InfoSecAI

A BRIEF FOR CISOS · BISOS · GRC LEADERSHIP

WHITE PAPER · PAPER 1 · 2026

AI Governance Is No Longer a Policy Problem

A practitioner's brief for CISOs, CIOs, CTOs, AI transformation leaders and board sponsors building governance that proves what is happening, not just what should happen.

AUTHORED BY

Paul Jolliffe

By Paul Jolliffe, Founder and Director, InfoSecAI Limited · MBA · CISSP · ISO 27001 Lead Auditor

01 · EXECUTIVE SUMMARY

Executive summary

Most organisations now have an artificial intelligence (AI) policy, an AI acceptable-use statement and, in many cases, an AI steering committee. Very few can answer the question that follows: what evidence proves the policy is operating?

The gap between AI policy artefacts and AI operational evidence is widening. Regulators, customers, internal audit functions and boards are no longer satisfied with the existence of governance documents. They want to see a functioning operating model: a current inventory of AI systems, classifications tied to risk and autonomy, controls mapped to specific evidence, named owners, and a reporting cadence that triggers decisions rather than just providing updates.

This brief is the first in InfoSecAI's five-part executive series, From AI Ambition to AI Assurance. It sets out why AI governance is now an operating-model problem, what a practical operating model looks like, where information security must be integrated, and how a chief information security officer (CISO) or chief information officer (CIO) can build the first thirty days of evidence-led AI governance without re-engineering the whole organisation.

The argument is direct. A policy tells people what should happen. An operating model proves what is happening.

02 · WHY THIS MATTERS NOW

Why this matters now

Three forces have converged in the last twelve months.

The first is regulatory pressure with teeth. The European Union Artificial Intelligence Act (EU AI Act, Regulation (EU) 2024/1689) entered into force on 1 August 2024. The prohibited-practices article and the AI literacy obligations applied from 2 February 2025. The general-purpose AI (GPAI) model obligations applied from 2 August 2025. The bulk of the high-risk system rules and the provider and deployer obligations apply from 2 August 2026.

The second is the maturity of voluntary standards. ISO/IEC 42001:2023, the world's first AI management system standard, was published in December 2023. The United States National Institute of Standards and Technology (NIST) published the AI Risk Management Framework (AI RMF 1.0) in January 2023 and the Generative AI Profile (NIST AI 600-1) in July 2024. Together these give a CISO a coherent management-system shape to point at when designing controls.

The third is operational reality. AI is no longer confined to the data science team. Generative AI features are embedded in everyday software-as-a-service (SaaS) tools, copilots are deployed across knowledge work, and agentic systems are moving into roles

that touch customer data, financial transactions and operational decisions. The attack surface is broader, the change cadence is faster, and the audit trail is thinner than for any prior technology category.

The organisations in the best position will not be the ones with the longest AI policies. They will be the ones with the shortest distance between policy intent and operational evidence.

03 · WHY CURRENT APPROACHES ARE FAILING

Why current approaches are failing

Three failure patterns recur, each rational in isolation, all three together produce AI governance theatre.

The first is policy-led governance. The organisation drafts an AI policy, circulates it for sign-off and publishes it on the intranet. The policy is competent. Nobody can demonstrate which AI systems it applies to, who has read it, or what changes it has driven. Policy without an operating model is not governance; it is a wish list with letterhead.

The second is committee-led governance. An AI steering committee is established. It meets monthly, reviews use cases brought forward by business teams, and minutes its discussions. The committee does not maintain an AI inventory, does not see shadow AI adoption, does not own a control catalogue, and does not produce assurance evidence between meetings. A committee without an operating model is a forum, not a control.

The third is project-led governance. The organisation runs an AI Act readiness project or an internal audit of AI controls. The project produces a gap analysis, a roadmap and a steering report. Six months later the gaps have moved, the inventory has drifted, and the roadmap has slipped behind organisational AI adoption. A one-off project cannot govern a moving estate.

The common cause is structural. AI is dynamic; policy is static. AI is distributed; committees are centralised. AI is continuous; projects have end dates. The fix is to move from artefact production to management-system operation.

04 · THE FRAGMENTED-OWNERSHIP PROBLEM

The fragmented-ownership problem

In most organisations AI ownership is fragmented across at least six functions. Legal and privacy own the regulatory interpretation. Information security owns the data and identity controls. Risk owns the enterprise risk register. Data and analytics own the models. Technology owns the infrastructure. The business owns the use case. Each function holds a partial view and assumes another holds the rest.

The result is predictable. When the board asks who owns AI risk, six functions can each give a partial answer. None can give the complete one. When internal audit requests evidence of control operation, the request bounces between functions until it is partially answered or quietly closed.

The fix is not to centralise AI ownership into a single function. The fix is to make ownership explicit at the control level rather than at the system level, supported by a responsible, accountable, consulted, informed (RACI) model that names individuals against each control activity. Ownership at the control level scales; ownership at the system level does not, because new AI systems appear every week.

05 · THE AI GOVERNANCE OPERATING MODEL

The AI governance operating model

The model below draws on ISO/IEC 42001:2023, NIST AI RMF 1.0 and the EU AI Act's structural expectations. It is not a certification scheme. It is a practical control-and-evidence system that can be operated continuously and surfaced into board reporting.

The model has seven activities. Each activity produces one canonical evidence artefact, named, owned and reviewed on a defined cadence.

Inventory. Maintain a current register of AI systems in use, including supplier-embedded AI features, internally developed models, agentic systems, and material AI capabilities procured as part of broader software contracts. Each entry carries an owner, a business purpose, a data-classification tag, an AI Act risk-tier preliminary classification and an autonomy-and-access classification.

Classify. Apply a single classification per system that resolves four questions in one pass: AI Act risk tier, operational criticality, data sensitivity and agent autonomy class (observe, advise, act with approval, act autonomously). Refreshed quarterly.

Control. Map each classification to a defined control set drawn from ISO/IEC 42001 Annex A, NIST AI RMF Govern-Map-Measure-Manage functions, and the organisation's existing information security management system (ISMS).

Evidence. For each control, define the operational artefacts that prove it is operating: access logs, training-data provenance records, model outputs, monitoring alerts, change records, human-oversight decisions. Refreshed monthly.

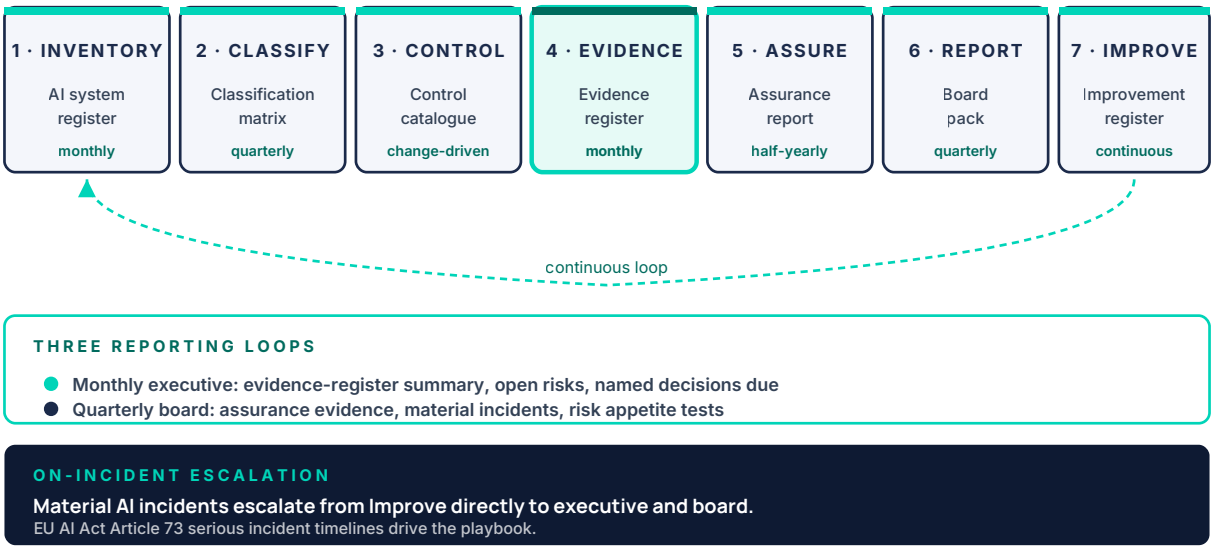
Assure. Run a periodic independent check on a sampled subset of controls. Internal first, external where the regulatory context requires it.

Report. Produce one quarterly board-ready report and one monthly executive report, both drawn from the same evidence register. The board report focuses on assurance evidence, incidents, and decisions required.

Improve. Operate a defined corrective and preventive action workflow against incidents, assurance findings and external changes.

FIGURE 1 · The AI governance operating model.

Seven activities. Seven canonical artefacts. Three reporting loops.



The discipline is the one-artefact-per-activity rule. Most organisations producing AI governance theatre have many overlapping artefacts nobody can reconcile. The operating model treats each artefact as a single source of truth and every other document as a derivative view.

06 · THE MATURITY LADDER

The maturity ladder

The ladder is a diagnostic, not a target. It is built around what operational evidence the organisation can produce at each rung.

Level 1, declared. The organisation has an AI policy. It cannot produce an AI inventory.

Level 2, scoped. The organisation has an AI inventory and a documented classification scheme. Controls are described but evidence is not consistently produced.

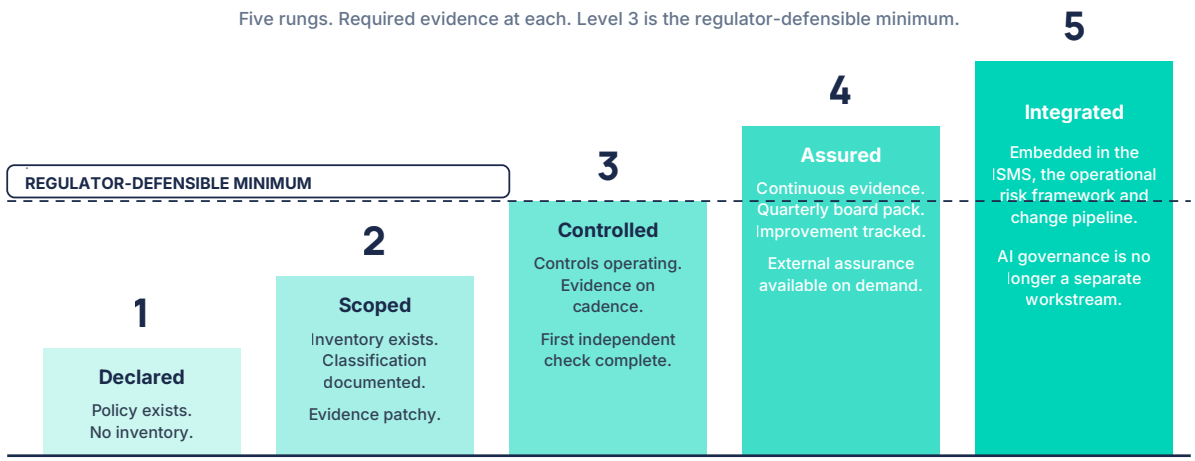
Level 3, controlled. Controls are operating against classified systems. Evidence is produced on a defined cadence. The first independent assurance check has been completed.

Level 4, assured. Continuous evidence production. Quarterly board reporting against the operating model. Improvement actions tracked to closure.

Level 5, integrated. The AI operating model is integrated with the broader ISMS, the operational risk framework and the business change pipeline. AI governance is no longer a separate workstream.

FIGURE 2 · The policy-to-operating-model maturity ladder.

Five rungs. Required evidence at each. Level 3 is the regulator-defensible minimum.



Most mid-sized organisations sit at Level 1 or 2 today.
Level 3 is the regulator-defensible minimum for high-risk AI under the EU AI Act from 2 August 2026.

A practitioner-grade diagnostic moves through five questions per level: can you produce the artefact, is it current, is it owned, is it reviewed, and does it drive a decision? Any "no" means the level is not held.

07 · THE USE-CASE CLASSIFICATION MATRIX

The use-case classification matrix

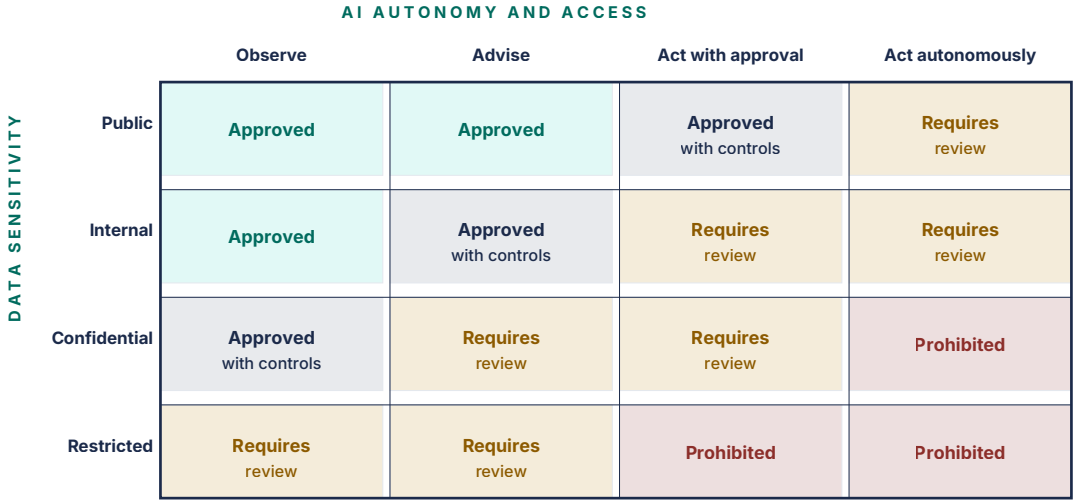
The matrix replaces the open-ended question "is this AI use case OK" with a structured one that gives a defensible answer in minutes.

Two axes. Data sensitivity on one axis, scored against the existing data classification policy. AI autonomy and access on the other, ranging from purely observational through advisory, act-with-approval, and act-autonomously.

Every AI use case lands in one cell. Each cell carries a default disposition: approved, approved-with-controls, requires-review, prohibited. The default is the starting point; it can be overridden by the named control owner with documented justification.

FIGURE 3 · The AI use-case classification matrix.

Default dispositions. Overrides require named owner and documented justification.



The practical benefit is decision velocity. A workflow that previously waited four weeks for a committee slot produces a defensible classification in one working day, with the committee reserved for genuinely ambiguous cases.

Worked example. Hartfield Asset Management plc, a mid-sized UK asset manager, deployed the matrix as an intake control in March 2026. The first quarter produced 47 AI use cases. Of those, 28 landed approved or approved-with-controls and went live within ten working days; 14 landed requires-review and went through structured assessment; five landed prohibited and were redirected to non-AI alternatives. The AI committee now reviews only the 14 ambiguous cases.

08 · THE AI OWNERSHIP RACI

The AI ownership RACI

Control-level ownership closes the fragmented-ownership problem. Figure 4 below is a starting template overlaid directly onto the operating-model activities; tune it to existing accountability lines. Two design rules matter. The accountable role stays stable across the activities (CISO in most cases, with the Chief Risk Officer accountable for classification and the Chief Audit Executive accountable for internal assurance). The responsible role rotates by activity, because the operational work shifts as the artefacts shift.

FIGURE 4 · AI ownership RACI overlaid on the operating model.

Accountable role is stable. Responsible role rotates by activity.

09 · INFORMATION SECURITY IMPLICATIONS

Information security implications

A common failure mode is to treat AI governance and information security as parallel workstreams. They are not. AI introduces specific risks the existing ISMS already has most of the controls for, provided integration is deliberate.

Six integration points matter most. Access control to AI systems and the data they consume extends naturally from existing identity and access management (IAM) if the AI inventory feeds the access-review cycle. Data leakage through prompts, outputs and embeddings is covered by existing data-loss-prevention controls if the AI use cases are inventoried; shadow AI bypasses sanctioned channels and is the subject of Paper 2 in this series. Prompt injection and adversarial inputs map onto the Open Worldwide Application Security Project (OWASP) Top 10 for Large Language Model Applications (2025 edition); controls include input validation, output filtering, sandboxed execution and least-privilege tool access. Model and prompt provenance is contractual for procured AI (training-data representations, supplier improvement-use clauses) and documentary for internal AI (training data lineage, model cards, evaluation records). Logging and audit trail under EU AI Act Article 12 require high-risk system logs to be retained, with a pragmatic minimum of six months. Incident response and serious incident reporting under EU AI Act Article 73 add tighter timelines than most general SOPs; the integration point is the incident-classification step, which needs an explicit AI-system branch.

The discipline is to refuse to build a parallel AI security function. AI security is information security with new attack patterns and new evidence requirements. The CISO who treats it as parallel doubles the cost and halves the coherence.

10 · EXECUTIVE DECISION POINTS AND THE 30-DAY BASELINE

Executive decision points and the 30-day baseline

Six decisions are usually pending in the organisations that ask InfoSecAI for help.

Who is the accountable executive for AI risk? What is the risk appetite for AI autonomy, specifically when an AI system is permitted to act without human approval? What is the approval path for new AI use cases? What evidence is required before a use case goes live? What goes to the board and how often? When is external assurance triggered?

The first thirty days of an AI governance operating model do not require a tooling investment, a certification programme or a new committee. They require five concrete artefacts.

Week one: the first version of the AI system register. Sources: existing IT asset register, SaaS contract list, procurement records, supplier AI feature disclosures, business AI use-case requests. Target completeness is 80 per cent of material systems.

Week two: run the classification matrix against the register. Default dispositions only; overrides come later. Output is a classified register with red, amber and green flags.

Week three: map the existing information security control set against the classified register. Identify which controls already cover AI, which need extension, which are missing.

Week four: name the owners. RACI in draft against the operating-model activities. Walk the RACI with each accountable role to confirm they accept the accountability or escalate.

By the end of week four, the organisation has an inventory, a classification, a control map and an ownership model. Level 3, controlled, follows as the operating-model activities are run; Level 4, assured, follows after the first independent check.

11 · QUESTIONS EVERY LEADER SHOULD ASK NOW

Questions every leader should ask now

Can you produce the current AI system register in under one hour? If not, the inventory activity is not operating.

For the three highest-risk AI systems, can you name the owner, the classification, the control set and the most recent evidence of control operation? If not, the control and evidence activities are not operating.

What changed in the AI estate in the last thirty days? If the answer is "I would need to check," improvement is not operating.

When did the board last see AI assurance evidence rather than AI strategy slides? If "never" or "more than six months ago," reporting is not operating.

If internal audit began an AI controls audit on Monday morning, which control would the organisation be least able to evidence? The honest answer is the priority action.

12 · CLOSING THOUGHT

Closing thought

AI governance is no longer a policy problem. It is an operating-model problem. The shift from policy artefacts to operational evidence is the single most important change executive teams need to make in 2026.

The distinctive thread across this series is straightforward: AI assurance evidence, not AI reassurance narrative. A policy tells people what should happen; an operating model proves what is happening. The organisations that build the operating model first will scale AI safely. The organisations that keep producing artefacts will spend the next two years explaining why their controls did not catch what their policies said would never happen.

The next paper, The Shadow AI Exposure Map, examines what to do when the AI estate turns out to be larger than the inventory.

13 · SOURCE REGISTER

Source register

All sources verified to primary publisher on 1 June 2026.

#	SOURCE	USE IN PAPER	LINK
1	Regulation (EU) 2024/1689 (the EU AI Act)	Risk tier framework, GPAI obligations, applicability timeline, Article 12 logging, Article 73 serious incident reporting	https://eur-lex.europa.eu/eli/reg/2024/1689
2	European Commission, AI Act regulatory framework	Implementation timeline references and role determination	https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
3	ISO/IEC 42001:2023, AI Management System Standard	Management-system structure, Annex A control families	https://www.iso.org/standard/42001
4	NIST AI Risk Management Framework (AI RMF 1.0), January 2023	Govern, Map, Measure, Manage functions	https://www.nist.gov/itl/ai-risk-management-framework
5	NIST AI 600-1, Generative AI Profile, July 2024	Generative AI risk taxonomy and recommended actions	https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

#	SOURCE	USE IN PAPER	LINK
6	NCSC, Guidelines for secure AI system development	Secure design, development, deployment and operation framing	https://www.ncsc.gov.uk/collection/guidelines-secure-ai-system-development
7	OWASP Top 10 for Large Language Model Applications, 2025	Prompt injection, sensitive information disclosure, excessive agency	https://genai.owasp.org/llm-top-10/

14 · ABOUT THIS SERIES

About this series

From AI Ambition to AI Assurance is a five-paper executive briefing series, 1 to 5 June 2026.

1. AI Governance Is No Longer a Policy Problem (this paper)
2. The Shadow AI Exposure Map
3. Securing Agentic AI Before It Acts
4. Why AI Transformation Fails After the Pilot
5. The Board Pack for AI Assurance

Each paper is published as a 12-page executive briefing under the InfoSecAI Blog Template. The full series is available at infosecai.net/insights for subscribers to the InfoSecAI insights list.

15 · PRACTITIONER NOTE

Practitioner note

This briefing is practitioner interpretation, not legal advice. For regulated deployments, validate final claims against current legal obligations, sector-specific requirements and the original primary sources before relying on them.

InfoSecAI

About InfoSecAI

InfoSecAI is an independent UK consultancy helping organisations turn security, regulatory, resilience and AI governance requirements into practical operating models, stronger controls and robust delivery.

We work across strategy, governance, risk, compliance, AI security, assurance, operations and engineering. Our services help leadership teams assess their current position, align to standards and regulation, define the target operating model, and deliver the governance, controls, artefacts and ways of working needed to move from intent to implementation.

Our toolkit capability accelerates structured work across ISO 27001, ISO 22301, ISO 42001, NIST CSF, NIST AI RMF, CIS Controls, Cyber Essentials, DORA, NIS 2, the EU AI Act, GDPR, UK GDPR, SOC 1 and SOC 2. The approach combines AI-enabled workflow support with senior practitioner judgement, so outputs remain proportionate, usable and connected to the way the organisation actually operates.

InfoSecAI was founded in **2025** by **Paul Jolliffe**. The company is built for organisations that need clarity, senior leadership and hands-on delivery across information security and AI governance, without adding unnecessary complexity or treating compliance as a paperwork exercise.

infosecai.net · paul.jolliffe@infosecai.net