

MANDATE *Suite*

Governance Architecture

42

EXECUTIVE OVERVIEW

Purpose-Built AI Governance

for Regulated Industries

Most AI governance frameworks answer one question: Are our AI systems compliant? The MANDATE Suite answers three. It is the only AI governance framework that governs the full authority lifecycle — from the moment the Board delegates machine authority to its ongoing obligation to ensure that authority remains admissible under current conditions.

Built specifically for regulated enterprises in financial services, insurance, healthcare, and government, the MANDATE Suite provides the constitutional architecture, normative instruments, and Board governance machinery that compliance standards alone cannot provide. It does not replace ISO/IEC 42001 or the EU AI Act. It makes them operational at the depth that regulated institutions require.

67	5	13	27
NORMATIVE INSTRUMENTS	COMPONENTS	PRODUCTS	BOARD KRIS

> *The governing distinction: Compliance answers whether a system should operate. Governance answers whether it remains in control — and whether its authority still has the right to bind consequence today.*

The MANDATE Suite is not a policy template. It is a complete governance architecture — constitutionally structured, Board-adopted, and operationally enforced. Every instrument traces to a Board resolution. Every control has a measurable threshold. Every delegation of machine authority has a ceiling.

THE PROBLEM

Three Questions.

Most Boards Ask One.

The board asks whether its AI systems are compliant. That question has an answer — ISO/IEC 42001, the EU AI Act, and APRA's prudential standards address it. What they do not address is the governance architecture that must exist beneath compliance for regulated enterprises operating autonomous AI systems at scale.

QUESTION ONE — ANSWERED

Are our AI systems compliant?

Addressed by ISO/IEC 42001 · EU AI Act · APRA CPS 220

QUESTION TWO — RARELY ASKED

How much machine authority have we delegated in total — and who has the power to stop it growing?

No aggregate measure exists in most organisations · No Board-approved ceiling · No named accountable executive

QUESTION THREE — NEVER ASKED

Does the authority we have delegated retain the right to bind consequence now — not merely when we originally granted it?

No governance framework addresses this · Conditions change · Exposure drifts · MANDATE does

> *An organisation can be technically compliant and still be structurally exposed. Compliance answers the past. Governance answers the present.*

The MANDATE Suite was built to answer all three questions — constitutionally, operationally, and continuously. It governs delegation as a live condition, not a historical record.

The Strategy-to-Execution Gap.

Where Governance Intent Evaporates.

Most regulated enterprises have four things: a Board resolution on AI governance, a CAIO-authored framework, a technology architecture built by the CTO, and AI systems operating in production. What they lack is a structurally enforced chain connecting all four. That gap — between governance intent and execution reality — is where machine authority operates without constitutional constraint.

The failure mode is precise. The Board adopts a risk appetite and sets a ceiling on machine authority. The CAIO authors a governance framework that reflects that appetite. The CTO builds systems that reference the framework as a design input. But the framework is a document. The resolution is a resolution. Neither has technical enforcement reach at the execution boundary. The result is a governance chain that looks coherent on paper but is structurally unenforced in production.

BOARD

Adopts risk appetite. Sets Autonomy Budget ceiling. Passes six constitutional resolutions.

Intent is constitutional. Enforcement reach is zero.

CAIO

Authors the governance framework. Defines delegation standards. Issues policies and standards.

Instruments are normative. Enforcement depends on others implementing them.

CTO → EXECUTION

Builds the systems. Deploys agents. Manages infrastructure. Governs the execution boundary.

Implementation is technical. Policy compliance is not structurally verifiable.

This is not a failure of intent. It is a failure of architecture. Three specific gaps define it:

First, no instrument structurally enforces authority limits at the execution boundary, independent of the systems being governed. Policy compliance is self-certified by the same delivery function that has commercial incentives to deploy.

Second, no mechanism continuously tests whether delegated machine authority remains admissible under current conditions—not merely whether it was correctly granted at the time of initial authorisation.

Third, no portfolio-level instrument aggregates machine authority across the estate against a Board-approved ceiling. Authority expands system by system; the Board has no view of the total.

> *The question MANDATE answers is not whether governance intent exists. It is whether governance intent is structurally enforced — at the execution boundary, continuously, at the portfolio level, and without management's ability to bypass it unilaterally.*

One Architecture.

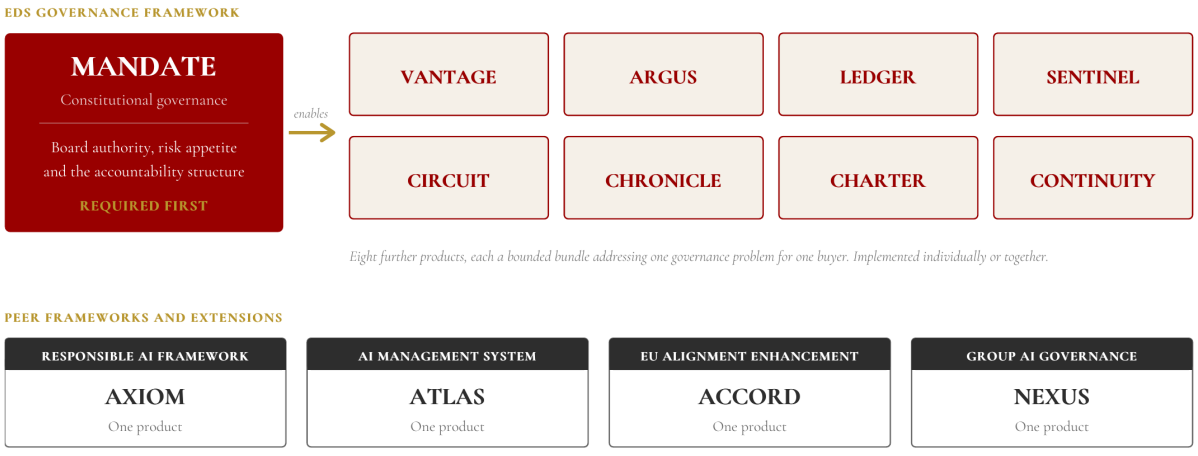
Five Components.

The MANDATE Suite has five components: two peer governance frameworks, a management system, a compliance enhancement and a structural extension. Neither governance framework overrides the other. All 67 instruments trace to a single constitutional hierarchy — the Governance Doctrine adopted by the Board.

Two governance frameworks are required because governing autonomous AI involves two distinct but inseparable obligations. MANDATE governs the delegation of machine decision authority — how much authority exists, who granted it, under what conditions it operates, and whether those conditions remain admissible. AXIOM governs the ethical and rights boundaries on what that authority may be used for — fairness thresholds that are measurable and binding, individual rights obligations that cannot be contracted away, and prohibited uses that no commercial pressure may override. ATLAS is not a third governance framework. It is the management system that keeps both operationally maintained — the competence programme, internal audit cycle, continual improvement obligation, and ISO/IEC 42001:2023 certification pathway. MANDATE and AXIOM are peers: neither is subordinate, and both must be satisfied concurrently for every AI system in scope. AIMS has no constitutional instrument of its own; ISO/IEC 42001:2023 sits above it.

The Product Map — What Has to Come First

A product is a bounded bundle of instruments solving one governance problem for one buyer.



Thirteen products across five frameworks. MANDATE establishes the Board authority from which every other product in its framework derives its mandate.

Component 01 · 32 Instruments

EDS Governance Framework — MANDATE

Decision authority governance · Autonomy Budget · Board Control Loop

Governs the delegation of machine decision authority from Board to system — establishing the constitutional ceiling on aggregate autonomous authority, the per-system authority limits, the five-level autonomy classification, and the eight-stage Board governance control loop that keeps machine authority aligned with Board intent at all times.

Component 02 · 14 Instruments

Responsible AI Framework — AXIOM

Fairness · Data provenance · Individual rights · Prohibited use

Governs the ethical and rights dimensions of AI deployment — establishing fairness assessment obligations, data provenance requirements, individual rights response processes, and the prohibited use boundaries that no commercial pressure may override. Operates as a constitutional peer to the MANDATE.

Component 03 · 12 Instruments

AI Management System — ATLAS

ISO/IEC 42001:2023 certification pathway · Management system discipline

Provides the ISO/IEC 42001:2023 management system envelope — delivering the competence programme, impact assessment framework, internal audit procedure, and continual improvement cycle that certification bodies require. ATLAS bridges MANDATE's constitutional governance and the international standard.

Compliance Enhancement · 6 Instruments

EU Alignment Enhancement — ACCORD

EU AI Act deployer compliance · FRIA · Article 73 serious incident notification

Extends the core suite for EU AI Act deployers — the Fundamental Rights Impact Assessment, high-risk classification, provider technical documentation, Article 50 transparency and serious incident notification, aligned with Regulation (EU) 2026/1744, in force since 27 July 2026.

Structural Extension · 3 Instruments

Group AI Governance Framework — NEXUS

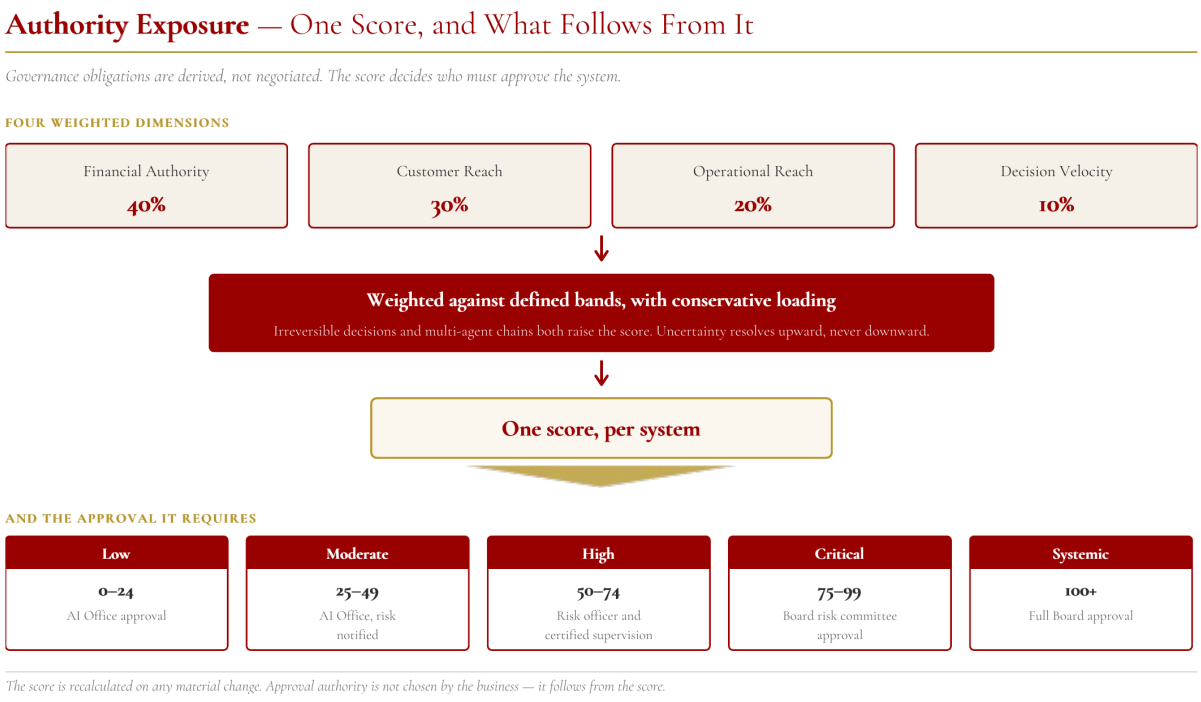
Holding company structure · Group Portfolio ADAE · Cross-entity authority governance

Extends the core suite for group structures — the Group Autonomy Budget ceiling, subsidiary sub-budgets, mandatory cross-entity loading, and the Group CAIO role. Individual entity compliance is not group governance.

The Autonomy Budget

and the Governance Maturity Index

The Autonomy Budget and the Governance Maturity Index are the two sides of the same constitutional constraint. The Autonomy Budget governs quantity — how much delegated machine authority the organisation holds in aggregate, measured continuously against a Board-approved ceiling. The Governance Maturity Index governs capability — whether the organisation's governance infrastructure is certified to support the autonomy level it operates at. Neither operates without the other: machine authority cannot exceed the Board-set ceiling, and it cannot increase in level without corresponding GMI certification. Together, they make autonomy expansion a Board-governed act, not a management decision.



The Autonomy Budget

The Autonomy Budget is the Board-approved aggregate ceiling on total delegated machine authority across the entire AI estate. It converts the abstract risk appetite statement into a single, quantified, continuously monitored exposure limit.

Every AI system in the organisation is scored using the Autonomous Decision Authority Exposure (ADAE) model — four weighted dimensions: financial authority, customer reach, operational reach, and decision velocity — producing a score that reflects the ceiling of what that system can cause, not the floor of what it typically does.

When the portfolio's aggregate ADAE utilisation approaches the Board-set ceiling, governance responds automatically. Only a Full Board resolution can raise the ceiling. Management cannot expand machine authority beyond what the Board has sanctioned.



Below 80% — Normal operations

Quarterly monitoring. No escalation required.

80–89% — Board review required

No new deployment without BRC sign-off.

90–99% — Deployment moratorium

No further deployment until utilisation falls below 90

100% — Full Board resolution required

Raise ceiling by resolution or decommission to return

The Governance Maturity Index

Autonomy may expand only where governance capability demonstrably supports it. The GMI is the Board's constitutional instrument for enforcing this principle.

The GMI certifies the organisation's governance capability across five levels. No system may operate at an Autonomy Level exceeding the organisation's certified GMI level — under any circumstances. This constraint cannot be waived by commercial pressure, delivery timeline, or operational necessity.

GMI 1 — AI estate registered; CAIO appointed; Safety Runtime deployed

GMI 2 — HOTL framework operational; supervisor certification active

GMI 3 — AI QA function independent; kill-switch tested in production

GMI 4 — Safety Runtime fully deployed; zero orphaned agent identities

GMI 5 — End-to-end traceability confirmed; external assurance

The ceiling is Board-set. The CRO certifies Levels 1–2, the Board Risk Committee Levels 3–4, and the Full Board Level 5 on external assurance. Neither may be circumvented by management action alone.

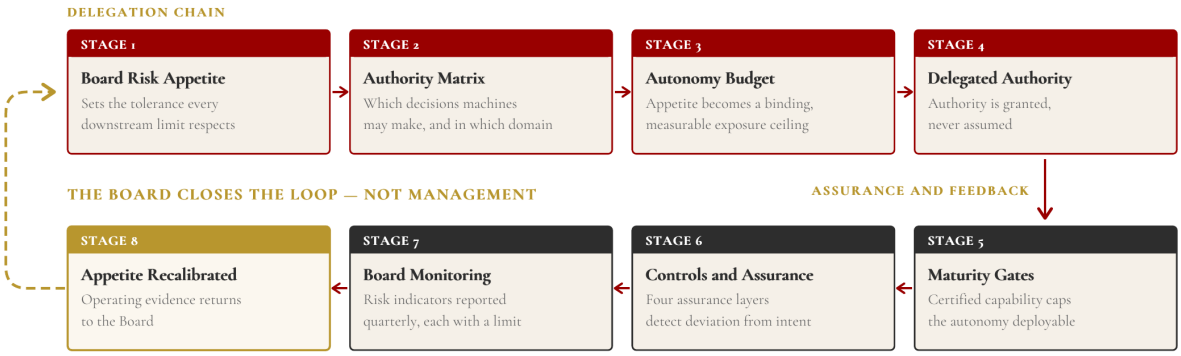
Governance starts at the Board.

Not at the system.

The MANDATE Suite begins where most frameworks end — at the Board. Six constitutional resolutions establish the governance mandate. The Board sets the Autonomy Budget ceiling. The Board certifies the Governance Maturity Index. The Board receives 27 Key Risk Indicators quarterly — 22 from the EDS Governance Framework and five from the Responsible AI Framework. Every instrument traces to a Board resolution.

Board Governance Control System — Eight-Stage Continuous Loop

Machine autonomy expands only when governance maturity demonstrably increases and operations remain within Board-approved risk appetite



The loop tightens under stress. A breached threshold accelerates the feedback pathway and brings an unscheduled report to the Board. Every expansion of machine authority re-enters the loop at Stage 1.

> *The eight-stage Board Governance Control Loop governs how the Board sets risk appetite, delegates machine authority within certified limits, monitors utilisation against the Autonomy Budget, receives evidence through the quarterly dashboard, and closes the loop by adjusting appetite and authority limits based on operational evidence. The Board is the primary governance actor — not an implied authority.*

Six Board Resolutions

Constitutional mandate adopted at Full Board level. Risk posture and Governance-Constrained Autonomy; the Governance and Quality Assurance Framework; the funding envelope and CAIO accountability; the Autonomy Budget; the Machine Decision Authority Matrix; and Board monitoring and reporting — all Board-resolved, none delegable to management.

27 Key Risk Indicators

Quarterly Board Monitoring Dashboard: 22 Key Risk Indicators across 13 governance domains, from Autonomy Budget utilisation to HOTL certification coverage, and four Responsible AI indicators reported alongside them. Every KRI has a defined threshold, a red, amber, green status, and a mandatory escalation pathway.

Constitutional Hard Ceiling

The Autonomy Budget ceiling cannot be waived by commercial pressure, delivery timeline, or operational necessity. Only a Full Board resolution can raise it. The governance framework is constitutionally enforced — not policy-enforced.

The eight stages of the Board Governance Control Loop are: (1) Board Risk Appetite — (2) Machine Decision Authority Matrix — (3) Autonomy Budget — (4) Delegated Machine Authority — (5) GMI Gates — (6) Operational Controls and Assurance — (7) Board Monitoring Dashboard — (8↺) Feedback to Risk Appetite. Stage 8 is not a reporting obligation. The Board closes the loop — not management. Human-on-the-Loop (HOTL) supervisors, referenced in the six resolutions, are certified individuals with real-time monitoring and intervention authority over agentic systems at Level 3 and above.

A governance framework without a Board-level constitutional architecture is a policy document. MANDATE is not a policy document. It is a constitutional governance system — adopted by the Board, enforced by the infrastructure, measured by the dashboard, and closed by the Board every quarter.

THE DISTINGUISHING CONCEPT

Continuous Admissibility.

The Governance Dimension No Standard Reaches.

Every AI governance framework governs the moment of deployment. Systems are authorised, assessed, and approved before they go live. What happens between deployment and the next review cycle is where governance frameworks stop — and where MANDATE begins.

Historical Validity

Was this authority correctly granted? The authorisation is on record. The governance process was followed. The documentation is complete. The system was approved on the day it went live.

Answered by every governance framework. Answers the past.

Current Admissibility

Does this authority retain the right to bind consequence now? Operating volumes may have shifted. Financial exposure may have drifted. Organisational context may have evolved. A delegation that was correct on the day it was granted may be inadmissible today.

Answered only by MANDATE. Governs the present.

The MANDATE Suite governs admissibility through four mechanisms operating between scheduled review cycles:

Continuous Admissibility Monitor

Detects live drift across seven dimensions — volume, customer reach, financial exposure, behaviour, retrieval corpus currency, external fact currency and output residency. Raises a state-progression alert when operating parameters diverge from the governance profile under which authority was granted.

ADAE Reassessment Obligation

A confirmed drift alert triggers mandatory ADAE reassessment within 10 business days. If the system's profile has changed materially, reauthorisation at the current band is required before it continues operating.

Structural Change Trigger

Any material change to organisational structure triggers an admissibility review of every affected system. An authorisation granted for one organisational context does not survive structural discontinuity without active reconfirmation.

> *The audit trail answers: what was authorised? The Continuous Admissibility Monitor answers: what remains admissible now? These are different questions. MANDATE answers both.*

REGULATORY POSITIONING

Built for the Regulatory Environment

You Operate In.

The MANDATE Suite does not create a parallel compliance programme. It creates the governance architecture that makes compliance obligations operational — measurable, evidenced, and Board-accountable. Every instrument maps to its regulatory basis. Every control is traceable to a prudential standard.

Regulatory Obligation	What it requires	What MANDATE Suite provides
APRA CPS 220 / CPS 230	Board risk governance · AI as operational risk · Accountable executive	**MANDATE — Six Board resolutions · CAIO Charter · Constitutional Governance Doctrine · Draft CPS 510 anticipated**
ISO/IEC 42001:2023	AI management system · Competence · Internal audit · Continual improvement	**ATLAS — full ISO/IEC 42001 certification pathway · 12 AIMS instruments**
EU AI Act (Deployer)	Fundamental rights impact assessment · Transparency duties · Serious incident notification · High-risk registration	**ACCORD — 6 instruments · Article 50 in force now · high-risk duties from 2 December 2027**
APRA CPS 220 / 230 / 510 · Corporations Act s181, s187	Group-level AI governance · Cross-entity machine authority · Intra-group conflicts of interest	**NEXUS — 3 instruments · Group Autonomy Budget · Concurrent Board sanction**
OAIC Privacy Act · ASIC RG 271	Automated decision transparency · Individual rights · Explanation obligations	**AXIOM — Individual Rights Framework · Transparency Standard · Fairness Assessment**

> *The MANDATE Suite is designed to absorb regulatory extensions as obligations mature — without requiring a framework rebuild. When Australian obligations crystallise, EUAE instruments are absorbed into core MANDATE instruments. The architecture anticipates regulatory evolution.*

What MANDATE Provides

That Standards Cannot.

Implementation Depth

ISO/IEC 42001 requires an AI risk assessment. It does not specify a methodology. MANDATE specifies the ADAE scoring formula, the four weighted dimensions, eight conservative loading rules, the five exposure bands, and the governance actions required at each band. The difference between a certifiable governance system and a governance aspiration is implementation depth.

Board-Level Architecture

Most frameworks imply Board involvement as the source of mandate. MANDATE starts at the Board. Six constitutional resolutions. A Board-adopted Governance Doctrine. A Board Monitoring Dashboard with 27 defined KRIs. A Board-set Autonomy Budget ceiling that only a Full Board resolution can raise.

Quantified Authority

Machine authority is quantified, not estimated. The ADAE model produces a precise score for every system, a portfolio aggregate for the Board, and a utilisation percentage against the Board-set ceiling. An organisation implementing MANDATE knows its exact machine authority exposure — as a number, every quarter.

Continuous Governance

The Continuous Admissibility Monitor and ADAE reassessment obligation govern what happens between deployment and the next review cycle. Delegation is governed as a live condition. Authority that was valid on the day it was granted is continuously tested for admissibility today.

MANDATE provides

67 normative instruments specifying exactly what to do, in what sequence, with what approval, measured against what threshold

6 Board resolutions constitutionally adopted — not implied

27 Board KRIs with defined thresholds, data sources, and escalation triggers

5 autonomy levels, with certified human oversight mandatory at Level 3 and above

2 assurance architectures — the Three Lines Model with External Assurance, and the Four-Layer Assurance Model of Safety Runtime, human oversight, AI QA Function and Board dashboard

11 gate conditions that must all pass before any Level 3 system enters production — partial compliance is not sufficient

1 published book setting out the architecture — Apress, 2026

16 peer-reviewable papers forming an open-access research record

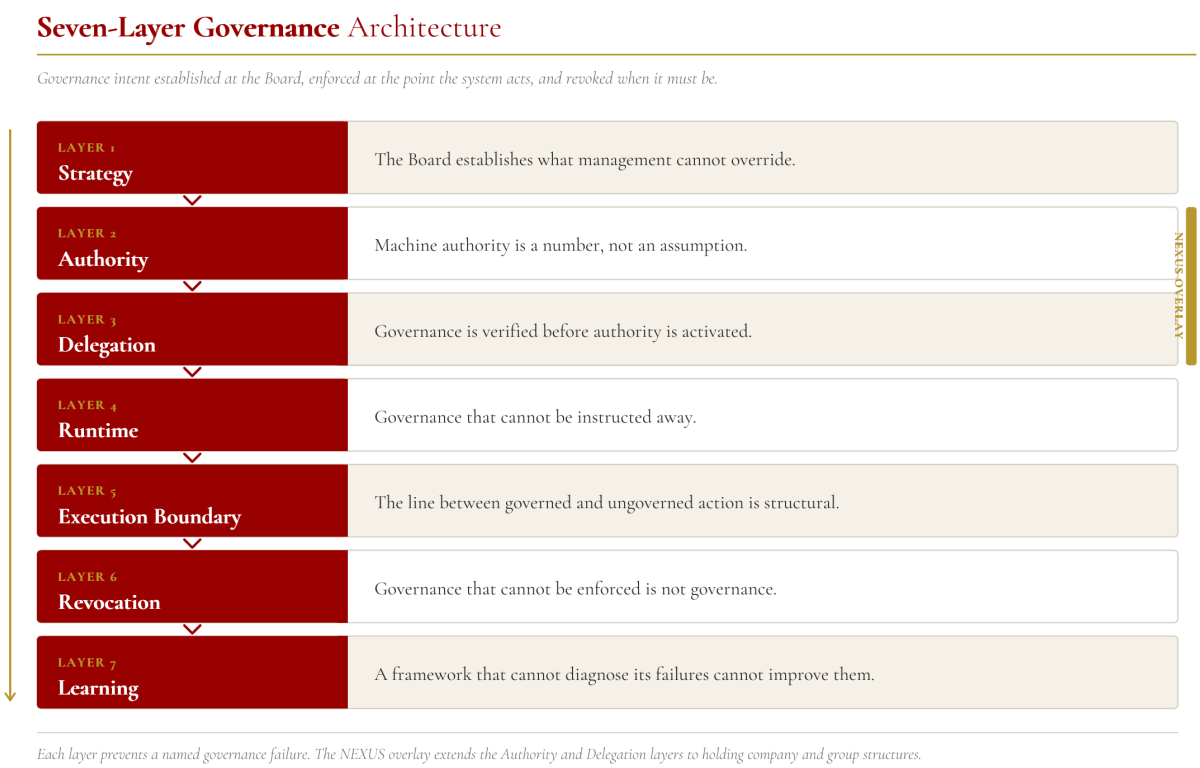
0 mechanisms by which management can unilaterally raise the Autonomy Budget ceiling

Seven-Layer Governance

Architecture.

The pages that follow are written for practitioners — technically literate assessors evaluating whether MANDATE's governance claims are real. Each layer page names the governance failure the layer prevents, then specifies the mechanisms that address it with enough precision to assess the claim. The architecture is a single integrated system: all seven layers operate simultaneously in production. A failure contained by one layer does not propagate to the next; simultaneous failures across multiple layers constitute a material governance breach.

The seven layers below describe the lifecycle of machine authority — how it is established, quantified, verified, enforced, bounded, revoked and improved. They are not the four authority layers used across the instruments themselves, where Layer 1 is Constitutional, Layer 2 Policy, Layer 3 Standards and Layer 4 Operational, and the numbering sets which instrument prevails over which. The two schemes answer different questions and are deliberately kept apart: one describes what governance does, the other which instrument governs. Nothing in the seven layers alters the precedence set by the four.



Strategy	Governance intent is constitutionally established and cannot be modified by management. The failure this layer prevents: governance that exists only at management discretion and evaporates under commercial pressure.
Authority	Machine authority is quantified, classified, and continuously monitored as a portfolio. The failure this layer prevents: authority that accumulates without a ceiling and without the Board having any aggregate view.
Delegation	Governance conditions are independently verified before authority is activated. The failure this layer prevents: systems reaching production before governance architecture is in place, with compliance self-certified by the delivery function.
Runtime	Governance is enforced at every execution moment independently of the model's own reasoning. The failure this layer prevents: governance that exists as policy but has no technical enforcement reach at the point of action.
Execution Boundary	The boundary between governed and ungoverned action is structurally enforced, not described. The failure this layer prevents: governance controls that can be bypassed under retry, high-load, or model self-instruction conditions.
Revocation	The ability to immediately halt machine authority is maintained continuously and exercisable independently by named authorities. The failure this layer prevents: governance that cannot be enforced once a system is in production.
Learning	Governance failures become governance improvements through a structured diagnostic and improvement cycle. The failure this layer prevents: governance that repeats the same failures because it cannot distinguish between what it specified and what it enforced.

> *The NEXUS overlay, described after Layer 7, extends the Authority and Delegation layers to group structures. Individual entity compliance is not group governance — NEXUS governs the aggregate exposure that no individual entity instrument can see.*

LAYER 1 OF 7

Strategy.

The Board establishes what management cannot override.

Every governance framework claims to start at the Board. The question is whether the Board's authority is constitutionally established — binding on management regardless of commercial pressure — or merely implied as the source of a policy mandate that management can quietly erode. The Strategy layer answers this question structurally. Management cannot amend its instruments, waive them under operational necessity, or circumvent them with delivery timelines. Nothing in Layers 2 through 7 can authorise action that exceeds the boundaries set here.

Mechanism	How it works
Board Governance Doctrine *Without a constitutional foundation, governance is only as strong as management's willingness to enforce it.*	The constitutional instrument adopted by the Full Board establishing three non-delegable doctrines: Decision Authority Governance, Governance-Constrained Autonomy, and Delegated Machine Authority. The Group Governance Doctrine adds a fourth doctrine, Group Decision Authority Governance, where a group structure applies. All downstream instruments derive authority from this Doctrine. Management cannot amend it.
Six Constitutional Resolutions *Without Board-level resolutions, governance rests on policy documents that management can revise without Board involvement.*	The Board adopts six resolutions establishing: the enterprise risk posture and Governance-Constrained Autonomy principle; the Governance and QA Framework including the AI QA Function and HOTL programme; the funding envelope and CAIO accountability; the Autonomy Budget as a standing Board instrument; the Machine Decision Authority Matrix as a living instrument; and the Board Monitoring Dashboard and annual Internal Audit obligation. None may be delegated to management.
Risk Appetite Statement *Without a Board-set risk appetite, authority limits are set by management and expand at management's discretion.*	Defines the Board's tolerance for autonomous decision risk: maximum credible loss per scenario type, customer exposure limits, domain-specific authority restrictions, and the enterprise risk posture of Guarded Innovation. All downstream instruments must operate within these parameters.

Mechanism	How it works
Autonomy Budget Ceiling *Without a portfolio ceiling, machine authority accumulates system by system with no aggregate governance visibility.*	The Board-approved aggregate ceiling on total delegated machine authority across the entire AI estate, expressed in ADAE points. Only a Full Board resolution can raise the ceiling. Management cannot expand machine authority beyond the Board-sanctioned limit under any circumstances.
GMI Certification Gates *Without a capability gate, authority can expand faster than governance infrastructure can support it.*	The Governance Maturity Index certifies the organisation's governance capability at five levels. No system may operate at an Autonomy Level exceeding the organisation's certified GMI level. This constraint is constitutionally absolute — it cannot be waived, suspended, or overridden by management action alone. GMI Level 5 requires external assurance.
Prohibited Use Register *Without constitutional prohibited use categories, every use case remains negotiable under commercial pressure.*	A Board-adopted register of AI use cases and decision domains that are absolutely prohibited regardless of commercial opportunity, technical feasibility, or regulatory absence. No MDAM entry may authorise operation in a prohibited domain. The register is a constitutional constraint, not a guideline.
External Assurance Obligation *Without independent assurance, governance quality is self-assessed by the same function responsible for delivery.*	The Board resolution establishing mandatory external assurance of governance controls at GMI Level 5, and Internal Audit as a standing scope item at all levels above GMI Level 1. Assurance is independently conducted and reported to the Board — not a management attestation.

> *A governance framework that begins at the CAIO level begins too late. The Board is not the implied source of mandate — it is the active constitutional actor. Every instrument in Layers 2 through 7 traces to a Board resolution adopted here.*

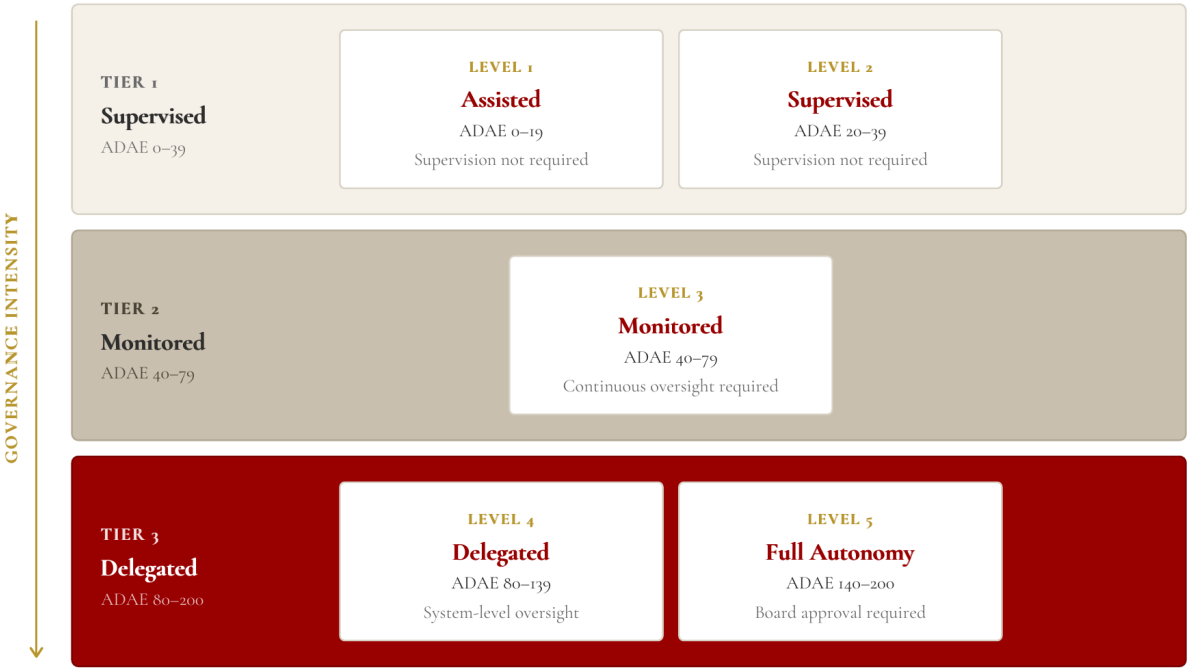
Authority.

Machine authority is a number, not an assumption.

Most organisations have a general sense of how much AI-driven decision-making they are doing. They do not have a number. Without a number, the Board cannot set a ceiling, monitor utilisation, or know whether delegated authority is approaching a level that exceeds its risk appetite. The Authority layer makes machine authority quantifiable, comparable across systems, aggregable at the portfolio level, and continuously monitored against a Board-approved ceiling. It turns the abstract concept of risk appetite into an enforceable governance instrument.

Autonomy Levels and Delegation Tiers

Governance intensity is set by the authority a system holds, not by the technology it uses.



Mechanism	How it works
ADAE Scoring Model *Without a scoring model, authority exposure is estimated qualitatively and cannot be aggregated or compared across systems.*	Scores each system on four weighted dimensions: Financial Authority (40% — maximum financial value affected per decision), Customer Reach (30% — maximum customers affected per execution cycle), Operational Reach (20% — number and criticality of systems accessible), and Decision Velocity (10% — maximum consequential decisions per hour). Eight conservative loading rules apply, four of them percentage loadings on the composite score: delegation chain depth up to +30%, multi-agent orchestration +20%, irreversibility +15%, and provider concentration, +10% by default. The model scores the ceiling of what a system can cause, not the floor of what it typically does.
Machine Decision Authority Matrix *Without a formal domain allocation, systems operate in authority domains that the Board has never explicitly sanctioned.*	Translates the Board's risk appetite into a formal delegation structure: which decisions may be made by machine actors, in which operational domains, at what autonomy level, and under what oversight model. Every production system must have a current MDAM entry. A system operating in a domain without an approved MDAM entry is a zero-tolerance governance breach regardless of how well its technical controls are configured.
Autonomy Register *Without a live register, the organisation's actual machine authority exposure is unknown and ungoverned between review cycles.*	The authoritative live record of every production AI system — its ADAE score, autonomy level, delegation tier, MDAM domain reference, HOTL supervisor assignment, NHI identifiers, and governance status. The Autonomy Register is the operational instrument through which the Board's delegation of machine authority is managed. It is maintained continuously, not updated periodically.
Five Autonomy Levels *Without a classification structure, oversight requirements cannot be calibrated to actual authority exposure.*	Systems are classified at one of five levels: Level 1 (Assistive), Level 2 (Workflow Automation), Level 3 (Supervised Autonomous Agents — HOTL required), Level 4 (Autonomous Execution within Defined Limits), Level 5 (Maximum Delegated Authority). At Levels 4 and 5, the minimum oversight model is HAL, with HOTL required at the High tier and HITL at the Critical tier. No system may operate at a level exceeding the organisation's certified GMI level.

Mechanism	How it works
Utilisation Bands *Without automatic governance triggers, the Board only knows about budget pressure after it has become a governance breach.*	Portfolio ADAE utilisation against the Board ceiling is monitored against four bands with automatic governance responses. Below 80%: normal operations. 80–89%: Board review required, no new deployment without BRC sign-off. 90–99%: deployment moratorium. 100%: Full Board resolution required to raise ceiling or decommission systems to return below it.
Maximum Action Magnitude *Without a per-action financial ceiling enforced at the infrastructure layer, a compromised or misconfigured agent has no technical barrier to causing unlimited financial harm.*	The MDAM sets the per-action financial authority ceiling for each registered system and enforces it at the infrastructure layer, independently of the model. A compromised or malfunctioning agent cannot cause a financial action that exceeds its MAM ceiling through any execution path—including prompt manipulation or model self-instruction.
Board Monitoring Dashboard *Without a structured quarterly instrument with defined thresholds, Board oversight is reactive rather than continuous.*	Presents 27 Key Risk Indicators across 13 governance domains to the Board Risk Committee quarterly. KRIs cover Autonomy Budget utilisation, GMI compliance, HOTL certification coverage, continuous admissibility events, incident trends, change velocity, NHI governance, vendor concentration, and assurance outcomes. Each KRI has a defined threshold, red/amber/green status, and a mandatory escalation pathway.
Eight-Stage Control Loop *Without a closed feedback loop, governance becomes retrospective — the Board receives evidence of what happened rather than governing what is happening.*	The eight-stage cycle set out earlier in this document, from Board risk appetite through to the quarterly dashboard. Stage 8 returns monitoring evidence to Stage 1 at each quarterly BRC meeting. The Board, not management, closes the loop.

LAYER 3 OF 7

Delegation.

Governance is verified before authority is activated.

The most common governance failure mode in AI deployment is not malicious circumvention — it is the ordinary commercial pressure of deployment timelines. Systems reach production before governance infrastructure is in place. The kill-switch is untested. The SRE is unconfigured. The HOTL supervisor is unassigned. The Delegation layer makes this impossible: every deployment passes through an independent gate that validates the complete governance baseline before executing a single decision in production. The gate authority sits outside engineering and delivery leadership — it cannot be overridden by commercial urgency.

Mechanism	How it works
AI QA Gate *Without an independent gate, deployment sign-off rests with the delivery function that has commercial incentives to ship.*	The independent pre-deployment quality assurance gate operated by the AI QA Function, which holds release gate authority independently of engineering and delivery leadership. Validates: ADAE scoring complete, MDAM domain entry approved, kill-switch tested, SRE configured and tested, HOTL supervisor assigned and certified (Level 3+), Autonomy Register entry complete, write-before-execute logging operational, vendor governance terms verified. The AI QA Function may not accept direction from engineering or delivery on release decisions.
Pre-Deployment Checklist *Without a mandatory baseline checklist attested to the Board, governance completeness is self-declared and unverifiable.*	For Level 3 and above, the CAIO must attest to the Board Risk Committee that all eleven gate conditions are confirmed operational before deployment is approved. Partial compliance is not sufficient — all eleven must pass. The checklist covers: Autonomy Register completeness; an independent AI QA Function; Safety Runtime with tested kill-switch and circuit-breaks; the HOTL framework and approved Threshold Register; GMI Level 2 certification; explainability logging at the 98% standard; model version pinning with tested rollback; NHI credentials under JML control; at least one certified HOTL supervisor per workflow; incident response briefing; and a signed Accountability Acceptance Record from the executive named in the MDAM entry.

Mechanism	How it works
HOTL Certification Framework *Without certified supervisors, human oversight of high-autonomy systems is nominal rather than substantive — supervisors rubber-stamp without genuine review capacity.*	A five-component certification programme required before overseeing a Level 3+ system: Foundation governance knowledge (4 hours, ≥80% pass); system-specific training (4–8 hours, ≥90% pass); intervention skills covering circuit-break, kill-switch, and escalation (4 hours); automation bias training (4 hours, ≥75% identification rate); and assessed live supervision exercise (4 hours minimum, Satisfactory required). Certification is valid for 12 months. A lapsed certification requires repeating the full programme.
GMI Certification Process *Without an independent certification process, GMI levels are self-assessed, and the constitutional gate between GMI level and autonomy level has no enforcement mechanism.*	The operational process through which governance capability is independently assessed and certified at each GMI level. The CRO certifies Levels 1 and 2, the Board Risk Committee certifies Levels 3 and 4 on Internal Audit opinion, and the Full Board certifies Level 5 on Internal Audit and external assurance. The process produces a signed Certification Artefact recorded in the Autonomy Register. GMI certification is a perpetual constraint, not a one-time approval event — the gate remains active for every subsequent deployment.
Change and Release Policy *Without a formal change classification process, model updates and configuration changes enter production without governance review, silently changing system behaviour under existing authorisations.*	Governs all production AI system changes through formal classification: Standard (expedited gate, two-day review), Significant (standard gate, five-day review), Material (full gate re-run, ten-day review, BRC notification), and Emergency (Section 6 protocol, with a mandatory retrospective gate within five business days). Undisclosed material changes by vendors are classified as Level 2 governance incidents. Version pinning and regression testing are required before any model or prompt change enters production.
Change Classification Rules *Without explicit rules specifying which change types trigger which governance response, classification is contested, and governance requirements can be argued down.*	Specifies which change types fall into which classification tier. Material changes — triggering full gate re-run — include: any change to a system's action domain or MDAM entry; any model version change affecting a Critical-tier system; any change that could affect kill-switch or circuit-break configuration; any change to SRE policy rule configuration.

Mechanism	How it works
Operational Implementation Framework *Without practical implementation guidance, governance policies are interpreted inconsistently across delivery teams, producing uneven governance coverage.*	Provides practical implementation guidance across four phases: Phase 1 Foundation (months 1–6), Phase 2 Assurance Build (months 7–12), Phase 3 Maturity and Scale (months 13–24), Phase 4 Steady State (month 25+). Guidance, not obligations — the binding obligations are in the policies and standards it references.
Supplier and Third-Party AI Governance Standard *Without mandatory contractual governance terms, vendor AI systems operate under the organisation's delegated authority with none of the organisation's governance controls accessible to it.*	Governs the full vendor AI system lifecycle: pre-procurement due diligence, mandatory contractual terms, onboarding, monitoring, and breach response. Mandatory terms for Level 3+ vendor systems include: real-time telemetry access, a Circuit-Break API accessible to the organisation, 30-day advance notice of material model changes, incident-notification obligations, and audit rights. A 40% vendor concentration limit applies — no single model provider may account for more than 40% of the ADAE budget.

LAYER 4 OF 7

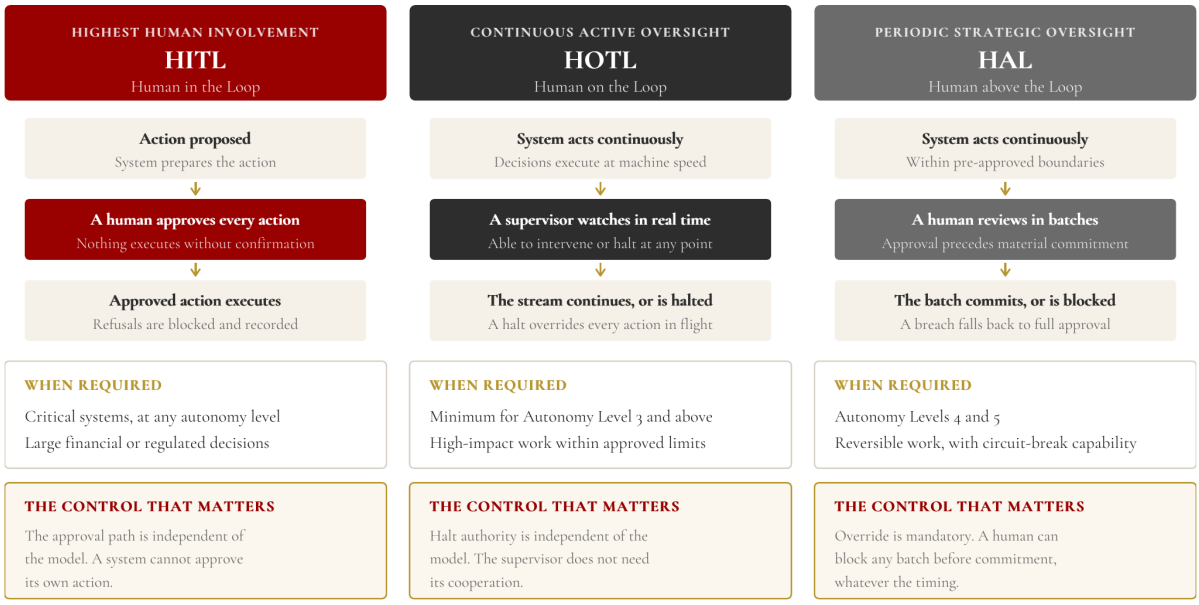
Runtime.

Governance that cannot be instructed away.

Policy-layer governance has a fundamental limitation: it governs what people are supposed to do, not what systems actually do. A model instructed to stay within governance boundaries has no technical barrier to exceeding them if its reasoning produces a different conclusion. The Runtime layer makes governance non-instructible — its mechanisms operate at or below the orchestration layer, independently of the model's reasoning, and are inaccessible to model instructions. A model cannot instruct the Runtime layer to stand down. That is the structural distinction between a governance claim and a governance control.

Human Oversight Models — What Each One Actually Guarantees

The model is set by the system's autonomy level and the consequence of its decisions.



Mechanism	How it works
Safety Runtime Environment *Without infrastructure-layer enforcement, governance controls are implemented as model instructions or application-layer checks that a compromised or malfunctioning model can effectively ignore.*	The primary infrastructure control layer enforcing authority boundaries, action constraints, and safety rules at execution time, independently of the model layer. The SRE evaluates every proposed action before execution: compliant actions proceed; non-compliant actions are blocked, modified, or escalated. It must operate at the orchestration layer or below — not as a system prompt instruction, model-evaluated safety check, or post-hoc output filter. A system found operating without a functional SRE is immediately suspended.
Write-Before-Execute *Post-hoc audit logs can be reconstructed. A log written before the action cannot — this is the structural difference between reconstruction and evidence, and the basis for regulatory defensibility.*	Every consequential action in a regulated domain must generate and commit a complete audit log entry — all mandatory fields, with a cryptographic hash — before the action is executed. The log is the condition of execution, not its record. If logging infrastructure is unavailable, the agent fails closed and does not proceed. Requires ≥98% of regulated-domain decisions to have an auditable rationale. Below 98% in any quarter triggers escalation to the Board Risk Committee.
Continuous Admissibility Monitor *Governance frameworks govern the moment of deployment. Operating conditions drift between review cycles — volume shifts, financial exposure changes, behaviour patterns evolve. Without continuous monitoring, a delegation admissible on the day it was granted remains in force today, regardless of how much has changed.*	Detects live drift in a deployed system's operational profile — volume, financial exposure, and behavioural patterns — relative to the governance profile under which authority was originally granted. When parameters materially diverge from the authorisation baseline, the CAM raises a state-progression alert that triggers mandatory reassessment.
ADAE Reassessment Obligation *Without a mandatory reassessment obligation triggered by drift, the CAM alert becomes advisory — systems can continue operating at their current authority level while drift persists.*	A confirmed CAM drift alert triggers mandatory ADAE reassessment within 10 business days. If the operational profile has materially changed across any of the four scored dimensions, reauthorisation at the current band is required before the system can continue operating at its existing autonomy level. Reassessment is not discretionary.

Mechanism	How it works
Structural Change Trigger *An organisational restructure changes the governance context in which a delegation was made. Without a trigger, systems continue operating under authorisations that were granted for an organisational reality that no longer exists.*	Any material organisational change — merger, acquisition, entity restructure, business unit reorganisation — triggers an admissibility review of every affected system. An authorisation granted under one organisational context does not survive structural discontinuity without active reconfirmation.
Kill-Switch *A governance control without a halt capability is not a governance control; it can observe failure but cannot stop it.*	Every production agentic system must have a kill-switch capability independent of the application layer and the model layer. A compromised or malfunctioning model cannot disable its own kill-switch. Activation timelines: Level 1–3 systems halt within 60 seconds; Level 4–5 within 30 seconds with hard stop and state preservation. Tested quarterly (monthly for Level 4–5). A non-functional kill-switch is a zero-tolerance Level 2 incident — immediate suspension.
Circuit-Break Controls *Without automated threshold-based controls, human supervisors are the only mechanism for catching governance drift — and human span-of-control has hard limits.*	Automated controls that shift a system from autonomous operation to human-in-the-loop mode when defined parameters are approached — before a governance breach occurs. Circuit-breaks monitor decision volume, error rate, financial exposure trajectory, and span-of-control thresholds. Activation timelines: 30 seconds for Level 4–5; 60 seconds for Level 1–3. Operates independently of the model layer.
Logging and Explainability Standard *Without a mandatory log schema and retention obligation, audit trails are incomplete, inconsistent across systems, and legally indefensible.*	Specifies the mandatory 24-field log schema, with five further fields for systems in Responsible AI Framework scope, the ≥98% auditable rationale standard, 7-year minimum retention obligation, and cryptographic integrity requirements for all decision logs. Establishes the explainability obligations for adverse automated decisions — the standard for regulatory defensibility under APRA CPS 230, ASIC RG 271, and the Privacy Act.
Model and Prompt Versioning Standard *A model whose behaviour has changed is effectively a different system operating under the governance profile of the original. Without version control, silent drift invalidates existing authorisations without triggering reassessment.*	Governs version pinning, regression testing requirements, rollback procedures, and model supply chain risk management. Prevents silent behaviour drift from unmanaged model or prompt changes — requires testing and validation before any version change reaches production.

Mechanism	How it works
NHI Security and Lifecycle Governance *An agentic AI system without a governed identity is an ungoverned actor — it can authenticate and act with no accountability chain.*	Every agentic AI system is assigned a unique Non-Human Identity with provisioning, modification, and deprovisioning governed under JML lifecycle controls. Credentials are scoped to the approved action domain with mandatory rotation schedules: short-lived tokens 24 hours; API keys 90 days; client certificates 12 months (6 months for Level 4–5). Zero-orphan enforcement: disable any active credential with no governance record within 1 hour of discovery.
Policy Conflict Resolution Engine *Without a constitutional priority hierarchy for policy conflicts, a system facing conflicting governance rules defaults to its own reasoning — which may prioritise efficiency over safety.*	Where governance rules conflict at runtime, the engine applies a constitutional priority hierarchy: Safety policies override all others; Regulatory compliance policies override Security and Efficiency; Security policies override Efficiency. The resolution is determined before execution, not after. Conflicts are logged as governance events for subsequent review.
Admissibility Decay Rate Score *Without a measure of how fast a delegation loses admissibility, drift tolerances are set uniformly across systems that decay at very different rates.*	Scores how quickly a system's delegation loses admissibility, on its own four factors rather than the ADAE dimensions. The score calibrates the drift tolerances the Continuous Admissibility Monitor applies to that system, so a fast-decaying delegation is tested more often than a stable one.
Governance Signal Convergence Monitor *Without convergence monitoring, several sub-threshold governance signals can arrive together, and none individually triggers a response, while jointly they indicate a material failure.*	Detects where separate governance signals converge on the same system or control within a defined window. Tier 1 convergence events must be assessed within five business days; failure to do so is reported as KRI-20 and escalated to the CRO within 24 hours.
Delegation Chain Integrity Assessment *Without chain-level assessment, each agent in a multi-agent architecture is individually compliant while the chain as a whole exercises authority no single agent holds.*	Confirms at deployment and at each ADAE review that every agent in a chain is registered and scored, that the orchestrator carries the delegation chain depth loading, that aggregate authority stays within the orchestrator's ceiling, and that every instruction carries verifiable provenance. The chain carries its own criticality tier, set by the consequence of chain failure rather than by the highest tier within it. Reported as KRI-21.

Mechanism	How it works
Foundation Model Characterisation *Without characterising the foundation model a system depends on, an unannounced provider update changes behaviour under an authorisation granted for the previous model.*	Records the foundation model identifier, provider, known failure modes and update channel for every dependent system, and is the reference against which overseer toolchain declarations are tested for shared failure modes. Reported as KRI-22.

LAYER 5 OF 7

Execution Boundary.

The line between governed and ungoverned action is structural.

Runtime governance controls can be architected correctly in design and still fail at execution. The failure mode is well understood: alternative execution paths, permissive timeouts, authority inheritance across agent boundaries, and prompt injection create gaps between the governance model and execution reality. The Execution Boundary layer closes these gaps structurally — not through additional policy rules, but through architectural constraints that make the gaps technically impossible. The question this layer answers is not "does governance apply here?" but "can the system act without governance applying here?" The answer must be no, under every condition, including conditions the governance designer did not anticipate.

Mechanism	How it works
Non-Bypassable SRE Control Surface *An SRE that can be bypassed under any condition — retry, high load, degraded mode — is not a governance control. It is a preference that yields under pressure.*	The SRE control surface at the execution boundary is structurally non-bypassable. There is no alternative execution path — not under normal operation, not under retry conditions, not under high-load degradation, not through model self-instruction. Every proposed action must pass through the SRE before reaching an execution system. An SRE that can be bypassed under any condition is a zero-tolerance design violation.
Timeout Fail-Closed Obligation *A system that defaults to permissive behaviour when governance infrastructure is unavailable treats governance as an enhancement rather than a condition of operation.*	If the SRE evaluation times out because the policy engine is unavailable, overloaded, or unresponsive, the action is blocked, not permitted by default. Governance unavailability is not a reason to proceed without governance. A system that defaults to permissive behaviour on SRE timeout is a zero-tolerance design violation.
Independent Authority Validation *In multi-agent workflows, authority inheritance from orchestrators creates a single point of governance failure — one compromised orchestrator can grant unbounded authority to all sub-agents.*	In multi-agent workflows, authority validation is performed independently at each delegation boundary — not inherited from the orchestrating agent. Each sub-agent validates that the instruction it receives falls within its registered authority scope, regardless of whether the orchestrator claims the authority is valid. An orchestrator cannot grant a sub-agent authority the sub-agent does not independently hold.

Mechanism	How it works
Injection-Resistant Inter-Agent Communication *Prompt injection is not a theoretical risk in multi-agent environments — it is an active attack surface that can cause agents to act outside their authority scope by manipulating the content of instructions passed between agents.*	Inter-agent communication must be resistant to prompt injection. Controls include: structured inter-agent messaging protocols that do not pass raw user or environment content as agent instructions; content validation before instructions are passed between agents; and SRE-level monitoring of inter-agent instruction patterns for anomalous authority requests.
Mutual Authentication (Level 3+) *Without mutual authentication, a receiving agent has no technical basis to verify that an instruction came from an authorised source rather than a compromised or spoofed agent.*	Systems at Autonomy Level 3 and above must implement mutual authentication for inter-agent communication: the receiving agent verifies the sending agent's identity before processing its instruction, using NHI credentials distinct from the sending agent's external resource access credentials. Shared credentials across authentication contexts are not compliant.
MAM Infrastructure-Layer Enforcement *An application-layer financial ceiling can be bypassed by a compromised or malfunctioning agent. The distinction between an infrastructure-layer and application-layer ceiling is the difference between a governance control and a governance aspiration.*	The Maximum Action Magnitude — the per-action financial ceiling for each system — is enforced at the infrastructure layer, not at the application layer. Infrastructure-layer enforcement means the agent cannot technically exceed the ceiling, regardless of its reasoning or model output. An agent cannot instruct itself past its MAM ceiling.
Output Validation Pipeline *An agent output that is acted upon before validation can cause harm even if the agent itself was operating within its authority scope — because the output may be malformed, outside the approved schema, or injected by adversarial content.*	Before any agent output is acted upon — by an execution system, downstream agent, or human operator — it is validated against the system's approved output schema and authority scope. Outputs outside approved parameters are blocked or quarantined. The pipeline operates after the SRE and before execution: the final structural check before consequence.

> *The claim that governance is enforced at the execution boundary is verifiable. A practitioner assessing this layer asks: can the agent execute an action that the SRE would block? Can it exceed its MAM ceiling? Can it grant a sub-agent authority it does not hold? If yes to any of these, the execution boundary is not governed — it is described.*

Revocation.

Governance that cannot be enforced is not governance.

The ability to halt machine authority is not a fallback mechanism — it is a primary governance control. Every delegation of machine authority must be revocable immediately and completely by named authorities, independently of one another, at the application, infrastructure, and identity layers. A governance architecture that can authorise but cannot revoke is constitutionally incomplete. The Revocation layer governs both the revocation mechanism and the conditions that preserve revocation capability — ensuring that, when a halt is directed, it executes within defined timeframes, propagates to all downstream delegations, and cannot be circumvented by the halted system.

Mechanism	How it works
Emergency NHI Shutdown *An application-layer kill-switch can fail — through compromise, configuration failure, or vendor inaccessibility. The identity layer provides a fallback that is independent of the application and effective regardless of the system's operational state.*	The CISO, CAIO, or CRO may each independently direct the immediate suspension of all NHI credentials associated with any production agentic AI system where continued operation presents material risk. No other approval is required. The CTO executes within 30 minutes of direction and confirms completion within 5 minutes. Effect: the system cannot authenticate to any resource and cannot take any autonomous action. This is the fallback when the application-layer kill-switch cannot be activated or has failed.
Downstream Delegation Propagation *A revocation that does not propagate to downstream systems creates orphan authority chains — sub-agents continuing to act on behalf of a system whose authority has been extinguished, with no accountability chain.*	When a system's authority is revoked — through kill-switch activation, NHI suspension, or MDAM entry withdrawal — revocation must propagate to all downstream systems and sub-agents operating under delegated authority from the revoked system. Propagation is a mandatory governance obligation, not a best-practice recommendation.
Credential Rotation and Expiry Standards *Credentials that are not rotated remain valid indefinitely — creating persistent access risk long after the governance justification for that access has expired.*	Credential validity periods by type and autonomy level: short-lived tokens 24 hours; API keys 90 days (shorter if provider requires); client certificates 12 months standard, 6 months for Level 4–5; signing keys 180 days standard, 90 days for Level 4–5. Rotation must be completed before expiry — a credential not rotated within 24 hours of the deadline is automatically suspended.

Mechanism	How it works
Privilege Scope Management *Excess privileges are the primary mechanism through which a compromised agent exceeds its authority scope — an agent with broader permissions than it requires can cause harm beyond its governance boundaries.*	Every NHI holds exactly the permissions required for its approved action domain — no more and no less. Privilege drift must be remediated within 4 hours of discovery. Drift that was exercised: Level 2 incident. Drift present but not exercised: Level 1 incident. Privilege scope reviews are conducted quarterly by the CISO, semi-annually by the CRO's function, and annually by Internal Audit.
Orphan NHI Detection *An orphan NHI — an active credential with no governance accountability — is an ungoverned actor that can authenticate and act without any organisational authority for its existence.*	Zero tolerance applies: orphan NHIs must be disabled within 1 hour of discovery regardless of active use, classified as a Level 2 incident, and reported to the CAIO within 24 hours. Quarterly automated discovery scans are mandatory across all identity sources. An orphan NHI discovered more than 24 hours after its governance coverage lapsed is evidence of a process failure requiring CAPA.
Independent Shutdown Authority *A single-authority revocation model creates a single point of failure — incapacitation of one authority removes the ability to halt machine authority in a crisis.*	Revocation authority is distributed across three independent executives — CISO, CAIO, and CRO — each of whom can independently direct an Emergency NHI Shutdown. This independence is structural: a governance failure that incapacitates one authority does not incapacitate the others. For Level 3 incidents, NHI re-enablement requires CAIO confirmation that the governance risk has been resolved, and CRO concurrence before systems are reinstated.

LAYER 7 OF 7

Learning.

A framework that cannot diagnose its failures cannot improve them.

Most incident management frameworks produce a record of what happened and a corrective action plan. They do not produce a structural diagnosis of whether the failure was a governance enforcement failure — a control that existed but did not operate — or a governance specification failure — a gap in the framework design itself. These two failure types require categorically different remediation. Conflating them produces CAPA plans that treat symptoms rather than root causes. The Learning layer makes this distinction mandatory, requires the framework to improve in response to both failure types, and automatically escalates when improvement does not occur.

Mechanism	How it works
Incident Severity Classification *Without formal severity classification, every incident is handled with the same urgency — or less urgency than warranted — and escalation obligations are unclear.*	Three severity levels: Level 1 (Minor — deviation from expected behaviour, no immediate harm risk, investigation required); Level 2 (Significant — behaviour outside approved authority limits causing financial loss, customer harm, or governance control failure; CRO notified within 24 hours, BRC at next meeting); Level 3 (Material — significant customer harm, material financial loss, regulatory consequences, or simultaneous failure of multiple foundational governance controls; kill-switch activation, full crisis protocol, Board advisory within 48 hours). Classification is performed by the CAIO; no business unit may direct reclassification to a lower level.
Enforcement vs Specification Failure *A CAPA plan that addresses only the immediate symptom without diagnosing whether the control existed but failed, or was never specified, will not prevent recurrence.*	A named diagnostic distinction applied to every Level 2 and Level 3 incident. An enforcement failure: a governance control existed and was specified correctly but was absent, bypassed, or inoperative at the point of failure — requiring control operability remediation. A specification failure: governance controls operated as designed but the design was insufficient to prevent the failure — requiring framework redesign. Conflating the two produces remediation that does not address the actual failure mode.

Mechanism	How it works
Root Cause Analysis Standard *An RCA that addresses only the immediate trigger without examining enabling conditions and systemic factors will prevent recurrence of that specific event but not of the class of events it represents.*	Every Level 2 and Level 3 incident requires RCA across three dimensions: immediate trigger (the proximate event), enabling conditions (the governance or technical gaps that allowed the trigger to cause harm), and systemic factors (the broader organisational or process conditions that created the enabling conditions). The root cause statement must be agreed between the CAIO and CRO before the CAPA plan is finalised.
CAPA Framework *Without a structured corrective and preventive action framework, incident closure produces a commitment to improve without specifying what will be done, by whom, by when, and how completion will be verified.*	Every Level 2 and Level 3 incident requires a CAPA plan with corrective actions that remedy the harm and close the immediate control gap, and preventive actions that address systemic factors to prevent recurrence. Every action requires a named owner, a target date, and a verification method. CAPA plans that address only the immediate trigger — not the root cause — are rejected and must be revised within 14 days.
Post-Incident Review *Without a structured review of incident response quality, the governance framework cannot assess whether its incident management process itself performed adequately.*	A structured review conducted within 10 business days of Level 2 or Level 3 closure, assessing detection quality, classification accuracy, response timeliness, escalation adherence, and investigation thoroughness — not only the technical cause of the incident. The PIR examines whether the governance framework responded as designed.
Annual Incident Trend Analysis *Individual incident reviews reveal specific failures. Only a portfolio-level annual analysis reveals systemic patterns — recurring root cause categories that indicate structural gaps in the governance architecture.*	An annual review of all classified incidents covering trend direction, root cause category distribution, enforcement vs specification failure ratio, CAPA completion rate, and repeat incident rate. Produced annually and submitted to the BRC annual governance framework review. Provides the evidence base for recalibrating Board risk appetite.
Lessons Learned Register *Without a maintained register, governance improvements derived from incidents are not systematically captured and cannot be verified as having been incorporated into policies and standards.*	A mandatory governance artefact capturing the primary improvement derived from each Level 2 and Level 3 incident. Internal Audit must verify completeness and accuracy as part of the annual governance audit scope. The register is reviewed at each ADAE methodology review and annual policy review cycle to identify learnings not yet reflected in framework instruments.

Mechanism	How it works
Framework Improvement Obligation *Without a mandatory improvement obligation triggered by recurring root cause categories, the annual trend analysis produces a diagnosis without producing a remedy.*	When annual trend analysis identifies a recurring root cause category — the same type appearing in two or more incidents within 24 months — the CAIO must propose a specific framework improvement and submit it to the CRO for approval within 30 days. A recurring root cause not addressed by a framework improvement is a significant non-compliance with this policy.
Recurring Root Cause Escalation Rule *Without automatic escalation for repeat incidents, a CAPA plan that fails to prevent recurrence can be accepted and closed without consequence.*	A repeat incident — substantially identical in root cause to a prior incident within 12 months — triggers automatic escalation: the CAIO must produce a framework improvement proposal within 30 days. If not produced within 30 days, the breach escalates to zero-tolerance classification. The rule makes the framework structurally self-correcting by obligation, not discretionary intent.

Group AI Governance Framework

NEXUS.

Individual entity compliance is not group governance. A holding company whose subsidiaries each satisfy their individual Autonomy Budget and GMI requirements may still be exposed at the group level — because cross-entity AI systems operate across legal-entity boundaries simultaneously, and no individual entity's governance instruments can see or govern that aggregate exposure. The governance failure that NEXUS prevents is invisible from any single entity's vantage point: machine authority accumulating across the group without any instrument holding it to a group-level ceiling, and systems crossing entity boundaries without any authority governing those boundary crossings.

Mechanism	How it works
Group Portfolio ADAE *Without a group-level aggregate score, the group's total machine authority exposure is the sum of numbers no single governance instrument has ever calculated.*	The sum of all Final ADAE Scores across all active production systems in all subsidiaries and the holding company, with cross-entity loadings applied. Calculated quarterly by the Group CAIO. The Group Board monitors this number against the Group Autonomy Budget ceiling.
Group Autonomy Budget Ceiling *Without a group ceiling, each subsidiary operates within its own ceiling while the group aggregate is unconstrained — compliant entities, non-compliant group.*	The Group Board-approved aggregate ceiling on total delegated machine authority across the entire group AI estate. Set by Full Group Board resolution. May only be raised by Full Group Board resolution — the Group CAIO, Group CRO, and CEO cannot raise it unilaterally. The sum of all subsidiary individual ceilings is the maximum permissible group ceiling.
Subsidiary Sub-Budget Allocation *Without allocated sub-budgets, subsidiary Boards set their own ceilings independently — the group ceiling has no mechanism for enforcement at the entity level.*	The Group Board allocates an ADAE sub-budget to each subsidiary as part of the Group Autonomy Budget instrument. Sub-budgets are not independently set by subsidiary Boards — they are allocated by the Group Board. The sum of all subsidiary sub-budgets plus the holding company's own ADAE allowance must not exceed the Group ceiling. The Group CAIO is notified at 80% subsidiary utilisation; the subsidiary Board at 90%.

Mechanism	How it works
Cross-Entity ADAE Loading (one band, +25 ADAE points) *A system that crosses entity boundaries amplifies governance exposure in ways a single-entity ADAE score does not capture — it affects customers, data, and operations across different legal, regulatory, and accountability frameworks simultaneously.*	Any cross-entity system is subject to a mandatory loading of one band, being +25 ADAE points, applied after every other loading. This loading is mandatory and cannot be waived — not by commercial pressure, not by the Group CAIO, not by the Group Board. Moderate becomes High; High becomes Critical; Critical becomes Systemic. A system already Systemic stays Systemic; that band is terminal, and points are still added.
Group CAIO Constitutional Role *Without a constitutionally designated Group CAIO, cross-entity classification authority is undefined and group-level governance has no accountable executive.*	The Group CAIO is a mandatory constitutional role and a precondition for deploying any cross-entity system. It is not a Governance Maturity Index criterion: the criteria for each GMI level are set exhaustively by the GMI Framework. The Group CAIO holds final determination authority over C1–C4 classification, cross-entity ADAE loading verification, Group Portfolio ADAE consolidation, and Group Board Monitoring Dashboard reporting.
C1–C4 Classification Criteria *Without explicit criteria, cross-entity system identification depends on self-declaration by system owners who may not recognise cross-entity characteristics — or may prefer not to.*	A system is cross-entity if it meets any of four criteria: C1 — reads from or writes to data stores containing records from more than one legal entity; C2 — executes actions affecting the operations, customers, or financial position of more than one entity; C3 — acts as orchestrator or sub-agent in a multi-agent architecture spanning more than one entity; C4 — outputs used as material inputs into decisions made by or on behalf of more than one entity, even if the system is registered to a single entity. The Group CAIO's classification determination is final.
Concurrent Sanction for Sanction-Constrained Entities *Without concurrent sanction, a Group Board authorises machine authority over a subsidiary whose directors cannot lawfully act in the holding company's interests, and the authorisation is ineffective for that entity.*	Where a subsidiary is not wholly owned, lacks constitutional authority to act in the parent's interests, holds its own licence, or sits in another jurisdiction, Group Board sanction is necessary but not sufficient. That subsidiary's Board must concurrently sanction the authority exercised over its operations, and both sanctions must be recorded before deployment. The Group CAIO determines sanction-constrained status with General Counsel concurrence and reconfirms it annually. Systems running without the concurrent sanction are counted as KRI-GGF-05 and escalated to the Group CRO within four hours.

Mechanism	How it works
Five Group KRIs *Without group-level KRIs on the Group Board Monitoring Dashboard, the Group Board has no structured visibility into cross-entity governance dimensions that individual entity dashboards cannot surface.*	The Group Board Monitoring Dashboard includes five Group KRIs reported quarterly: KRI-GGF-01 — Group Portfolio ADAE utilisation; KRI-GGF-02 — cross-entity system count; KRI-GGF-03 — subsidiary sub-budget compliance; KRI-GGF-04 — Group CAIO escalation events; and KRI-GGF-05 — unsanctioned cross-entity systems. The Group dashboard is a distinct report from the entity-level dashboard, which the GGF does not amend. Each Group KRI carries a defined threshold, a Red, Amber, Green status and a mandatory escalation pathway.

> *A holding company whose subsidiaries are individually compliant but whose cross-entity systems operate without group-level governance is exposed in exactly the way that is hardest to see from any single entity's vantage point. NEXUS governs what no individual entity instrument can reach.*

Complete Mechanism Inventory

by Governance Layer.

Every mechanism listed below is specified in a normative instrument with defined ownership, thresholds, escalation pathways, and breach classifications. Each mechanism below is named with the failure it prevents and how it works, in the layer pages above.

STRATEGY

- Board Governance Doctrine
- Six Constitutional Resolutions
- Risk Appetite Statement
- Autonomy Budget Ceiling
- GMI Certification Gates
- Prohibited Use Register
- External Assurance Obligation

AUTHORITY

- ADAE Scoring Model
- Machine Decision Authority Matrix
- Autonomy Register
- Five Autonomy Levels
- Utilisation Bands (4)
- Maximum Action Magnitude
- Board Monitoring Dashboard
- 27 Key Risk Indicators
- Eight-Stage Control Loop

DELEGATION

- AI QA Gate
- Pre-Deployment Checklist
- HOTL Certification Framework
- GMI Certification Process
- Change and Release Policy
- Change Classification Rules
- Operational Implementation Framework
- Supplier / Third-Party Standard

RUNTIME

- Safety Runtime Environment
- Write-Before-Execute
- Continuous Admissibility Monitor
- ADAE Reassessment Obligation
- Structural Change Trigger
- Kill-Switch
- Circuit-Break Controls
- Logging and Explainability Standard
- Model and Prompt Versioning Standard
- NHI Security and Lifecycle Governance
- Policy Conflict Resolution Engine
- Admissibility Decay Rate Score
- Governance Signal Convergence Monitor
- Delegation Chain Integrity Assessment
- Foundation Model Characterisation

EXECUTION BOUNDARY

- Non-Bypassable SRE Control Surface
- Timeout Fail-Closed Obligation
- Independent Authority Validation
- Injection-Resistant Inter-Agent Communication
- Mutual Authentication (Level 3+)
- MAM Infrastructure-Layer Enforcement
- Output Validation Pipeline

REVOCATION

- Emergency NHI Shutdown
- Downstream Delegation Propagation
- Credential Rotation and Expiry Standards
- Privilege Scope Management
- Orphan NHI Detection
- Independent Shutdown Authority

LEARNING

- Incident Severity Classification
- Enforcement vs Specification Failure
- Root Cause Analysis Standard
- CAPA Framework
- Post-Incident Review
- Annual Incident Trend Analysis
- Lessons Learned Register
- Framework Improvement Obligation
- Recurring Root Cause Escalation Rule

NEXUS OVERLAY

Group Portfolio ADAE · Group Autonomy Budget Ceiling · Subsidiary Sub-Budget Allocation · Cross-Entity ADAE Loading (one band, +25 ADAE points, mandatory, non-waivable) · Group CAIO Constitutional Role · C1–C4 Cross-Entity Classification Criteria · Concurrent Sanction for Sanction-Constrained Entities · Five Group KRIs

No mechanism in this inventory is aspirational. Each is specified in a normative instrument—with defined ownership, thresholds, escalation pathways, and breach classifications.

GLOSSARY

Terms Used

in This Document.

ADAE · Autonomous Decision Authority Exposure
The composite score produced by applying the four-dimensional decision authority scoring methodology. Determines a system's Autonomy Level, delegation tier, and the approval authority its deployment requires.

ADAE Bands
The five exposure bands into which a composite ADAE score falls: Low (0–24), Moderate (25–49), High (50–74), Critical (75–99) and Systemic (100 and above). The band sets the deployment approval authority and the minimum governance intensity. A system reaches the Systemic band only where a conservative loading applies.

Admissibility Decay Rate Score
A measure of how quickly a delegation loses admissibility, scored on its own four factors rather than the ADAE dimensions. It calibrates the drift tolerances the Continuous Admissibility Monitor applies, so it tests a fast-decaying delegation more often than a stable one.

AI QA Function
The independent assurance function operating within the first line, run by the Chief AI Officer and structurally independent of engineering and delivery. It holds release-gate authority as a hard stop that delivery teams cannot bypass and may recommend immediately suspending a production system.

Autonomy Budget
The Board-approved aggregate ceiling on total delegated machine authority across all production agentic AI systems in the organisation's estate. Expressed in ADAE composite score points. Prevents the invisible concentration of institutional authority in production AI systems at the portfolio level, a risk that no individual system limit can detect. No deployment that would push Portfolio utilisation above 80% of the ceiling may proceed without Board Risk Committee review.

Autonomy Level
A five-point classification (Level 1 through Level 5) that specifies the degree of machine autonomy a registered AI system is authorised to exercise, the mandatory oversight model that applies, and the maximum decision authority it may exercise without human confirmation. Determined by the ADAE composite score. Managed through the Autonomy Register.

Autonomy Register
The EDS framework's live inventory of all authorised AI deployments. Every AI system must have a registered Autonomy Register entry before it may be deployed to production. The register records each system's Autonomy Level, Criticality Tier, oversight model, accountable executive, and Authority Matrix domain authorisation.

BRC · Board Risk Committee

The Board committee with standing delegated oversight of the Autonomy Budget and the Board Monitoring Dashboard. It receives the quarterly dashboard, approves deployments in the Critical band, and reviews any proposed ceiling change. It cannot amend the Autonomy Budget ceiling or any constitutional principle.

CAIO · Chief AI Officer

The senior executive responsible for AI governance strategy, Authority Matrix domain authorisation at or below Board-approved limits, and compliance with the EDS framework across the organisation. The CAIO owns the Autonomy Register.

CAM · Continuous Admissibility Monitor

The runtime component that detects drift between a system's live operating parameters and the governance profile under which its authority was granted, across seven dimensions. A CAM alert is a governance signal requiring CAIO assessment, not an execution block.

CAPA · Corrective and Preventive Action

The structured remediation plan required after every Level 2 and Level 3 incident. Corrective actions close the immediate control gap; preventive actions address the systemic factors behind it. Every action carries a named owner, a target date and a verification method.

Concurrent Sanction

The requirement that both the Group Board and the Board of a sanction-constrained subsidiary sanction cross-entity machine authority exercised over that subsidiary's operations, with both recorded before deployment. Group sanction alone is necessary but not sufficient where the subsidiary's directors cannot lawfully act in the holding company's interests.

Continuous Admissibility

The principle is that a delegated authority's fitness to bind consequences must be tested continuously in the present, not treated as permanently conferred by the original authorisation. Governed through the Continuous Admissibility Monitor (CAM), the ADAE reassessment obligation, and the structural change trigger.

Criticality Tier

The classification of a system by the consequence of its failure — Critical, High, Moderate or Low. The tier sets the human oversight span-of-control threshold and, combined with the Autonomy Level, determines the mandatory oversight model. It is distinct from the Responsible AI Framework tiers used for fairness assessment.

Cross-Entity System

An AI system that reads from or writes to systems, data, or processes across more than one legal entity in a group. Identified against four criteria, loaded by one ADAE band, and deployable only with Group CAIO written approval and a recorded Group Board sanction.

Delegation Chain Integrity

The condition that every agent in a multi-agent chain is registered and scored, the orchestrator carries the chain depth loading, aggregate authority stays within the orchestrator's ceiling, and every instruction carries verifiable provenance. The chain carries its own criticality tier, set by the consequence of chain failure.

EDS · Enterprise Decision Systems
The governance framework that establishes the constitutional doctrines, governance principles, and operational instruments through which machine decision authority is delegated and controlled.

Foundation Model Characterisation
The record of the foundation model identifier, provider, known failure modes and update channel for every dependent system. It is the reference against which supervisor toolchain declarations are tested for shared failure modes between the overseer's tools and the system overseen.

FRIA · Fundamental Rights Impact Assessment
The assessment required of deployers of high-risk AI systems under the EU AI Act, covering the system's impact on the fundamental rights of affected individuals. Delivered through the EU Alignment Enhancement.

GMI · Governance Maturity Index
The EDS framework's five-level certification scale (Level 1 through Level 5) that measures an organisation's readiness to govern autonomous AI systems. The GMI is a binding constraint on autonomy scaling: no system may operate at an Autonomy Level exceeding the organisation's certified GMI level. GMI Level 2 is the minimum governance readiness gate before any Level 3 system may be deployed; GMI Level 3 before Level 4; GMI Level 4 before Level 5.

Guarded Innovation
The Board's constitutional risk posture toward autonomous AI. Guarded Innovation is the commitment to supporting autonomous AI where it demonstrably improves performance, while enforcing hard limits on uncontrolled autonomy, irreversible actions without circuit-breakers, and unmonitored authority. It is not risk avoidance; it is risk-managed velocity; the constitutional foundation of Certified Operational Velocity.

HAL · Human-above-the-Loop
The minimum oversight model for all systems at Autonomy Levels 4 and 5. HAL supervisors review decision batches at a defined cadence rather than monitoring individual decisions. At those levels HOTL is additionally required for High-tier systems and HITL for Critical-tier systems, as set by the criticality and autonomy intersection in the Operating Manual. HAL supervisors must hold the Critical-tier endorsement within the HOTL Certification Framework. This endorsement constitutes the HAL certification pathway.

HITL · Human-in-the-Loop
The oversight model in which a human must approve each individual action before the system may execute it. Required for Critical-tier systems at Autonomy Levels 4 and 5, and triggered automatically when a supervisor reaches the span-of-control threshold for their tier.

HOTL · Human-on-the-Loop
The EDS oversight model for Autonomy Level 3 systems. A certified HOTL supervisor monitors the decision stream in real or near-real time and can intervene, redirect, or halt the system at any point. HOTL oversight is a constitutional governance obligation; a system operating at Level 3 or above without a currently certified HOTL supervisor breaches governance.

HOTL Certification Framework

The EDS framework's mandatory certification program for all primary and secondary HOTL supervisors. Comprises five sequential components: Foundation, System-Specific Training, Intervention Skills, Automation Bias Training (≥75% scenario identification pass standard), and Assessed Supervision Exercise (minimum four hours). Certification is valid for 12 months; renewal requires an updated automation-bias scenario assessment. The Critical-tier endorsement within this framework constitutes the HAL certification pathway.

JML · Joiner-Mover-Leaver

The identity lifecycle process governing provisioning (Joiner), modification (Mover), and deprovisioning (Leaver) of Non-Human Identities (NHIs) in agentic AI deployments. Adopted by the EDS framework as the primary lifecycle control for all NHIs. Every system lifecycle event triggers an NHI lifecycle review. A model update that materially changes a system's decision logic is a Mover event requiring an NHI privilege scope review before the updated model reaches production.

KRI · Key Risk Indicator

A metric signalling whether a governance control is operating within its Board-approved appetite. Each carries a defined threshold, a Red, Amber, Green status and a mandatory escalation pathway. The EDS Governance Framework specifies 22, and the Responsible AI Framework specifies 5.

Logging and Explainability Standard

The mandatory technical specification governing audit trail completeness, retention obligations and rationale capture for every registered system.

MAM · Maximum Action Magnitude

The single-transaction ceiling on the financial authority or operational impact that a registered autonomous agent may exercise without human confirmation. Set by the CAIO within each Authority Matrix-authorised domain, at or below the domain limit. Technically enforced at the infrastructure level; not policy-declared and not overridable by the agent's own reasoning.

MANDATE · Machine Authority Non-Delegable Autonomous Technology Enforcement

The Board-level instrument through which machine decision authority is formally delegated, explicitly bounded, and continuously monitored. MANDATE is not a software product or a technical standard; it is the highest-order governance act the Board performs in relation to AI, and the constitutional foundation from which all other EDS controls are derived. No EDS product — no policy, standard, operating manual, or assurance framework — can operate without the constitutional foundation MANDATE provides. The three Governance Doctrines (Decision Authority Governance, Governance-Constrained Autonomy, and Delegated Machine Authority) form MANDATE's core and define its scope.

MDAM · Machine Decision Authority Matrix

The Board instrument through which decision authority is formally delegated to autonomous AI systems across defined organisational domains. Specifies which operational domains are authorised for machine decision-making, the maximum decision authority permitted in each domain, the oversight model required, and the aggregate ceiling (Autonomy Budget) on total delegated machine authority. No autonomous system may operate in a domain not covered by a Board-authorised MDAM entry.

NHI · Non-Human Identity

A digital identity assigned to an autonomous AI agent, service account, or automated script that grants it the ability to act within enterprise systems. In the EDS framework, NHIs follow the JML lifecycle framework, must be registered in the NHI Lifecycle Register before production use, and may never operate with privilege scopes exceeding the Maximum Action Magnitude specified in the Authority Matrix domain entry. An unregistered NHI is a zero-tolerance governance breach.

PIR · Post-Incident Review

A structured review conducted within 10 business days of closing a Level 2 or Level 3 incident, assessing detection quality, classification accuracy, response timeliness, and escalation adherence — that is, whether the governance framework performed as designed.

Prohibited Use Register

The Board-adopted register of AI use cases and decision domains that are absolutely prohibited regardless of commercial opportunity, technical feasibility or regulatory absence. No Machine Decision Authority Matrix entry may authorise operation in a prohibited domain.

RAG · Red, Amber, Green

The three-state classification applied to every Key Risk Indicator on the Board Monitoring Dashboard. Green is within appetite; Amber is approaching the threshold and requires a root-cause statement and remediation plan; Red is outside appetite or a zero-tolerance breach and requires out-of-cycle notification to the Board Risk Committee Chair within 24 hours.

SRE · Safety Runtime Environment

The always-on infrastructure layer that evaluates every agent decision against configured thresholds, oversight requirements and authority limits. It does three things a per-decision policy engine cannot: shifts automatically to HITL as span-of-control thresholds are approached, detects anomalies across the decision stream, and fails closed when logging is unavailable. A policy engine decides single actions; the SRE governs the operating mode.

Statutory Accountable Person

The individual to whom accountability for an area of responsibility is assigned under the applicable statutory or regulatory accountability regime. The executive named in a Machine Decision Authority Matrix entry must be that person; an unapproved divergence voids the delegation.

Verification Burden Assessment

The pre-deployment estimate of the time, cognitive effort and judgement a certified supervisor needs to assess one decision log entry, and the resulting load per hour at the system's projected volume. Deployment may not proceed where the load would exceed the cognitive sustainability threshold.

The Governance Conversation

Starts Here.

42 Consulting.AI designed, built, and published the MANDATE Suite. It is the product of a single governing conviction: that AI governance at the depth regulated enterprises require cannot be delivered by a compliance framework. It requires a constitutional architecture — adopted by the Board, enforced by the infrastructure, and continuously tested against current conditions.

"Governance is not the opposite of intelligence. It is what makes intelligence trustworthy."

If your organisation is approaching the governance threshold—where autonomous AI systems make or materially influence decisions that affect customers, financial outcomes, or regulatory standing—the MANDATE Suite provides the architecture to cross it correctly.

Contact

Dr M Maruf Hossain, PhD, GAICD

Chief AI Strategist · Architect, MANDATE *Suite*

42 Consulting.AI

enquire@42consulting.ai

www.42consulting.ai

Academic Foundation

Governance-First AI ·. Board Authority. Architectural Precision. Governed AI at Enterprise Scale.

https://doi.org/10.1007/979-8-8688-3118-8

The MANDATE is constitutional. 15 architectural blueprints make it enforceable.

Apress · 2026

The MANDATE Suite rests on an open-access research record of sixteen papers setting out the theoretical basis for each governance mechanism, published for independent scrutiny and citation.

The full research record is available for independent review at https://zenodo.org/communities/mandate/records

MANDATE Suite · EDS Governance Framework · Responsible AI Framework · AI Management System · EU Alignment Enhancement · Group AI Governance Framework

© 2026 42 Consulting.AI · All rights reserved · PUBLIC