VIANORDIS

Adoption & Measurement Guide

Piloting governed AI actions, and measuring whether the control chain works

Open Review

Version	Version 0.1.4
Status	Open Review
Date	26 August 2026
Author	Yves Frédéric N'Soussoula, Founder & CEO
Publisher	GoSec Cloud OÜ, Tallinn, Estonia
Licence	Creative Commons Attribution 4.0 International
Web	www.vianordis.eu

Open Review. This document is published for criticism. Corrections of fact are applied in the next edition and listed; substantive critique is credited unless the contributor prefers otherwise. Contact: www.vianordis.eu/en/contact · Law as at 26 August 2026. Not legal advice.

Contents

> **Publisher's disclosure.** This guide is published by GoSec Cloud OÜ (Tallinn, Estonia), which develops and operates **Vianordis**, an implementation of the architecture class it describes. We have a commercial interest in the argument. The measurement protocol in §6 is deliberately written so that an organisation can run it, and publish the result, without using Vianordis — including a result that goes against us.
>
> **Status.** This guide is non-normative. It contains no field measurements. Where a figure or a scenario appears, it is illustrative and constructed unless explicitly stated otherwise.

1. Status and how to read this document

1.1 One of three documents

This guide is one of three documents that are intended to be read together and maintained on separate version series:

- ***Position Paper 01 — The AI Control Gap*** (v0.5.4) carries the argument and the regulatory case: what changes when a language model's output becomes a business event, what a control layer does not solve, which existing controls already do part of the job, and how the EU AI Act, NIS2 and the BSI Act, DORA, the GDPR and the Cyber Resilience Act bear on it. It also defines the vocabulary the other two documents use, including the impact tiers and the provenance assurance classes.
- The ***GAIP Candidate Specification*** (0.4) is the normative technical profile: the Governed Action Envelope data model, the mandatory invariants, the processing procedure, canonicalisation and cryptographic profiles, target-adapter assurance modes, and the required negative tests. It defines no conformance scheme and no conformance levels.
- **This guide** (0.1.4) is the practical layer: how to pilot a governed action, how to route an existing estate towards enforcement, how leadership can assess where it stands, and how to measure whether any of it works.

1.2 What this guide assumes

This guide assumes the reader already has two definitions from *Position Paper 01*.

The **impact tiers 0–5** (*Position Paper 01*, §2.4) classify what an action can do, from *inform* at tier 0 to *critical or irreversible* at tier 5. The tier is an attribute of the action, not of the model, and it is value-sensitive: the same payment operation may sit at tier 3 below a threshold and tier 4 above it.

The **provenance assurance classes P0–P4** (*Position Paper 01*, §2.5) classify how trustworthy the facts behind an action are, from *independently verified* at P0 to *mixed, unknown or untraceable* at P4. The class that matters most in practice is **P3 — derived from untrusted content**: a value a model or parser extracted from mail, documents, web content or tool descriptions.

Neither taxonomy is a legal classification. Both are engineering instruments for deciding how much control an action needs and when automation must stop. This guide uses them throughout and does not restate them.

1.3 No field data exists yet

There is no field data behind this guide. We do not have this data. We have not identified published cross-organisational field measurements using comparable definitions as of 26 August 2026; that statement rests on the sources reviewed for this edition and should not be read as proof that no such data exists. We did not run a systematic literature search, and we do not claim one. The scorecard in §6.6 is therefore published with blank values. Those blanks are deliberate. This guide does not contain field measurements and must not manufacture them.

The measurement protocol in §6 should be read as an invitation to gather that data, not as a report of it. It is designed to produce results that can be compared between organisations, that include the outcomes unfavourable to the underlying thesis, and that can be published by an organisation that has never spoken to us.

---

2. A worked example: AI-assisted invoice processing

This example is illustrative and constructed. It is not a customer case study, and no claim is made that the figures or failure modes are drawn from a documented incident.

It is included because it is the most concrete material in the set: it shows what the tiers, the provenance classes and the envelope actually mean in one process, and it ends with the case the architecture does not save.

2.1 The scenario

An organisation wants an AI agent that reads invoices from a mailbox, identifies supplier and invoice data, matches against the purchase order in the ERP, assesses discrepancies, prepares a payment, and executes payment for uncritical cases.

2.2 Typical integration without a shared control layer

The agent holds access to the invoice mailbox, read rights on purchase orders, write rights on creditor master data, access to the payment workflow, and a general service account.

The agent logs its model calls. The ERP logs the posting. The identity system logs the service account sign-in.

Questions that remain open:

- Which employee instructed the agent?
- Was that person entitled to authorise payments for this legal entity?
- Was the agent instructed only to check, or to pay?
- Which model version was used?
- Which data were transmitted to the model provider?
- Which policy version applied?
- Was a discrepancy detected?
- Which specific payment did a human approve?
- Does the executed transaction match the transaction presented?
- **Where did the bank details come from?**

2.3 Execution with Governed Intelligence

Step 1 — Intent. An entitled person or a scheduled business process creates the intent: *"Check invoice 4711 against purchase order 9204 and prepare payment."* The intent is structured against a schema; its origin is marked.

Step 2 — Identity. The agent uses its own identity and instance identity. The instructing person remains attached to the transaction as principal.

Step 3 — Mandate. The mandate permits: read the specific invoice; read the associated purchase order; check the supplier; produce a payment proposal. It does not yet permit payment.

Step 4 — Context. The control plane resolves tenant and legal entity, cost centre, invoice amount, data classification, target system and risk indicators.

Step 5 — Policy. For example: amounts up to 500 euro may be processed automatically on a full three-way match; amounts between 500 and 5,000 euro require approval; **new or changed bank details always require a second approval sourced independently of the invoice document**; data from this invoice may only be sent to a model approved for this data class; discrepancies block automatic payment.

Step 6 — Approval. The entitled person is shown invoice number, supplier, bank details, amount, purchase order, detected discrepancies and the planned action — **each field annotated with its provenance**: which values were extracted by the model from the document, which were read from supplier master data, and which were independently verified. Fields derived solely from the untrusted document are visually distinguished. The approval is bound technically to exactly these values, their provenance, the supplier-master and purchase-order versions used for the decision, and a short expiry.

Step 7 — Execution. The action broker receives a short-lived, transaction-bound right to execute exactly this payment. Immediately before commit it revalidates mandate and policy status and requires the supplier and purchase-order versions to match the approved preconditions. The agent holds no standing payment access.

Step 8 — Evidence. The evidence record links principal, agent, mandate, policy version, model and tools, approval, transaction parameters, parameter provenance, approved and committed precondition state, idempotency key, target receipt and ERP result. Depending on assurance level it is hash-chained, signed and stored in a separate evidence repository.

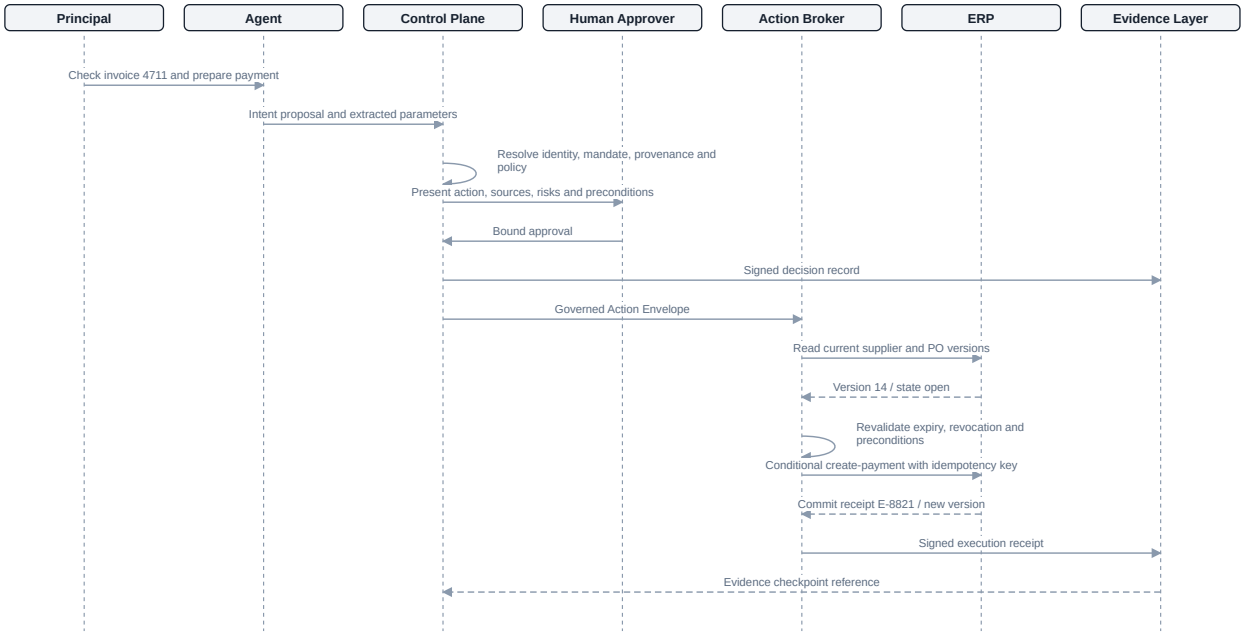

Figure 1. A governed invoice payment, end to end.

The envelope, the idempotency rules and the conditional-commit behaviour shown in this diagram are specified normatively in the *GAIP Candidate Specification*; this section shows only what they look like in one business process.

The diagram abstracts the control plane into one participant. In a real tier 4 deployment it is not one: the *GAIP Candidate Specification*, §5.2, requires the Mandate Authority, the Approval Service, the Action Broker and the evidence signing key custody to sit in distinct trust domains.

2.4 The difference at audit

Without a shared control layer:

> "The service account created a payment at 14:32. The agent log shows a model response a few seconds earlier."

With Governed Intelligence:

> "Agent A, instance A-91, acted on behalf of person B under mandate M-4711. Policy P-27 required approval because of the amount. Person C approved at 14:31 an action bound to amount, payee and invoice. The bank details in that action were extracted by model M v3 from document D (hash H) received from sender S, and were verified against supplier master record V-238 before presentation. The action broker executed exactly this action at 14:32 in the ERP. The ERP confirmed the transaction under reference E-8821. The signed evidence checkpoint shows the event chain has not been altered since."

The second record does not prove that every legal or commercial decision was correct. It makes authority, process, execution and subsequent integrity verification reconstructable — and, critically, it records where each decisive value came from.

2.5 What still fails in this example

Assume the attacker controls the inbound document and the supplier is new, so there is no master record to verify against. The model extracts the attacker's IBAN. Master-data verification returns "no prior record". The approval dialogue shows the IBAN as document-derived and unverified. If the approver approves anyway — a foreseeable outcome under time pressure or poor interface design — the chain executes an attacker-controlled payment with a perfect audit trail.

The controls that would have stopped it are not in the control chain at all: a supplier onboarding process that establishes bank details out of band before the first invoice, and a policy that refuses tier 4 execution where a payee is both new and document-derived. The architecture can *enforce* that policy. It cannot *invent* it.

This is what the input-manipulation limit (*Position Paper 01*, §3.2) means in a concrete case, and it is why we would rather show the failure than only the success. A reader who takes only the success path from §2.3 and §2.4 has taken the wrong half of this section.

2.6 What happens when state changes after approval

Assume the payment is approved at 14:31 against supplier record version 14 and an open purchase order. At 14:31:30 the fraud team blocks the supplier and the ERP advances the record to version 15. The broker attempts execution at 14:32.

A parameter-only design still sees the approved amount and IBAN and proceeds. A precondition-bound design detects the version mismatch — **provided the adapter can act on it**. On `conditional-commit` or `locked-transaction` (GI-6P), the envelope becomes stale, the broker records `precondition_mismatch`, does not call the payment operation, and returns the transaction for policy re-evaluation or human review. On `broker-ledger` (GI-6D), the payment commits and the divergence is found by reconciliation within the declared interval — useful evidence, but the money has moved. Which of those two an organisation gets is determined by the target system's primitives, not by the control plane, and it should be decided and stated before the pilot rather than discovered during it.

If the ERP supports a conditional commit, the same version check can be enforced atomically by the target. If it only supports a read followed by an unconditional write, a race remains between the final read and commit. The adapter must disclose which mode it is operating in. At tier 4 the organisation does not have a free choice here: where the target offers neither a conditional commit nor a lock, the *GAIP Candidate Specification*, §15.2, requires the `broker-ledger` mode — an authoritative broker-side intent ledger with a declared reconciliation interval — and forbids `immediate-reread`, whose residual race is not acceptable at that tier. `immediate-reread` remains available below tier 4. The full set of assurance modes is in §15.1 of that document.

This case is important because it shows the difference between three proofs:

1. **the approver saw and approved these values;**
2. **those values came from these sources;**
3. **the material state that made the action permissible still held when the target committed it.**

A mature evidence record needs all three.

3. An adoption path

Durations are deliberately omitted. Any figure presented without knowledge of identity, target-system and integration maturity would be an estimate disguised as a plan. The sequence below is designed to produce measured scope before schedule commitments.

3.1 Entry principle — no big-bang control-plane migration

The programme unit is an **action type**, not an application and not an AI model. An organisation should not wait until every system is integrated before governing its highest-impact actions. Nor should it route low-impact activity through high-assurance controls that make the platform unusable.

Start with the action map, select the highest combination of impact and feasible control, and move that action through M0 to M4 as defined in §4 of this guide.

3.2 Phase 1 — Make AI actions visible

Inventory the actual AI-assisted actions, not only the models. For each use case, document: business purpose; users and owners; models used; data processed; connected systems; possible actions; impact tier per the tier model (*Position Paper 01*, §2.4); existing approvals; existing logs; third parties.

The result is not a model list but an **AI action map**.

3.3 Phase 2 — Control one bounded process

Choose a clearly bounded pilot: invoice preparation, document approval, support ticket classification, user onboarding, contract drafting, internal publication.

The pilot receives an agent identity, a mandate, bounded tools, a versioned policy, defined approval points, a governed action envelope, an execution receipt and an evidence record.

Choose a pilot at tier 3 or 4. Tier 0–2 pilots often prove integration but not authority binding.

A pilot is not complete when the happy path works. Acceptance criteria should include:

- mandate revoked between approval and execution;
- policy changed after approval;
- target state changed after approval;
- repeated request with the same idempotency key;
- compromised or unavailable evidence producer;
- control-plane outage for every supported tier;
- parameter derived from untrusted input;
- attempted target-system bypass;
- partial commit and failed compensation;
- independent reconstruction by someone outside the implementation team.

The corresponding negative tests are specified in the *GAIP Candidate Specification*; the point here is that they belong in the pilot's acceptance criteria rather than in a later hardening phase.

3.4 Phase 3 — Reuse shared control functions

Standardise the reusable building blocks: identity schema, mandate format, policy interface, approval mechanism, Governed Action Envelope profile, action broker, execution receipt, evidence schema including parameter provenance and preconditions, tier and provenance definitions, and adapters for central systems.

3.5 Phase 4 — Cross provider and application boundaries

Extend the control layer to further models, applications and business units. Subject-matter responsibility stays decentralised while central minimum controls are enforced uniformly.

Sequence by impact tier, not by system. A tier 4 action in a peripheral system matters more than a tier 1 action in a core one.

3.6 Phase 5 — Continuous recertification

Re-evaluate agents and AI processes regularly or on events: new model version, new tool, changed data class, new regulatory requirement, changed corporate policy, security incident, changed business role, mandate expiry.

3.7 Phase 6 — Independent review and community feedback

For critical components, make architecture, threat model, evidence schema and reference policies available for independent review: open review of the position paper and the candidate specification, public architecture decision records, community security reviews, external penetration tests, auditor and industry workshops, published corrections and lessons learned.

3.8 Establish a measurement baseline before expansion

Before the pilot changes the process, sample comparable existing transactions and record:

- whether principal, purpose and policy version can be reconstructed;
- time and staff effort to reconstruct;
- number of systems and people needed;
- proportion of approvals bound to the executed values;
- proportion of parameters with reconstructable provenance;
- existing duplicate, stale-state and exception rates;
- current human approval time;
- current operational and audit cost.

Repeat the measurement after the pilot. Without a baseline the organisation can demonstrate architecture but not value. The definitions to use for these measurements, so that the result is comparable with other organisations' results, are in §6.

3.9 Exit criteria for scaled adoption

Scale only when the pilot has shown:

1. no undeclared or undetected alternative path for the declared high-impact action path — every alternative route inventoried, each with a detective control;
2. for adapters in `conditional-commit` or `locked-transaction` mode, stale approvals are rejected before commit in test and production-like conditions; for adapters in `broker-ledger` mode, injected divergences are detected and escalated within the declared reconciliation interval;
3. an independent reviewer can reconstruct a random sample within the target time budget;
4. evidence verifies after key rotation and simulated component compromise;

5. safe degradation and break-glass behaviour are tested;
6. false blocks, approval delay and exception rates are operationally acceptable;
7. data protection and employee-representation controls are agreed where applicable;
8. the measured benefit justifies the additional operating dependency.

Criterion 8 is the one most often skipped. An organisation that cannot state the measured benefit has bought an operating dependency on the strength of an architecture diagram.

---

4. Brownfield: routing an existing estate

Migration is an adoption problem before it is an architecture problem, which is why the enforcement modes are set out here rather than alongside the reference architecture. They describe how far a given action type has actually moved, and therefore what may honestly be claimed about it.

4.1 Enforcement modes M0–M4

A mature estate cannot place every integration behind a hard enforcement point at once. The migration path should be explicit rather than hidden behind a binary "governed / ungoverned" label.

Mode	Behaviour	What can be claimed
M0 — Inventory	Action is mapped; no runtime telemetry added	Visibility only
M1 — Observe	Existing identity, model, application and target logs are correlated after the fact	Reconstruction coverage can be measured; no prevention claim
M2 — Shadow decision	PDP evaluates the action but cannot block it; differences from actual execution are recorded	Policy readiness and false-block analysis
M3 — Soft enforcement	Unclear or high-risk actions are reduced to draft, warned or routed to manual processing; some bypass remains	Partial runtime control
M4 — Enforced governed path	Agent, application and declared automated execution credentials for the assessed action path are mediated by the PEP; unauthorised envelopes are blocked. **Stale-state prevention is not a property of the mode — it follows from the target adapter's assurance mode**	For the declared action set, with a stated scope and date: unauthorised envelopes cannot commit, and alternative paths are inventoried and monitored. On `conditional-commit` or `locked-transaction`, declared stale-state changes are prevented before commit (GI-6P). On `broker-ledger`, a divergent action **may** commit and is detected within the declared reconciliation interval (GI-6D); at tier 5 `broker-ledger` is not available at all, and the fallback is a manual target-native process. Not an absolute no-bypass claim

Movement between modes should be per action type and impact tier, not per application. A large ERP may have tier 1 actions in M1 and payment actions in M4 simultaneously.

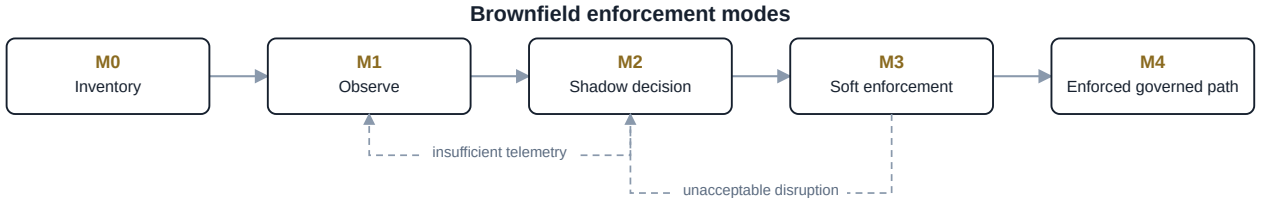

Figure 2. Movement between modes is per action type, not per application.

The two dotted transitions are not failures of the programme. Insufficient telemetry at M2 and unacceptable disruption at M3 are the two most common reasons to step back, and a migration plan that does not allow for them will produce a governed label that the estate does not support.

4.2 Integration patterns

Different target systems require different control points. Four patterns cover most estates:

Pattern A — Credential choke point

Remove standing target credentials from the agent and business application. Only the broker can obtain a short-lived credential. This is the strongest path where the target supports scoped API access.

Pattern B — Network and API mediation

Place the target API behind a gateway, service mesh policy or network path that requires the governed envelope or a derived proof. This can provide M4 enforcement without changing the business application, but only if alternative network paths and credentials are actually removed.

Pattern C — Target-native workflow binding

Use the target system's own approval and transaction mechanisms, but inject the principal, mandate, provenance and decision references into the native object. This is often preferable for ERP and HR systems with mature workflow controls. The integration task is then cross-system binding and evidence correlation, not replacement.

Pattern D — Observe and reconcile

Where no enforcement point is initially possible, correlate actions after the fact, compute reconstruction coverage and compare a shadow policy decision with the actual outcome. This is M1 or M2, not runtime governance, but it creates the evidence needed to justify the next migration step.

The same programme may use all four patterns. What matters is that the assurance mode is declared per action type.

---

5. A maturity model for leadership

A numerical score is useful for showing progress, but dangerous if critical failures can be averaged away. This guide therefore combines a score with mandatory gates.

5.1 How to use this

Score one point per question answered **yes, with evidence a third party could verify**. "We think so" scores zero. Total 0–21.

Then apply the mandatory gates in §5.3. A high numerical score cannot compensate for a direct path around enforcement, self-issued authority or unprotected evidence keys.

5.2 The questions

A. Identity (4 points)

1. Does every production agent have a distinct identity, separate from any shared service account?
2. Can we determine, for a past action, on whose behalf the agent acted?
3. Are individual agent *instances* distinguishable, with parent-child lineage where agents call agents?
4. Can we revoke an agent's authority immediately, and have we tested it during an in-flight process?

B. Authority (4 points)

5. Is the agent's mandate separate from its technical credentials?
6. Are entitlements purpose-bound and time-bound rather than standing?
7. Does the mandate carry a monetary or impact ceiling and an explicit sub-delegation rule?
8. Can a business application or model issue mandates, widen its own scopes or approve its own request? (Score the point if the answer is **no**.)

C. Enforcement (6 points)

9. Are all paths to a tier 4 or 5 target operation outside the enforcement point inventoried — native interfaces, administrative consoles, other integrations, batch jobs, vendor support — with a detective control declared for each? (Score the point if the answer is **yes**.)
10. Do we know which data are transmitted to which model and tool, per data class and transaction?
11. Are policies enforced technically at execution time, not only documented or evaluated in shadow mode?
12. Are human approvals bound to the concrete action, parameters, parameter provenance, material preconditions and expiry — and, for tier 4 and 5, does the approval bind to the artefact the approver's own client rendered (GI-5H) or only to the artefact a central service rendered (GI-5A)? Score the point for a documented binding either way, and record which; GAIP requires GI-5H at tier 5.
13. **Transaction validity** — does the broker reject revoked, replayed and expired envelopes, and is the target operation idempotent or otherwise duplicate-safe? 13a. **State assurance** — does the adapter either prevent a changed-state commit (GI-6P) or detect and escalate it within a policy-approved interval (GI-6D)? Score the point for either, and record which.

D. Evidence (4 points)

14. Can we demonstrate which policy version and decision record applied to a past action?
15. Can someone outside the implementation team reconstruct the chain from instruction to committed system state within the declared time budget?
16. Is that chain cryptographically tamper-evident, including an independently verifiable checkpoint for the highest assurance class used?

17. Are keys, signing services, trusted-time services and the evidence store separated from ordinary business applications and service accounts?

E. Organisation (3 points)

18. Do management, IT, data protection, security and the business share one view of responsibility, and are employee representatives involved where the evidence design engages co-determination rights?
19. Are agents, models, tools and action adapters part of ICT risk management, supply-chain assessment and incident response?
20. Do we know which actions must block, degrade or use break-glass when a control component is unavailable, and have we tested each mode?

Question 20 refers to the safe-degradation matrix in the *GAIP Candidate Specification*, §21.3, and the break-glass profile in §19 of the same document, which define per tier what must block, what may degrade and what an emergency path has to satisfy before it counts as a design rather than a hidden administrator credential.

5.3 Mandatory gates and score caps

Gate	Requirement	Consequence if failed
G1 — Authority separation	Question 8 scores its point (a business application or model **cannot** mint, widen or self-approve a mandate)	Maturity is capped at **Partial**
G2 — Bypass declaration and detection	Question 9 scores its point, for the action set claimed as governed	Maturity is capped at **Partial**
G3 — Transaction binding (tier 4)	Questions 12, 13 and 13a all score. Questions 12 and 13a accept **either** documented property — GI-5A or GI-5H, GI-6P or an explicitly accepted GI-6D exception — because a tier 4 `broker-ledger` deployment legitimately holds GI-6D and cannot prevent every stale commit. The assurance actually held must be recorded either way	Maturity is capped at **Governed in parts**
G3-5 — Transaction binding (tier 5)	For any action in the declared scope at tier 5: **GI-5H** (the approver's own client renders and signs), **GI-6P** (`conditional-commit` or `locked-transaction`), and approvals from **at least two distinct natural-person credentials in distinct roles**, each distinct from requester, beneficiary and each other. GI-5A, GI-6D or a one-of-one quorum do not satisfy this gate. Where they cannot be met, the tier 5 action must not be automated	Maturity is capped at **Governed in parts**, and the tier 5 scope must be declared as not governed
G4 — Independent reconstruction	Question 15 = yes	Maturity is capped at **Governed in parts**
G5 — Evidence and key separation	Questions 16 and 17 = yes for tier 4/5 evidence	Maturity is capped at **Governed in parts**
G6 — Safe failure	Question 20 = yes	Maturity is capped at **Governed in parts**

These gates apply only to the action scope being assessed. An organisation may be M4 and "Governed" for invoice payment while remaining M1 and "Partial" for infrastructure changes. A single enterprise-wide label is therefore less informative than an action-level maturity map.

5.4 Score bands

Score	Band	What it means
0–5	**Unmapped**	AI actions exist but are not governed as actions. The first task is an action map, not a platform decision.
6–10	**Partial**	Controls exist per system. They do not compose. Reconstruction is possible but slow, manual or dependent on the implementation team.

Score	Band	What it means
11–16	**Governed in parts**	A working control chain exists for some processes. The principal risk is uneven coverage, stale-state handling and ungoverned high-tier actions.
17–21	**Governed**	The chain holds for the declared action scope, all gates pass and the result is independently demonstrable. Residual risks remain, especially manipulated input, privileged-domain compromise and operational dependency.

A score of 21 does not mean safe or compliant. It means the architecture has made its remaining failures more explicit and testable.

5.5 Four measurements that test the thesis

The central thesis of *Position Paper 01* is presented as falsifiable. These four measurements test different parts of it:

1. **Reconstruction coverage** — the proportion of AI-initiated transactions for which the human or organisational principal, purpose, policy version, decisive parameter provenance and execution result can be reconstructed from existing records.
2. **Time to reconstruct** — median and p95 elapsed time to reconstruct a complete action chain for a randomly selected transaction, measured by someone who did not build the system.
3. **Approval and state-binding rate** — the proportion of tier 4 and 5 approvals bound to the executed parameters, their provenance and the material preconditions. Report the state result in two separate figures, never one: the **preventive stale-rejection rate** for adapters that hold GI-6P, and the **detective divergence-detection rate with median and p95 detection latency** for adapters that hold only GI-6D. A single blended number conceals which actions were prevented and which were merely noticed.
4. **Governance-constrained initiatives** — the proportion of AI initiatives stopped, deferred, reduced in autonomy or materially delayed for reasons recorded as authority, accountability, auditability or control integration rather than model quality, cost or business case.

Measurements 1 to 3 test whether the operational gap exists. Measurement 4 tests whether it is a binding constraint on scaling AI. The gap may be real while some other constraint matters more.

The thesis is weakened if organisations with mature conventional controls score highly on measurements 1 to 3 without a dedicated control-plane pattern, or if they fail to scale while measurement 4 remains near zero. Section 6 of this guide defines the sampling and reporting method. We invite organisations to publish comparable results, with or without Vianordis.

5.6 Operational measures for a production deployment

The thesis metrics are not sufficient to operate the system. A production service should also track:

Measure	Why it matters
Governed-action coverage by tier	Shows whether the highest-impact actions are actually on the controlled path
Bypass exceptions	Reveals residual privileged routes and temporary waivers
Policy-decision p50/p95/p99	Quantifies machine-path overhead
Human approval p50/p95 and queue age	Shows where business latency accumulates
False-block and false-escalation rate	Detects policy that is technically correct but operationally unusable
Preventive stale-rejection rate (GI-6P adapters)	Measures real TOCTOU exposure and the value of precondition binding where the target can enforce it
Detective divergence-detection rate and latency (GI-6D adapters)	Measures how much of that exposure is only found afterwards, and how long the window is
Break-glass frequency and review closure	Detects whether emergency access has become routine
Evidence verification success	Confirms that signatures, checkpoints, keys and retained validation material still work

Measure	Why it matters
Partial-commit and compensation-failure rate	Exposes target adapters that cannot provide the claimed transaction assurance
Cost per governed action	Makes the control architecture an economic decision rather than a compliance abstraction

6. Measurement protocol and pilot scorecard

6.1 Purpose

This section defines a minimum protocol for testing the central thesis and the operational value of a runtime authority pattern. It is designed so that an organisation can publish results without using Vianordis.

6.2 Unit of analysis

The unit is a completed, denied, stale, partially completed or indeterminate **AI-initiated or AI-assisted action transaction**. Purely informational model calls may be sampled separately but should not dominate a study intended to measure execution governance.

Each study must define:

- action types included;
- impact tiers and provenance classes;
- systems and business units;
- observation period;
- whether the path is M0, M1, M2, M3 or M4 as defined in §4.1;
- whether the sample is pre-pilot, post-pilot or both.

6.3 Sampling

For each action type, draw a random or systematically sampled set large enough to include normal, denied, exceptional and failed transactions. Convenience samples selected by the implementation team are not sufficient for an independent reconstruction measure.

The reviewer measuring reconstruction time should not have built the integration and should receive only the access and documentation that an internal auditor or incident responder would normally receive.

Record exclusions and missing transactions. An unexplained exclusion is a failed reconstruction, not an absent denominator.

6.4 Thesis metrics

6.4.1 Reconstruction coverage

Numerator: sampled transactions for which the reviewer reconstructs principal, actor, purpose, mandate, policy version, decisive parameter provenance, approval status and execution outcome.

Denominator: all sampled transactions, including denied, failed and indeterminate outcomes.

Publish both full reconstruction and partial reconstruction by missing field.

6.4.2 Time to reconstruct

Start when the reviewer receives the transaction identifier or business reference. Stop when the reviewer produces a complete evidence package or declares the reconstruction incomplete.

Publish median, p95, maximum, staff minutes and number of systems or people consulted.

6.4.3 Approval and state-binding rate

For tier 4 and 5 sampled actions report:

- proportion with approval bound to executed parameters;
- proportion also carrying parameter provenance;
- proportion also carrying material preconditions;
- proportion with commit-time state verification;
- number and rate of envelopes rejected on **expiry**, reported separately from state freshness;
- number and rate of approvals rejected before commit because a **precondition changed**, for GI-6P adapters;
- number and rate of divergences **detected after commit**, with median and p95 detection latency, for GI-6D adapters;
- number of approvals executed after expiry or mismatch.

Commit-time state validation and the adapter assurance modes that determine how strong it can be are defined in the *GAIP Candidate Specification*; a study should report which mode each sampled action type actually used. At tier 4 and 5 the comparison that matters is `conditional-commit` or `locked-transaction` against `broker-ledger` — a prevented commit and a commit detected afterwards are not the same claim, and blending them produces a number that means nothing. (`immediate-reread` is not available at those tiers.)

6.4.4 Governance-constrained initiatives

For the relevant planning period, classify AI initiatives stopped, delayed, reduced in autonomy or materially rescoped. Record the primary and secondary reasons using at least:

- model quality;
- business case;
- cost;
- data availability;
- data protection;
- cybersecurity;
- authority/accountability;
- auditability/evidence;
- integration complexity;
- workforce or co-determination concerns;
- supplier or sovereignty risk.

The category should be assigned from decision records, not retrospective intuition.

6.5 Operational metrics

Metric	Definition
Machine-path latency	Identity through evidence emission, excluding human and queue time; p50/p95/p99
Approval latency	Presentation to decision; p50/p95 and oldest queue age
False block	Action later determined permissible but blocked by control logic
False permit	Action permitted despite a policy or mandate condition that should have blocked it
Expired-envelope rejection	Envelopes rejected because the envelope or approval expired. A time-validity control, **not** state freshness — report it separately
Preventive stale rejection	Envelopes rejected before commit because a material precondition changed (GI-6P)
Detective divergence	Actions committed under changed state and found by reconciliation, with detection latency (GI-6D)
Bypass exception	Action completed through a path outside declared enforcement

Metric	Definition
Replay prevented	Duplicate attempt rejected or mapped to the original outcome
Partial commit	Multi-step action not fully completed atomically
Compensation failure	Partial action for which declared compensation did not restore the intended state
Break-glass use	Count, purpose, approvers, duration and review closure
Evidence verification success	Percentage of sampled records whose signatures, timestamps and checkpoints validate
Cost per governed action	Allocated service, storage, approval and operating cost divided by governed actions

6.6 Pilot scorecard

A pilot report should include:

Area	Before	After	Target	Result
Reconstruction coverage	—	—	Defined by organisation	Pass / fail
Median reconstruction time	—	—	—	—
p95 machine-path latency	—	—	—	—
Median approval latency	—	—	—	—
Approval parameter-binding rate	—	—	100% for declared tier 4/5 scope	—
Precondition-binding rate	—	—	100% for material preconditions	—
Expired envelopes rejected	—	—	100% of injected test cases	—
Precondition-stale approvals rejected before commit (GI-6P adapters)	—	—	100% of injected test cases	—
Divergences detected after commit (GI-6D adapters)	—	—	100% of injected test cases, within the declared reconciliation interval	—
Median / p95 divergence detection latency (GI-6D adapters)	—	—	Within the declared reconciliation interval	—
Replay tests passed	—	—	100%	—
Undeclared or undetected alternative paths	—	—	0 for the declared M4 action scope, with the inventory and reconciliation interval stated	—
Evidence verification success	—	—	100% of sample	—
False-block rate	—	—	Organisation-defined	—
Partial/indeterminate outcomes	—	—	Organisation-defined and reviewed	—
Break-glass exercises reviewed	—	—	100%	—
Cost per governed action	—	—	Organisation-defined	—

Blank values are deliberate. This guide does not contain field measurements and must not manufacture them.

6.7 Publication template

A published result should state:

1. organisation type and approximate scale, without disclosing sensitive identity where necessary;
2. action scope, tiers and provenance classes;
3. enforcement modes and target-adapter assurance modes;

4. sample method and size;
5. metric definitions and any deviations from this protocol;
6. results with denominators;
7. known bypasses, exclusions and failed tests;
8. who performed the review and whether they were independent of the implementation;
9. date and software/profile versions;
10. conflicts of interest and vendor involvement.

6.8 Interpretation

A result showing excellent reconstruction without a dedicated control plane is evidence against a product-centric reading of the position paper and in favour of the architecture being assembled from existing controls. That is a valid outcome.

A result showing that governance rarely constrains AI adoption weakens the commercial urgency claim even if the control gap exists technically. That is also a valid outcome.

The objective is not to produce a benchmark that Vianordis wins. It is to determine whether the problem is material, which controls solve it, and at what operational cost.

---

7. An invitation

The three documents in this set make a claim that can be checked. *Position Paper 01* states the hypothesis: that materially fewer organisations can demonstrate authority, provenance and state binding for a specific AI-initiated action than can demonstrate control over which model was used. The *GAIP Candidate Specification* states what an implementation meeting the listed profile requirements would have to do. This guide states how an organisation would find out whether either is worth anything in its own estate.

None of that is settled by publishing more architecture. It is settled by measurements taken in real organisations, by reviewers who did not build the system, and reported with their denominators, their exclusions and their failed tests intact.

We therefore invite organisations to run the protocol in §6 and publish the result — with or without Vianordis, and whether or not the result supports us. A published finding that mature conventional controls already deliver high reconstruction coverage, or that governance is not the binding constraint on AI adoption, is more useful to this discussion than another paper that assumes the opposite. The contact route for corrections, counter-evidence and comparable results is in *Position Paper 01*, and at the canonical address above.

Vianordis Adoption & Measurement Guide, version 0.1.4 — Open Review. Derived from *Vianordis Position Paper 01 — The AI Control Gap*, v0.4, Chapters 8–10, §7.13 and Annex H. Published by GoSec Cloud OÜ, 9 Järvevana tee, 11314 Tallinn, Estonia, under CC BY 4.0.