Adaptive Operating System (AOS) · The Governance Intelligence Tests

Document: governance-intelligence-tests · Source revision or record: Governance Intelligence Tests, unnumbered, original source-manifest digest matches

AOS Specifications v1.0 · Edition dated 3 November 2026

Original architecture and specifications: Mark Sweeny, Governance Intelligence. © 2026 Mark Sweeny.

AOS-authored material is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/

Preserve third-party credits. Referenced external works remain subject to their own terms.

Canonical collection: https://www.governanceintelligence.co.uk/aos

Publication copy with editorial notices, public references and disclosure edits where applicable.

The Governance Intelligence Tests

A Mode 1 / Mode 2 Framework for Governance Quality in the AI - Native Organisation

Introduction

Mode 1 oversight is the operational intelligence function — sensing, monitoring, learning, responding. It asks whether the governance system has the architecture to see clearly: the data infrastructure, the risk intelligence, the compliance assurance, the performance picture. These are, in principle, progressively automatable. AI substantially enhances a board's capacity to perform Mode 1 functions, and a well - designed AI - native organisation should be able to produce evidence against most Mode 1 tests continuously rather than periodically.

Mode 2 oversight is the ethical intelligence function — adjudicating between incommensurable values, maintaining constitutional authority, earning and sustaining legitimacy with those whose lives the organisation shapes. These are not automatable. AI can surface trade - offs and model consequences, but the judgement about which good should yield to which other good — between safety and autonomy, efficiency and dignity, individual need and collective resource — remains irreducibly human work. In an AI - native organisation, the board's Mode 2 function does not diminish; it sharpens, because the consequences of those judgements propagate faster and further than previous generations of governors had to reckon with.

The tests that follow reflect this architecture. The Mode 1 tests are assessable largely through documentation, systems and observable behaviour — a sufficiently capable reviewer, or a sufficiently well - designed platform, can produce a reliable picture. The Mode 2 tests require something different: a human being in the room, able to read what the minutes do not say, to notice when dissent is performed rather than felt, to distinguish an organisation that has encoded its values from one that has learned to describe its optimisation targets in the language of values. That distinction is where experienced consultancy and regulatory judgement will always sit — and where it becomes more valuable, not less, as Mode 1 assurance grows cheaper and more continuous.

These tests also serve a second purpose within the Holding the Line series. The proposals set out in Volume Four — the Adaptive Operating System, the Five Intelligences and a Conscience, and the Governance Hub — are presented not merely as design ideas but as a set of claims about what good governance looks like when built from the ground up. The Governance Intelligence Tests are the quality standard against which those claims should be held. If the architecture proposed in Volume

Four cannot produce an organisation that would pass its own tests, the architecture is insufficient. The tests are not decoration. They are the accountability mechanism the book applies to itself.

Mode 1: The Operational Intelligence Tests

These tests concern the board's capacity to sense, monitor, learn and respond. A board that fails them is ungoverned in the technical sense — it cannot see clearly enough to act reliably. In a mature AI - native organisation, most should be continuously verifiable rather than periodically inspected.

M1.1 Strategic Coherence

The board operates from a clearly articulated strategy that is continuously tested against the organisation's mission and values. The strategy is not a document produced annually but a living framework that updates as the operating environment changes, with the board able to see and explain the current state of that framework at any time.

M1.2 Sensing Architecture

The organisation has the data infrastructure to provide the board with continuous, accurate intelligence about operational performance, resident experience, asset condition and financial health. The board receives signal, not summaries assembled by executives who have already decided what the news means.

M1.3 Risk Intelligence

The risk management framework is dynamic. Risks are identified and escalated based on emerging patterns, not just crystallised events. The board can see which risks are approaching threshold and which indicators are trending in concerning directions — before they become problems requiring reactive management.

M1.4 Composition and Capability

The board's size, skills and composition are appropriate to the demands of AI - native oversight. This includes sufficient technical literacy to interrogate algorithmic outputs, sufficient operational experience to contextualise data, and the cognitive diversity to prevent premature convergence on system - generated recommendations.

M1.5 Performance Legibility

The board has a clear, granular and current picture of how the organisation is performing — both financially and operationally — and understands not just where it is but where it is heading and what the range of plausible futures looks like.

M1.6 Compliance Assurance

The organisation can demonstrate compliance with all regulatory and statutory obligations. In an AI - native system this is a structured comparison — declared controls against documented reality — and should require minimal human effort to produce.

M1.7 Governance Documentation

The documentation of governance arrangements is accurate, accessible and current. Delegations are clear and free of drift. Policies are consistent with practice. Minutes reflect reasoning, not just conclusions.

M1.8 Accountability Architecture

The board can demonstrate clear accountability to its key stakeholders — residents, funders, regulators — and has mechanisms through which stakeholder feedback demonstrably influences decisions. The feedback loop closes.

M1.9 Value for Money

The board can demonstrate value for money in the achievement of its strategic objectives, with the basis of assessment transparent and contestable rather than self - reported.

M1.10 Human - AI Interaction Norms

The organisation has clear, documented norms governing how AI systems inform decisions — including escalation thresholds, override protocols, and the conditions under which algorithmic recommendations require human adjudication before action is taken. These norms are followed in practice.

M1.11 Learning Metabolism

The organisation learns from experience in a systematic rather than incidental way. Decisions are reviewed against outcomes. Assumptions are tested retrospectively. Near - misses are treated as intelligence rather than incidents to manage. The board can show how its understanding has changed over a given period.

M1.12 Continuous Self - Assessment

The board does not rely solely on periodic external review to understand its own effectiveness. It has mechanisms for continuous self - assessment — feedback on decision quality, culture health indicators, directorial development — that make the triennial review an integration point rather than the only diagnostic moment.

M1.13 Emergent System Behaviour

The organisation has governance mechanisms designed to detect emergent behaviours produced by the interaction of intelligence layers — behaviours that no individual layer would produce alone and that may not be visible through component - level monitoring. The board receives assurance about the system as a whole, not only about its parts. Where emergent behaviours are detected that were not anticipated in the architecture design, there is a defined process for governance review and response.

M1.14 Golden Thread Integrity

The organisation maintains the Golden Thread of building safety information required under the Building Safety Act 2022 as a living governance instrument rather than a documentation exercise. The board has a named accountability for the accuracy and completeness of building safety information across the portfolio. The information is current, structured, and accessible — demonstrable to a resident or regulator without prior preparation. The board receives regular

assurance on Golden Thread integrity as a distinct item, not as a sub - point within general compliance reporting.

Mode 2: The Ethical Intelligence Tests

These tests concern the board's capacity to adjudicate between incommensurable values, maintain its constitutional function under pressure, and remain genuinely accountable to those whose lives it shapes. They require observing the system under conditions that reveal whether stated commitments hold when tested. A board that fails them may be technically compliant and operationally competent while having quietly ceased to govern in any meaningful sense.

M2.1 Values Architecture

The organisation's values are encoded in its governance machinery, not merely declared in its communications. Decision thresholds, escalation routes, AI interaction norms and performance frameworks reflect stated commitments in their structure. Where values and efficiency conflict, the architecture creates visible tension rather than allowing values to be quietly overridden.

M2.2 Coherence Under Pressure

The board's stated values remain visible in its decisions when circumstances are difficult — when financial pressure mounts, when regulatory challenge arrives, when efficiency and dignity conflict. The board can point to decisions where it chose the harder path because it was the right one, and can explain why.

M2.3 Incommensurable Trade - off Capacity

The board has a demonstrated capacity to adjudicate between goods that cannot be simultaneously optimised — safety and autonomy, individual need and collective resource, short - term performance and long - term stewardship. These deliberations are visible in the record, not smoothed away in minute - taking.

M2.4 Epistemic Brake Function

The board has identifiable mechanisms — cultural, procedural or structural — that slow or stop algorithmic recommendations when ethical concerns are present but not yet fully articulated. Directors can describe occasions when they felt unease about a technically sound recommendation and the governance system treated that unease as worth investigating.

M2.5 Directorial Independence

Individual directors exercise genuine independent judgement. The board is not captured by executive framing, algorithmic authority or chair dominance. There is evidence of substantive dissent that influenced outcomes, and the culture treats such dissent as contribution rather than obstruction.

M2.6 Social Legitimacy

Those affected by the board's decisions — primarily residents — recognise the organisation as worthy of the authority it exercises. This is not measured through satisfaction surveys alone. The

board asks whether residents believe their voices change outcomes, whether they engage when genuinely invited, and whether they extend the benefit of the doubt when mistakes occur.

M2.7 No - Exit Obligation

The board explicitly recognises and acts on the moral obligation created by the no - exit condition. It does not mistake low complaint volumes for satisfaction, passive acceptance for consent, or disengagement for approval. It has designed its accountability architecture around the absence of market discipline rather than in spite of it. The board applies the no - exit condition as a standing design filter on resident experience architecture — testing each element of how resident experience is sensed, measured, and responded to against the question of whether it would generate genuine understanding in a context where exit signals are unavailable.

M2.8 Radical Transparency

The reasoning behind significant decisions is legible to those affected by them — not in curated summaries but in a form that allows genuine interrogation. Residents can ask why a decision was made, how their circumstances were weighted, and what the board worried about. The transparency is built in as a right, not offered as a courtesy.

M2.9 Moral Drift Detection

The board has mechanisms for detecting moral drift — the gradual widening of the gap between what the organisation says it values and what its systems actually optimise for. It treats coherence review not as a compliance exercise but as an integrity diagnostic, and can show how such review has produced changes in architecture or behaviour.

M2.10 Purpose Stability

The board maintains clarity about why the organisation exists — its purpose in the deep sense — even as strategy, structure and technology evolve. It can articulate, in plain language and with operational specificity, what constitutes a life worth living for the people it serves, and can show how that articulation shapes current decisions.

M2.11 Murmuration Capacity

The board functions as a distributed cognitive system capable of collective sense - making that exceeds the sum of individual contributions. It does not depend on the Chair or a dominant executive to provide the interpretive frame. Insight emerges from the group's interaction with information, and the culture rewards truth - seeking over position - defending.

M2.12 Constitutional Authority

The board understands and exercises its role as constitutional mediator rather than operational manager or passive approver. It knows which decisions are properly its own and holds that boundary even when it would be easier to delegate, approve, or defer. Its legitimacy derives from the quality of its judgements, not from its position in the hierarchy.

M2.13 AI Constitutional Framework Governance

The board has assessed the constitutional frameworks of the AI systems it deploys — the values, commitments, and design principles embedded by their developers — and understands where those

frameworks diverge from the organisation's own values. The assessment precedes deployment. The monitoring is continuous. The board can demonstrate, for each material AI system in use, whose values are doing the governance work and on what basis the organisation has determined that arrangement to be acceptable.

M2.14 Purpose - Led Design

The board can demonstrate that the organisation's governance architecture was designed from purpose rather than toward compliance. It can articulate, for its major governance design decisions, the purpose - led reasoning that preceded the compliance mapping — and can show where that mapping confirmed the design rather than drove it. Where a governance element exists primarily to satisfy a regulatory requirement rather than to serve organisational purpose, the board has made that determination explicitly and can explain why the compliance obligation was not sufficient reason to redesign from purpose.

M2.15 Telos / Function Distinction

The board can demonstrate that it maintains an active distinction between the organisation's conferred telos and the operational function of the AI systems it deploys. It can identify, for material decisions shaped by AI system outputs, whether the outcome is traceable to the organisation's own governance determination or to the optimisation objective of the system that informed it. Where those two things have aligned, the board can show that the alignment was verified rather than assumed.

A note on the tests and the architecture

The Governance Intelligence Tests do not stand alone. They are the third element of a three - document quality framework: the Foundational Principles establish the philosophical settlement from which the architecture is derived; the AOS Reasoning Principles express that settlement as operational logic applied continuously to every design decision; and these tests provide the structured accountability mechanism against which the architecture's claims are held. Each document presupposes the one before it. The tests can only be properly interpreted in light of the reasoning principles. The reasoning principles only make sense if the foundational commitments are accepted as given.

The Mode 1 tests are, in principle, progressively automatable as the architecture matures — a well - designed AI - native organisation should eventually be able to produce continuous evidence against most of them. The Mode 2 tests are permanently human. That is not a limitation of the framework. It is the point of it.