AN EXECUTIVE BRIEFING

The AI IT Security Implementation & Strategy™

The audit told you what was true on Tuesday. This is the program that keeps it true.

Stephen R. Jordan

Founder, SRJ Consulting & Services LLC

The audit is finished.
The environment kept moving.

WHERE MOST CISOS ARE

The audit produced a photograph. You are accountable for a moving object.

The inventory is complete and dated. The identity count is real. The vendor map is signed. Then an agent is deployed on Tuesday, a vendor activates an AI feature on Wednesday without telling you, and a business unit adopts a tool on Thursday.

Nothing in the audit was wrong. The problem is that an audit is an instrument for describing a state, and what a security leader owns is a rate of change.

Four questions the audit cannot keep answering:

01 What is running today that was not running at audit close?

02 Which agent actions are permitted without a human?

03 Can the suspension capability actually be executed, and how fast?

04 Is the vendor still acceptable, or only acceptable as of last year?

THE DEADLINE YOU DO NOT CONTROL

Four parties will ask you to prove it is governed. On their schedule.

4

The regulator Examines AI governance posture directly. The EU AI Act security obligations and SEC disclosure rules already reach AI exposure when it is material.

The enterprise customer The security questionnaire now carries an AI section. The answer either closes the deal or stalls it in procurement.

The insurance carrier Renewal now asks about containment capability, and adds AI exclusions where none can be demonstrated.

Your own audit committee Accepted the audit last cycle. This cycle they will ask what changed and what you did about it.

None of these are going away. The only variable is whether you arrive with a program or with intentions.

PILLAR 1

Governing Organizational Intelligence

> *Enterprise security is no longer protecting information.*
> *It is governing organizational intelligence.*

For thirty years the discipline organized itself around data. Where it lives, who can reach it, how it is encrypted, how you know when it left. Those questions remain valid and are no longer sufficient.

What runs in the environment now does not only hold data. It reasons over it, forms conclusions, and acts on those conclusions, frequently without a person in the loop and always faster than a person could intervene.

You are not securing a filing cabinet. You are supervising a decision maker that never sleeps, does not read policy, and holds credentials.

THE STRUCTURE OF THE PROGRAM

Four domains. Three chapters each. The same three questions in the same order.

GOVERNANCE & RISK

Decision rights, supervision, and the sanctioning pathway.

SECURITY OPERATIONS

Authority chains, circuit breakers, and continuous evaluation.

THIRD-PARTY & SUPPLY CHAIN

Live trust scores, bill of materials, concentration.

DATA PROTECTION & PRIVACY

Provenance, purpose enforcement, erasure engineering.

01 **WHAT AI IS BREAKING**

Working from the artifacts the audit handed over. Not theory.

02 **HOW AI REBUILDS IT**

The same technology creating the exposure is the only thing fast enough to govern it.

03 **THE ENDGAME**

Solution categories and a one-year operating model with an owner, a cadence, and a budget line.

PILLAR 2

Per Action Class. Never Per Tool.

> *Supervision policy written per tool is obsolete the moment a vendor ships a feature.*

A per-tool policy says this assistant is approved and that one is not. It has to be reopened on every release, every acquisition, and every feature activation you did not ask for.

A per-action-class policy says reading a record is one thing, drafting a communication is another, sending it externally is another, and moving money is another. Those classes are stable. The tools underneath them are not.

This distinction is the one external reviewers miss on first read. It is the difference between a policy that survives the year and one that does not.

WHAT CARRIES THE PROGRAM

Three instruments convert intention into a defensible capability.

01 The Human Supervision Matrix™

Which actions the machine may take alone, which require confirmation, and what evidence must exist before autonomy expands. Built per action class, reviewed monthly, expanded only on measured accuracy.

02 The Red-Button Protocol™

Suspension converted from a policy paragraph into a rehearsed procedure with a named executor, a dependency map, and a measured time to confirmed stop.

03 The Agent Authority Chain™

Identity, intent, authority, context, action, impact. Delegation traced end to end so that who authorized this has an answer that survives an inquiry.

PILLAR 3

A Claim Is Not a Capability

The distinction the auditor, the carrier, and the regulator are all testing for.

THE CLAIM	THE CAPABILITY
A policy stating AI systems can be suspended.	**A rehearsal record with a measured time to confirmed stop.**
A signed vendor questionnaire from last year.	**A live trust score with a direction and a threshold action.**
A statement that agents are monitored.	**An authority chain naming who delegated the right and when it expires.**
A responsible AI principles page.	**A supervision matrix signed per action class, dated.**
An assurance that erasure is honored.	**An erasure register with a sampled verification result.**

SELF-DIAGNOSTIC

Three signs the audit has already gone out of date.

01 **Nobody can name what changed since audit close.**

Ask for the list of AI systems that entered production after the audit was signed. If the answer is reconstructed rather than read from a register, the inventory is a document rather than a control.

02 **The kill switch has never been pulled.**

A suspension capability that has not been executed on a real system has no known time to stop and no known dependency list. It is a paragraph, and the carrier will read it as one.

03 **Vendor trust is still an annual attestation.**

The questionnaire tracks something that changes annually. Your vendors are shipping AI features weekly. The gap between those two rates is your exposure.

PILLAR 4

Categories Outlive Vendors

> *Not one vendor is named anywhere in this program.*
> *The entire solution landscape is described in categories.*

This market consolidates faster than a security architecture can be rebuilt. A procurement decision anchored to a named product has to be revisited every time that product is acquired, rebranded, or discontinued.

Anchored to a solution category and an architectural pattern, the decision survives. The frameworks are cited only where they actually bind: NIST AI RMF, ISO/IEC 42001, the EU AI Act, the OWASP LLM, Agentic, and MCP top tens, AIVSS, MITRE ATLAS, and Google SAIF.

The evidence is the work. The framework is only the citation.

THE CONNECTIVE TISSUE

The Agent Authority Chain™

A log that records a service account acted tells you what happened. It does not tell you who is accountable, which is the question you will actually be asked. Six elements, each with a domain owner and an evidence source.

IDENTITY	INTENT	AUTHORITY	CONTEXT	ACTION	IMPACT
Which system acted, and is it enumerated?	What objective, recorded before the work began?	Who delegated the right, and when does it expire?	Which sources shaped the reasoning, at what trust level?	What was done, which tools, which sub-agents?	What changed, and which human decisions relied on it?
Governance	**Governance**	**Governance**	**Third-Party Risk**	**Security Operations**	**Data Protection**

Authenticated is not authorized. The chain is what makes the difference provable.

PILLAR 5

The Twelve-Month Operating Model

Each of the four domains closes on a target operating model that names an owner, a cadence, a quarterly milestone, and a budget line. Every review ends with a decision, an owner, and a due date.

Q1

ACCOUNTABILITY

Decision rights ratified. Model registry and non-human identity inventory reconciled. One platform decision made.

Q2

COVERAGE

Continuous examination extended across control families and the vendor contract population. Coverage becomes the reported metric.

Q3

ENFORCEMENT

Circuit breakers, purpose enforcement, and the tool registry move from advisory into the execution path.

Q4

PROOF

Rehearsed suspension, verified erasure, concentration analysis, and an evidence pack that refreshes without a project.

Year one ends with installed capability, not a roadmap for year two.

DOMAIN 01

Governance & Risk Management

Decision rights ratified before the incident, not allocated after it.

WHAT THE PROGRAM INSTALLS

Decision Rights RACI

Who approves, who suspends, and who is accountable, per system class. Ratified by signature, not by committee minute.

Human Supervision Matrix™

Autonomy defined per action class with the evidence threshold that gates any expansion.

Shadow AI Sanctioning Pathway™

Unapproved use treated as a demand signal. Discovery, tiering, migration, disposition, and a published intake time.

Living Risk Register

Machine-maintained fields separated from human-adjudicated ones so entries stay current between committee cycles.

DOMAIN 02

Security Operations

Supervised autonomy, measured. Not detection timing carried over from a human-speed world.

WHAT THE PROGRAM INSTALLS

Agent Circuit Breaker Envelope

Twelve bounded dimensions including token spend, delegation depth, and loop detection. Enforced at the gateway, never in the agent prompt.

Red-Button Protocol™

Pre-authorized triggers, a named executor, a dependency map, and a rehearsal record with a measured time to stop.

Continuous Adversarial Evaluation

A standing test set run monthly against production so behavioral drift appears as a trend rather than an incident.

Non-Human Identity Pyramid™

Four tiers set by what an identity can do, driving credential lifetime, review cadence, and required controls.

DOMAIN 03

Third-Party & Supply Chain Risk

Vendor trust as a managed variable rather than a dated annual attestation.

WHAT THE PROGRAM INSTALLS

Adaptive AI Trust Scores™

Attested posture, monitored signals, contract strength, and incident history, weighted by the enterprise and carrying a direction.

AI Bill of Materials

Foundation model provider, version policy, inference routing, subprocessors, and unlearning capability, required at every renewal.

Concentration Analysis

Twelve vendors on one foundation model is one dependency. Quantified as a seventy-two hour outage figure the CFO can act on.

AI Clause Library

Required, prohibited, and monitored positions encoded once so procurement enforces rather than renegotiates.

DOMAIN 04

Data Protection & Privacy

The heaviest engineering load of the four. There is no way to soften that.

WHAT THE PROGRAM INSTALLS

AI Decision Provenance Chain™

What data went in, what processed it, what the decision was, and what could have altered it along the way.

Purpose Enforcement Map

Consent basis bound to a pipeline enforcement point that blocks rather than logs.

Erasure Engineering Register

Every location a request must reach: stores, backups, vector indexes, caches, agent memory, and model artifacts.

Memory Governance Standard

Retention, scoping, isolation, and a purge procedure for agent memory that has historically had no lifecycle at all.

THE NUMBER THE BOARD CAN FUND AGAINST

AI Security Debt™

Every program carries accepted risk. Adding an age dimension turns a static list into a trend the audit committee can read without being briefed.

Unregistered systems

In production, absent from the model registry.

Ungoverned identities

Capable of action with no named human owner.

Stale authority

A grant past expiry or past review, still active.

Unassessed connections

A tool or server reachable by an agent, not on the allowlist.

Undocumented versions

A production model version the enterprise cannot state.

Unratified rights

No named approver, suspender, or accountable executive.

A rising count with a rising burn-down rate is a program discovering faster than it closes. That is the expected year-one state, and it is the one to report.

WHAT THE BOARD RECEIVES

The AI Security Scorecard™

Assembled continuously, presented quarterly. The question is never what the number is. It is whether the number is moving in the right direction and what is being done about it.

SECTION	WHAT IT REPORTS	SOURCE
Population	AI systems in scope, and drift since last quarter	Model registry
Accountability	Systems with ratified decision rights, and the gaps	Decision Rights RACI
Supervision	Autonomy distribution and expansions with evidence	Supervision matrix
Containment	Rehearsal date and measured time to stop	Red-Button file
Supply chain	Trust score distribution and the concentration figure	Trust score model
Data protection	Erasure verification result and enforcement block rate	Erasure register
The measured number	One figure the program produced itself this quarter	Chosen deliberately

THE ENGAGEMENT

How the program is stood up

Five stages across the first four quarters, run against your calendar rather than a generic template.

01	Sequence	The dependency order is set against your environment. Four domains on one calendar, not four programs launched at once.
02	Ratify	Decision rights taken into the room with the business owners and signed per system class. The accountability question is settled first.
03	Install	The supervision matrix, the authority chain, the circuit breaker envelope, and the trust score model built and put into the execution path.
04	Rehearse	Suspension executed on a real system with a timed record. Erasure verified by sample. The findings are the deliverable.
05	Report	The scorecard assembled and presented to the audit committee with four quarters of trend structure already in place.

DELIVERABLES

What you operate afterward

Twenty-two working instruments, each with a named owner and a cadence. Not a report that ages.

01 **The Operating System Map**

Four domain programs held as one, with shared decision rights and one calendar.

02 **Human Supervision Matrix™**

Autonomy per action class with the evidence gate for every expansion.

03 **Red-Button Protocol™ File**

A rehearsal record with a measured time to confirmed stop.

04 **Agent Authority Chain™ Worksheet**

Applied at design for every agent in the top two authority tiers.

05 **Adaptive AI Trust Score™ Model**

Weighting, thresholds, and the named action at each threshold.

06 **AI Security Scorecard™**

The quarterly board instrument, with metric definitions and trend fields.

All twenty-two instruments ship as editable working files, not as figures in a book.

THE PLAN YOU LEAVE WITH

The first ninety days

Sequenced by dependency rather than preference. Four programs launched at once with one team and one budget become zero programs by month four.

DAY 1–30

Settle

The accountability conversation with General Counsel. Model registry and non-human identity inventory run in parallel. Everything downstream depends on both.

DAY 31–60

Ratify

Decision rights signed per system class with the business owners in the room. Exactly one platform decision, not four concurrent procurements.

DAY 61–90

Prove

The supervision matrix built. The kill switch rehearsed on a real system. One measured number that did not exist on day one.

The ninety-day close needs one number, not a status update. Activity is not evidence, and the people you report to can tell the difference.

TO SET EXPECTATIONS CLEARLY

What this program is not

This is the operating discipline that runs after the audit. It does not replace the audit, and it does not replace the security program.

✕ Not a repeat of the audit

Volume V produced the evidence base. This converts it into a running program. If the audit has not been run, the first chapter of each Part names what has to exist before the work can start.

✕ Not a penetration test

No production systems are attacked. Continuous Adversarial Evaluation is a standing test set against AI behavior, and red-team engagements are scoped separately.

✕ Not a compliance certification

The artifacts a certification asks for are produced here. The certifying body is a separate party and relies on evidence of exactly this kind.

✕ Not a vendor recommendation

No products are named anywhere in the program. No tools are sold and no referral fees are taken from any AI vendor in the environment.

BUILT FOR

Who this program is for

Written for the practitioner who owns the work, not for the board and not for a general business reader.

SPONSORS

Chief Information Security Officer

Security director

Security manager who owns the work

GRC and third-party risk directors

SOC and detection engineering leads

Sized for mid-market through large multinational organizations.

PRECONDITION

AI is already running in the environment, whether or not anyone approved it.

Every page assumes the reader knows what a control is, what an identity is, and what happens when neither one is inventoried. If the organization is still deciding whether to permit AI, this is the wrong instrument and Volume I is the right one.

ABOUT THE AUTHOR

Stephen R. Jordan

Founder, SRJ Consulting & Services LLC

Author of The AI IT Security Implementation & Strategy™, the sixth volume of The Operating Discipline for AI Library™ and the second 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.

He is also the publisher of theworldofai.org, an independent AI reference atlas rebuilt daily from primary sources, tracking state AI legislation, AI litigation, compliance frameworks, and peer-reviewed research.

BACKGROUND

Citi

Intel

McAfee

Optiv

Based in Dallas–Fort Worth, working with executives nationwide.

Ready to convert the audit into a program that produces evidence?

Book a conversation with Stephen R. Jordan.

srjconsultingservices.com

SRJ Consulting & Services LLC

Dallas–Fort Worth · Nationwide