AGENTIC RISK ARCHITECTURE FRAMEWORK · REFERENCE DOCUMENT

The Six ARAF Dimensions

The governance assessment framework for autonomous system deployments — how ARAF classifies, measures, and benchmarks governance posture across the six dimensions that determine institutional readiness.

PUBLISHED BY

Institute for Autonomous Governance Pty Ltd

DATE

March 2026

VERSION

1.1 — D6 two-layer evidentiary model

REFERENCE

araf-standard.org

OVERVIEW

Six dimensions. One composite score.

D1 1

Autonomy Gradient

What decisions does the system make without human authorisation for each?

D2 2

Data Sensitivity Exposure

What data does the system process, and where does it come from?

D3 3

Contract Infrastructure

What legal architecture governs the deployment?

D4	D5	D6
4	5	6
Liability Architecture	**Commercial Leverage**	**Adaptive Stability**
How is liability for the system's decisions structured and governed?	How commercially dependent is the organisation on this system?	Can governance be maintained as the system and its deployment evolve?

GOVERNANCE BENCHMARK INDEX (GBI) — CERTIFICATION TIERS		
ARAF ASSESSED	ARAF COMPLIANT	ARAF CERTIFIED
Any score	**GBI ≤ 2.50**	**GBI ≤ 1.75**
Assessment completed. Dimensional profile available for institutional use.	Governance posture meets institutional baseline across all six dimensions.	Governance posture meets the highest institutional confidence threshold.

Lower GBI scores indicate stronger governance. Scores are produced on a 1.0–5.0 scale. Three structural multipliers — Systemic Escalation, Infrastructure Collapse, Leverage Collapse — can compound scores where dimensional combinations exceed activation thresholds.

1
D1

Autonomy Gradient

What decisions does the system make without requiring human authorisation for each individual decision?

The Autonomy Gradient is the foundational dimension. It establishes the governance burden that all other dimensions must address. A system operating at low autonomy with weak contract infrastructure faces a different risk profile than a system operating at high autonomy with the same weakness.

WHAT IT ASSESSES

- Operational autonomy level and the categories of action the system takes without per-decision human authorisation
- Commitment authority — the maximum financial or operational commitment the system can make autonomously
- Exception handling — what happens when the system encounters decisions outside its trained parameters
- Scope boundary enforcement — how operational scope boundaries are enforced technically and contractually
- Human oversight adequacy proportionate to the autonomy level at which the system operates

INSTITUTIONAL IMPLICATION

D1 is foundational. Its score determines the governance burden all other dimensions must address. A system operating at high autonomy with weak liability architecture or contract infrastructure creates compounding exposure that the multiplier logic is designed to capture.

> **Machine-enforceable.** D1 is one of three dimensions that can be directly evidenced by runtime enforcement infrastructure. Deterministic enforcement logs — FAIL_CLOSED behaviour, HARD_BLOCK on scope violations, PASS_CONFIRMED on benign inputs — provide Tier 1 D1 evidence. Human oversight structure requires institutional-layer evidence in addition.

2
D2

Data Sensitivity Exposure

What data does the system process, and where does that data come from?

Data Sensitivity Exposure assesses two distinct risk categories that require separate governance treatment. Operational data creates immediate exposure at the point of processing. Training data provenance creates latent

exposure that is carried into every decision the system makes from deployment forward.

WHAT IT ASSESSES

- Operational data categories — sensitivity classification, access governance, handling controls
- Training data provenance — legal basis, IP clearance, consent frameworks for all training and fine-tuning data
- Foundation model provider representations — what the model provider asserts about training data governance
- Retraining cycles as new provenance governance triggers — each retraining event creates new potential liability
- RAG architecture provenance — where deployed, inference-time data retrieval creates additional exposure

INSTITUTIONAL IMPLICATION

For boards and legal counsel, D2 maps the latent liability created before deployment. The EU Product Liability Directive's lifecycle strict liability approach makes each retraining event a new potential liability trigger. A provenance gap identified post-deployment cannot be remediated retroactively — it is embedded in the system's operational history.

> **Partial runtime evidence.** Enforcement logs evidence operational data handling at the point of evaluation. Training data provenance requires separate documentation: foundation model provider representations, fine-tuning data legal basis, and IP clearance records. These are institutional-layer artefacts that enforcement infrastructure does not produce.

3
D3

Contract Infrastructure

What legal architecture governs the deployment?

Contract Infrastructure assesses the legal architecture through which liability is allocated, obligations are defined, and accountability is assigned across the deployment chain. It is the dimension most directly connected to the four-link accountability chain: design, deployment, operational, and outcome.

WHAT IT ASSESSES

- Customer agreements — AI-specific provisions, scope limitations, liability carve-outs
- Vendor agreements — model and infrastructure provider contracts, SLAs, indemnification
- Data processing agreements — legal basis for data use across the deployment chain
- Liability provisions governing autonomous action consequences (AE3)
- Contractual allocation of accountability across the Decision Supply Chain

INSTITUTIONAL IMPLICATION

D3 is a foundational liability dimension. The contractual layer is the mechanism through which provenance risk, AE3 exposure, and operational liability are allocated. An organisation without adequate contract infrastructure has no legal architecture through which to manage adverse autonomous decisions. Enforcement infrastructure — however strong — cannot substitute for the contractual layer.

> **Institutional-layer dimension.** No enforcement log satisfies D3. Contract infrastructure is assessed through legal document review. Its absence is not visible in enforcement telemetry; it only becomes visible when an adverse outcome occurs and the contractual allocation of liability is interrogated.

4
D4

Liability Architecture

How is liability for the system's autonomous decisions structured, documented, and governed?

Liability Architecture assesses how liability for autonomous decision consequences is structured. The central concept is AE3 — autonomous action consequences: the outcomes produced by autonomous decisions without per-step human authorisation. AE3 consequences accumulate at operational speed, creating liability exposure that traditional liability frameworks were not designed to absorb.

WHAT IT ASSESSES

- AE3 liability structure — how liability for autonomous decisions is defined and capped
- Carve-out provisions for specific autonomous decision categories
- Insurance coverage — whether AE3 exposure is adequately covered under current policies
- Liability allocation across the accountability chain: design, deployment, operational, outcome
- Governance of the liability structure — who reviews it, at what intervals, triggered by what events

INSTITUTIONAL IMPLICATION

For insurers, D4 is the coverage architecture dimension. It determines what the policy must cover, what can be excluded, and what the exposure scale is. A system generating high AE3 volume with inadequate liability architecture creates insurance exposure that current AI policies may not price adequately.

> **Enforcement evidence is the precondition, not the structure.** Tier 1 enforcement logs demonstrate that governance controls operated at the time of each decision — the evidentiary foundation that liability proceedings require. They do not establish that the liability structure itself is adequate. The structure requires independent assessment.

5
D5

Commercial Leverage

How commercially dependent is the organisation on this system, and how does that dependency constrain governance remediation?

Commercial Leverage assesses the degree to which the organisation is operationally and commercially dependent on the system. High commercial leverage limits governance remediation options: the disruption required to fix a governance gap may exceed what the organisation is willing to accept, making the gap structural rather than contingent.

WHAT IT ASSESSES

- Operational dependency — what proportion of core operations depend on the system
- Revenue concentration — direct or indirect revenue tied to system availability and performance
- Remediation commercial resistance — the commercial cost of governance-driven remediation
- Technology lock-in — switching cost, portability constraints, vendor dependency
- Customer relationship embedding — whether customer commitments depend on the system's continued operation

INSTITUTIONAL IMPLICATION

For investors, D5 is a structural risk signal. A high D5 score alongside a high D3 or D4 score indicates a governance deficit that the organisation's commercial structure is working against correcting. Remediation cost and timeline are determined not only by the size of the governance deficit but by the commercial disruption remediation requires. Commercial leverage is the dimension that most directly explains why governance gaps, once opened, tend to remain open.

6
D6

Adaptive Stability

Can governance be maintained as the system and its deployment evolve?

Adaptive Stability assesses the organisation's capacity to maintain adequate governance as the system evolves. An organisation that has adequate governance at the point of initial deployment may have inadequate governance six months later if its governance architecture does not evolve with the system. A system that generates extensive performance metrics but no governance records has monitoring without accountability.

WHAT IT ASSESSES

- Governance maintenance processes and formal reassessment triggers
- Change governance — how updates to the system and its rule sets are authorised
- Monitoring architecture and behavioural drift detection
- MKP version lineage and rule evolution governance
- AIOC authority over governance rule set changes

The Two-Layer Evidentiary Model

D6 evidence adequacy requires two components that serve distinct functions for distinct audiences. These components are assessed separately and should be maintained separately. Collapsing them into a single record would contaminate the evidentiary integrity of both.

COMPONENT 1 — RUNTIME ENFORCEMENT LOG

RECORDS

MKP identifier governing each decision; enforcement outcome; drift score per evaluation

PRIMARY AUDIENCE

Insurer or regulator interrogating a specific enforcement decision

EVIDENCE TIER

COMPONENT 2 — MKP REGISTRY

RECORDS

Version graph; update authorisation chain; rule change history; change timestamps

PRIMARY AUDIENCE

Assessor evaluating rule adequacy and evolution over time

EVIDENCE TIER

`Tier 1` — if infrastructure-generated and tamper-evident

WHAT IT SATISFIES

Rule attribution: which rule set governed this decision at this time

`Tier 1` if append-only and tamper-evident; `Tier 2` if documented institutional record

WHAT IT SATISFIES

Rule evolution: how the governing rule set changed, when, and who authorised it

> **Separation is the correct architecture.** An insurer or regulator interrogating a specific enforcement decision needs the runtime log without the noise of version history. An assessor evaluating governance adequacy over time needs the registry without the volume of per-decision logs. The separation is not just cleaner audit structure — it is the correct evidentiary architecture for two distinct institutional audiences.

The Traversal Path — D6 Closing Criterion

The runtime log records the MKP identifier. The registry records the version history of that identifier. The D6 closing criterion is whether a traversal path exists between the two: an assessor must be able to move from any runtime decision to the full governance history of the rule that governed it.

D6 Evidence Criterion	Assessment
Runtime log records MKP identifier per decision	Verifiable from enforcement log. Required for rule attribution.
MKP registry exists and is maintained	Requires registry access. Assessor verifies tamper-evidence and append-only architecture.
Traversal path: MKP identifier to version graph	**The closing criterion.** If absent, the two records exist in isolation and rule lineage cannot be reconstructed from enforcement evidence alone.

> *A deterministic system enforcing the wrong rules is deterministically wrong.* Rule attribution proves the rules were applied. Independent assessment determines whether the rules were adequate for the risk profile. That is the layer ARAF occupies.

INSTITUTIONAL IMPLICATION

For insurers, D6 is the ongoing governance maintenance signal across the coverage period. An ARAF assessment captures governance posture at a point in time. D6 assesses whether the organisation has the infrastructure to maintain that posture between assessments. A high D6 score means the governance posture priced at policy inception may not be the posture operating at the time of a claim.

Organisations deploying enforcement infrastructure with a functioning MKP registry and traversal path are structurally positioned for stronger D6 scores: the evidence is contemporaneous and infrastructure-generated rather than reconstructed.

COMPOUND RISK

The Multiplier Logic

The six dimensions are not independent. Weaknesses in one dimension compound weaknesses in others. ARAF incorporates three structural multipliers that capture this compounding effect. Individual dimensional weaknesses are manageable. Compound weaknesses create systemic exposure because compounding reduces remediation options and increases the cost of each option.

MULTIPLIER 1

Systemic Escalation

```
D1 score ≥ 4
AND D4 score ≥ 4
```

High autonomy combined with

MULTIPLIER 2

Infrastructure Collapse

```
D3 score ≥ 4
AND D1 score ≥ 3
```

Significant autonomy combined with

MULTIPLIER 3

Leverage Collapse

```
D5 score ≥ 4
AND D4 score ≥ 3
```

High commercial dependency with

inadequate liability architecture. AE3 volume is high and no liability framework is designed to absorb the consequences. The most severe compound exposure.

inadequate contract infrastructure. Liability allocation for autonomous decisions at the execution boundary is ungoverned contractually.

inadequate liability architecture. The system is structurally resistant to remediation because commercial disruption required to fix the liability architecture exceeds what the organisation will accept.

GBI scoring direction: higher numbers represent greater governance risk, not stronger performance. A score of 4 on any dimension indicates high governance risk on that dimension. See GBI Methodology for the complete scoring scale and multiplier calculation.

INSTITUTIONAL USE

The dimensional profile serves each audience differently

A composite GBI score indicates whether overall governance posture sits above or below an institutional threshold. The dimensional profile shows where the risk sits: whether the score reflects consistent moderate governance across all dimensions, or concentrated weakness on two with strong performance on four.

FOR BOARDS

Identifies which governance questions require immediate attention and which dimensions are producing the multiplier exposure the board must address.

FOR INSURERS

D4 weakness determines AE3 coverage requirements. D1 weakness determines commitment authority exposure. D6 weakness determines

ongoing monitoring conditions the insurer should require.

FOR INVESTORS

Distinguishes structural risks (D3 and D4, requiring contractual and architectural remediation) from operational risks (D1 and D6, responding to process and monitoring improvements). Determines remediation timelines and costs.

FOR REGULATORS

Provides a comparable, independently verified governance posture classification across the regulated population — aligned to EU AI Act, NIST AI RMF, ISO 42001, and APRA CPS 230 requirements.