware playbook. If the IR plan handles them as generic incidents, the runbook has not been retuned for AI-native threats.

03 **The last AI vendor review was a questionnaire.**

The vendor sent back yes to every control. No evidence, no data flow, no training rights, no incident notification clause verified. The signed questionnaire is not evidence. It is the absence of evidence.

PILLAR 4

A Framework Is Not a Deliverable

> Citing NIST AI RMF is not the same as producing the artifacts NIST asks you to keep.

Seven binding frameworks appear throughout the audit: NIST AI RMF, ISO/IEC 42001, OWASP LLM Top 10, OWASP Agentic Top 10, OWASP AIVSS, MITRE ATLAS, HITRUST AI, and Google SAIF. They are cited only where they actually bind, never as a compliance tour.

The evidence is the work. The framework is the citation. A regulator, insurer, or audit committee looks past the framework name and asks for the underlying artifact. That is what this audit produces. Framework alignment maps every clause to the artifact that satisfies it.

THREE REASONS THE QUESTION IS FORMING NOW

Why AI security exposure is now personal

THE REGULATOR

SEC cybersecurity disclosure rules under Item 1.05 already require material incident disclosure within four business days of a materiality determination. AI-related events qualify when the exposure is material. NYDFS Part 500 applies to covered financial services entities in New York. EU AI Act Article 15 is the security clause the audit answers to most directly.

THE ENTERPRISE CUSTOMER

Security questionnaires now include an AI section with ten to twenty-five specific questions covering non-human identity, agent boundaries, vendor training rights, and prompt-injection defenses. The response either lands the deal or stalls it. Most CISOs answering it today are writing evidence during the response window.

THE INSURANCE CARRIER

Renewal questionnaires now include AI exposure questions, and carriers are adding AI-specific exclusions and sub-limits to policies whose CISOs cannot demonstrate operational containment capability. A tested Red Button procedure with a timed shutdown record is the artifact the carrier is looking for.

PILLAR 5

The Audit Framework

The AI IT Security Audit examines six domains. Each produces its own evidence base, and together they form the exposure map.

01	02	03	04	05	06
Governance	**Security Operations**	**Architecture**	**Application Security**	**Third-Party Risk**	**Data Protection**

Every domain closes the same way: the Board Question your leadership must answer, the Evidence Required to answer it, the Common Failure Pattern that breaks the answer, and a 30-Day Move with a named output, a named owner, and a completion condition.

DOMAIN 01

Governance

The infrastructure that distinguishes a managed AI program from a tolerated one.

THE ARTIFACTS THIS DOMAIN PRODUCES

Named ownership

A specific human accountable for AI risk, with reporting line and escalation.

Written AI policy

Enforceable, reviewable, tested against NIST AI RMF and ISO/IEC 42001.

Regulatory crosswalk

A dated map of every regime that applies, kept current instead of rebuilt.

Board-ready summary

The four-page board pack, in the language the audit committee reads.

DOMAIN 02

Security Operations

Detection coverage retuned for AI-native threats the SOC was never built to catch.

WHAT THE AUDIT EVALUATES

Detection coverage

MITRE ATLAS techniques mapped against existing SIEM and EDR rules.

AI incident response

A tested playbook that routes prompt injection, model exfiltration, and agent misuse to named owners.

Red Button procedure

A rehearsed containment path with a timed shutdown record.

Threat intelligence

AI-specific feeds integrated into the SOC brief cycle, not treated as a side channel.

DOMAIN 03

Architecture

The non-human identity perimeter that traditional IAM was never designed to see.

WHAT THE AUDIT EVALUATES

Non-human identity inventory

Every service account, agent API key, retrieval token, and MCP credential enumerated.

Agent boundary matrix

Authorization scope, tool access limits, credential scoping, and human-override controls.

Zero Trust for AI traffic

Existing controls tested against agent-generated calls and cross-tenant flows.

Model gateway posture

Where LLM traffic enters and leaves the environment, and who governs the path.

DOMAIN 04

Application Security

Prompt injection audited as a privilege escalation vector, not as a research curiosity.

WHAT THE AUDIT EVALUATES

OWASP LLM Top 10 coverage

Testing methodology mapped to every top-10 category with named test cases.

Agentic Top 10 coverage

Excessive agency, tool integration risks, and memory poisoning tested in production paths.

Output validation

Regulated content and sensitive data intercepted before it reaches the user or downstream tool.

MLOps governance

Model deployment, retraining, and rollback paths reviewed against the audit trail expected.

DOMAIN 05

Third-Party Risk

AI vendors operating in a contractual blind spot the vendor questionnaire was never built to see.

WHAT THE AUDIT EVALUATES

AI vendor tier map

Every AI-enabled vendor classified by access depth, data handling, and blast radius.

Data flow map

What leaves the boundary, in what form, to whom, and under what contract.

Training rights posture

Whether the vendor is training on customer data, and how the contract reads today.

Silent-upgrade monitoring

A practice that catches AI features quietly activated inside platforms already in use.

DOMAIN 06

Data Protection

Prompts, outputs, embeddings, and retrieval indexes classified and governed as data.

WHAT THE AUDIT EVALUATES

Data classification for AI

Sensitive data identified across prompts, retrieval indexes, embeddings, and model outputs.

Lifecycle and retention

Documented controls for what an AI system stores, for how long, and how it forgets.

DLP for generative AI

Prevention rules tuned for the paths generative outputs actually take.

Cross-border inference

Every AI path that crosses jurisdictional lines, and the transfer mechanism relied upon.

THE ENGAGEMENT

How the audit is performed

01 Scope

Sessions with the CISO and named domain owners. The exposure profile the audit needs to survive is defined here.

02 Discovery

Sanctioned, shadow, embedded, and agentic AI inventoried across the six domains, in that order.

03 Evidence

Every finding scored on the Visibility Triangle. Undetectable gaps get the tooling and process changes required to bring them into view.

04 Baseline

The Defensible AI Security Baseline scored across seven areas with a named owner for each. Dated. Signed.

05 Readout

Executive readout with the CISO, the operating team, and the audit committee sponsor. The four-page board pack is handed over.

DELIVERABLES

What you receive

Six artifacts built to survive a regulator, an insurer, an auditor, an enterprise customer, or your own audit committee.

01 AI Exposure Map

Every AI system in the business, placed on one map across six operating domains.

02 Non-Human Identity Inventory

Every service account, agent API key, retrieval token, and MCP credential, enumerated and scoped.

03 AI Red Button Procedure

A rehearsed containment procedure with a named owner and a timed shutdown record.

04 AI Vendor Tier Map and Data Flow Map

Every AI-enabled vendor classified by access depth, with the data flow the audit can point to.

05 Regulatory Crosswalk

A dated document, not a fire drill. Kept current as regimes move.

06 Four-Page Board Pack

Exposure Map, Roadmap, Regulatory Posture, and Baseline. The four pages the audit committee reads.

THE PLAN YOU LEAVE WITH

The 90-day operating plan

Specific. Owned by name. Executable without additional headcount.

DAY 1–30

See

- Sanctioned, shadow, embedded, and agentic AI inventoried across the six domains.
- Non-human identities enumerated and scoped.
- Visibility Triangle marked for every finding.
- The four-page board pack drafted.

DAY 31–60

Score

- Defensible AI Security Baseline scored across seven areas with named owners.
- Regulatory Crosswalk dated and signed.
- AI Red Button procedure written and rehearsed.
- AI Vendor Tier Map and Data Flow Map completed.

DAY 61–90

Sign

- Board pack presented to the audit committee.
- Remediation roadmap sequenced and owned.
- Baseline signed by the CISO with a review cadence set.
- Program moved from audit to standing operation.

TO SET EXPECTATIONS CLEARLY

What this audit is not

This is a CISO-grade audit of AI security exposure, not a substitute for the security program itself.

✕ **Not a penetration test**	We do not attack production systems. The audit examines controls, evidence, and posture. Red-team engagements are scoped separately.
✕ **Not a compliance certification**	We produce the artifacts a certification asks for, but we are not the certifying body. Certification bodies rely on evidence like the artifacts this audit produces.
✕ **Not a general IT audit**	The AI Business Enablement Audit™ is a separate engagement. That audit is CEO-and-board grade. This one is CISO-grade.
✕ **Not a vendor recommendation**	No products are recommended. No tools are sold. No referral fees are taken from any AI vendor in the environment.

BUILT FOR

Who this audit is for

Designed for the executive who owns the answer, not the engineer who runs the tool.

SPONSORS

- Chief Information Security Officer
- Chief Information Officer
- Chief Risk Officer
- General Counsel
- Audit committee sponsors

SECTORS

- Financial services and insurance
- Healthcare and life sciences
- Public company IT and technology functions
- Professional services and legal
- Manufacturing, distribution, and regulated industry

Sized for mid-market through large multinational organizations.

ABOUT THE AUTHOR

Stephen R. Jordan

Founder, SRJ Consulting & Services LLC

Author of The AI IT Security Audit™, the fifth volume of The Operating Discipline for AI Library™, and the opening volume of Pillar II, AI Risk Governance & Security™. Stephen draws on three decades of senior security and operations leadership at Citi, Intel, McAfee, and Optiv.

Based in Dallas–Fort Worth, working with executives nationwide. SRJ Consulting & Services LLC delivers AI Business Services™ and AI Risk Governance & Security™ to organizations from mid-market to large multinational conglomerates.

BACKGROUND

Citi

Intel

McAfee

Optiv

Ready to prove your AI exposure is known, controlled, and governed?

Book a conversation with Stephen R. Jordan.

srjconsultingservices.com

SRJ Consulting & Services LLC

Dallas–Fort Worth · Nationwide