AI GOVERNANCE PAPER SERIES

MyConsultancy

AI ACCOUNTABILITY, LIABILITY AND VENDOR RISK

Why responsibility is shared, but accountability is not

MANOJ TAVARAJOO

Executive Summary

AI is no longer built and operated within clear organisational boundaries. It is assembled across vendors, platforms, foundation models and infrastructure providers. Responsibility becomes distributed across that chain. Control becomes indirect. Visibility becomes partial.

Accountability does not follow the same pattern.

Organisations remain accountable for outcomes, regardless of how many external dependencies are involved. Regulators assess outcomes, not vendor structures. Customers experience impact at the point of use, not at the point of development.

This creates a structural gap between how AI systems are delivered and how accountability is assigned.

That gap is where risk concentrates.

Courts and regulators have already begun testing this position, and the legal landscape is shifting in directions that increase, rather than reduce, organisational exposure.

Across the vendor chain, control reduces as dependency increases. Behaviour is shaped by layers that sit outside direct enterprise visibility. Changes can occur without notification. Behaviour can shift without explicit approval, often without any clear signal that the underlying system has evolved.

> **The system you approve is not the system that runs.**

Contracts allocate responsibility, but they do not transfer accountability. Governance that relies on documentation, design-time validation or contractual assurances is therefore incomplete by design.

Accountability must be enforced operationally through defined behaviour, continuous monitoring and evidence of control.

The Accountability Illusion

As discussed in the opening paper of this series, the 2024 Civil Resolution Tribunal ruling in Moffatt v Air Canada established a clear principle: the organisation that deploys an AI system is accountable for what it does, regardless of how that system was built or whose technology it runs on. The airline's argument that the chatbot was a separate entity responsible for its own actions was rejected without qualification.

Despite this, many organisations continue to operate under a different assumption. AI systems are consumed as services, integrated through APIs and governed through vendor contracts. This creates a belief that responsibility for behaviour can be transferred along the delivery chain.

In practice, it cannot.

Responsibility can be distributed. Accountability cannot.

The illusion is reinforced by abstraction. Systems appear stable. Interfaces remain consistent. Contracts define responsibilities. From the outside, the system looks controlled, predictable and contained.

Beneath that surface, it is neither static nor fully visible.

Models are updated. Data evolves. Routing logic adapts. Guardrails are modified. System prompts are adjusted. Third-party tools are introduced or replaced. Many of these changes occur without explicit notification, and often without the organisation having a mechanism to detect them.

The endpoint remains the same. The behaviour may not.

This creates a gap between perceived control and actual control. Organisations operate under the assumption that what was approved continues to operate. In reality, the system may have changed in ways that are not immediately visible, measurable or understood.

This is the accountability illusion.

The Vendor Responsibility Chain

Every AI system is delivered through a chain. Responsibility is distributed across that chain. Accountability is not.

VENDOR RESPONSIBILITY CHAIN

Responsibility is shared across the chain. Accountability remains with the organisation.

CLOSER TO THE BUSINESS ↑ / ↓ FARTHER FROM THE BUSINESS

THE CHAIN	WHAT THIS PARTY IS RESPONSIBLE FOR	CONTROL VISIBILITY	KEY LIMITATIONS	COMMON FAILURE POINTS (WHERE THINGS BREAK)
ORGANISATION (DEPLOYER)	• Define use cases and business context • Ensure appropriate governance and controls • Human oversight and decision authority • Monitor system behaviour and outcomes • Comply with laws and regulatory obligations • Accept residual risk	**High** Full visibility of usage, integration, decision impact and outcomes within the organisation.	• Dependent on supplier transparency • Cannot change underlying models • Cannot access all data or training sources • Relies on supplier transparency	• Assumes vendor is responsible • Incomplete inventory of AI use and tools • Weak monitoring of behaviour • Unclear internal ownership
SYSTEM INTEGRATOR / PLATFORM PROVIDER	• Integrate and configure AI solution • Manage data pipelines • Ensure platform availability and performance • Implement security controls • Provide operational support	**Medium** Visibility into platform configuration, data flows and service performance.	• Limited visibility into model internals • Dependent on vendor and model provider • Cannot change model behaviour	• Over-reliance on platform defaults • Limited transparency from downstream providers • Integration assumptions not tested or validated
AI VENDOR (APPLICATION PROVIDER)	• Provide AI application functionality • Manage product updates • Apply safety controls and guardrails • Ensure service reliability • Notify material changes (where agreed)	**Medium** Visibility into application behaviour, outputs, usage patterns and service performance.	• Limited visibility into model training data • Proprietary models and techniques • May update models without notice (unless contractually restricted)	• Model updates without notification • Insufficient audit and access rights • Overbroad data usage and training rights
FOUNDATION MODEL PROVIDER	• Train and run foundation models • Manage model weights and training processes • Provide inference infrastructure • Apply safety filters and system controls	**Low** Limited to what the provider chooses to disclose.	• Opaque training data • Proprietary model architecture • Frequent model updates and silent versioning • Limited explainability	• **Silent versioning changes behaviour without visibility** • Lack of transparency in training data • Restricted or no auditability
INFRASTRUCTURE LAYER	• Provide compute, storage and networking • Ensure uptime and scalability • Maintain physical and environmental infrastructure	**Low** Visibility limited to service performance and availability metrics.	• No visibility into model behaviour • Geographic and jurisdictional constraints • Shared infrastructure limitations	• Performance variability affects outcomes • Data residency and transfer risks • Dependency concentration risk

KEY TAKEAWAY

Responsibility is distributed across the chain. Accountability is not.

The organisation remains accountable for the outcomes of AI-enabled decisions.

WHEN THE CHAIN BREAKS

The organisation absorbs the impact: customer harm, regulatory action, reputational damage and financial loss.

Contracts may allocate risk but do not remove accountability.

ACCOUNTABILITY ANCHOR: ORGANISATION (DEPLOYER)

DESIGN FOR ASSURANCE — Map dependencies. Define boundaries. Demand transparency. Monitor behaviour. Escalate early.

Figure 1: Vendor Responsibility Chain — Responsibility is distributed, accountability is not

Each layer in the chain contributes to how the system behaves in practice. The organisation defines the use case and makes decisions based on outputs. Vendors provide applications and platforms. Foundation model providers shape the underlying behaviour. Infrastructure providers enable execution.

Responsibility is shared across all of them. Visibility is not.

As you move down the chain, visibility reduces. Control reduces. Dependency increases. The organisation becomes further removed from the mechanisms that determine behaviour, while remaining fully exposed to the outcomes that behaviour produces.

Risk therefore flows in the opposite direction. It concentrates back at the organisation.

This dynamic becomes particularly important when systems change over time.

> **Silent versioning changes behaviour without visibility.**

A model update, a data refresh, a routing adjustment or a change in safety controls can all alter how a system behaves. These changes may occur within vendor environments or at the foundation model level. They are not always communicated in a way that allows the organisation to reassess risk or validate control.

From a contractual perspective, nothing has changed. From a behavioural perspective, the system may be materially different.

This is where accountability begins to detach from control, and where governance based on static assumptions begins to fail.

Hidden Dependencies and Loss of Control

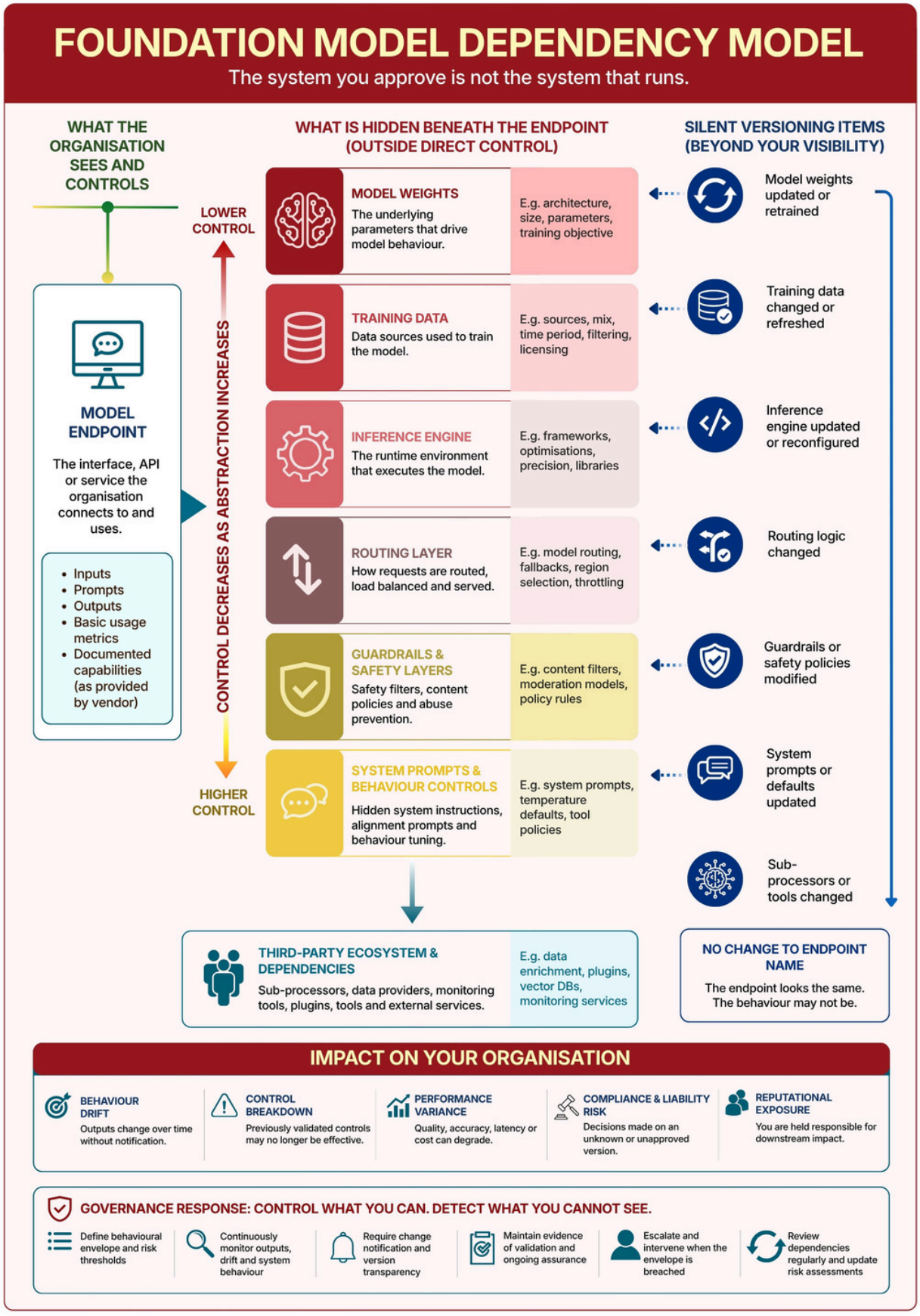

Figure 2: Foundation Model Dependency Model — The system you approve is not the system that runs

The system an organisation interacts with is only the visible surface. Beneath it sits a layered stack that determines behaviour in practice.

Model weights define capability. Training data shapes bias and performance. Inference engines determine how models execute. Routing layers decide which model responds to which request. Guardrails apply safety constraints. System prompts influence outputs. Third-party tools extend functionality.

Each layer introduces a dependency. Each dependency reduces direct control.

Industry practitioners have publicly described this pattern as silent versioning, where the same model endpoint can refer to materially different underlying systems over time.

The organisation does not control how the model was trained, how it evolves, or how underlying components are modified over time. Even where configurations are visible, the underlying behaviour may be influenced by changes beyond the organisation's line of sight.

> **The system you approve is not the system that runs.**

This is not a governance failure. It is a structural characteristic of modern AI systems.

Control therefore shifts. It moves away from configuration and towards observation. Static approval becomes less meaningful. Continuous monitoring becomes essential.

There is an additional dimension that is often overlooked.

AI capabilities are increasingly embedded within SaaS platforms. These capabilities are introduced through product updates, feature enhancements or default configurations, rather than explicit organisational decisions.

This is the embedded and shadow AI problem.

Organisations may be exposed to AI-driven behaviour without recognising it as a dependency, without governing it explicitly, and without understanding how it affects outcomes.

Dependencies extend beyond what is formally mapped.

Accountability therefore cannot rely on structure alone. It must be anchored in behaviour that can be observed, measured and controlled.

Defining the Accountability Boundary

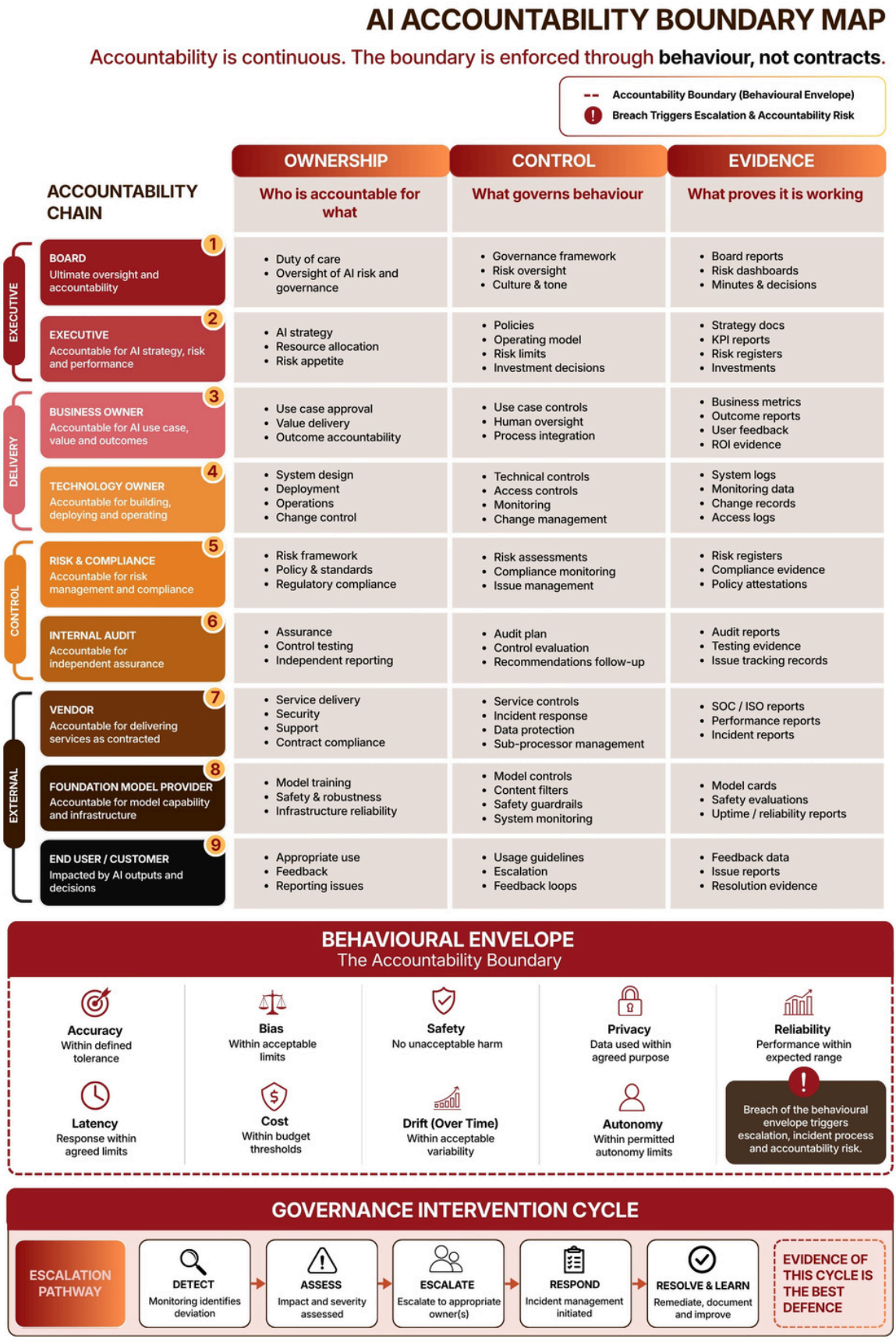

Figure 3: AI Accountability Boundary Map — Defining where accountability sits and how it is enforced

Accountability becomes real when it is tied to behaviour.

The behavioural envelope defines the limits within which an AI system is expected to operate. These limits translate governance intent into operational reality. Accuracy, bias, latency, cost, drift and autonomy define what acceptable behaviour looks like in practice.

This is not simply a governance tool.

> **It is the line at which liability attaches.**

When behaviour remains within the envelope, the system is considered controlled. When it moves outside, a breach has occurred. At that point, accountability becomes enforceable and visible.

Without defined behavioural boundaries, organisations cannot demonstrate control. Without demonstrable control, accountability remains theoretical.

The boundary connects governance intent to operational reality.

Liability in Practice

Liability follows outcomes, not architecture.

In financial services, AI-driven decisions that produce biased or inconsistent outcomes create regulatory exposure. ASIC has already signalled its focus on misleading conduct involving automated systems, while APRA CPS 230 extends this further by requiring institutions to manage operational risk across third-party dependencies.

In healthcare, AI-assisted decisions that lead to harm remain the responsibility of the provider organisation.

The legal landscape is shifting.

The EU AI Liability Directive was withdrawn in February 2025. The revised Product Liability Directive (EU) 2024/2853 extends strict liability to software and AI systems, effective December 2026.

The same trajectory is visible in the United States.

At board level, the standard of oversight remains unchanged, but expectations are rising.

The Caremark line of cases, including Marchand v Barnhill and In re McDonald's, reinforces the obligation for effective oversight of critical risks.
This is no longer hypothetical. The SEC has brought enforcement actions against firms for misleading claims about AI capabilities, signalling that AI-related disclosures and conduct now sit within the standard liability perimeter.

At the same time, the financial backstop is changing.

D&O insurers have begun introducing exclusions for AI-related losses into policies. The protection directors have historically relied on may not respond to AI-related claims.

Liability is becoming more direct, more visible and less transferable.

From Accountability to Control

Figure 4: AI Accountability Control Framework — From accountability to operational control and evidence

Accountability must be operationalised.

Ownership ensures accountability is assigned to named individuals. Data and model control ensure dependencies are understood. Monitoring ensures behaviour remains within defined limits. Response ensures deviations are addressed. Assurance ensures control can be demonstrated.

These elements are interdependent.

Without ownership, accountability is unclear. Without monitoring, drift is invisible. Without response, deviations persist. Without evidence, none of it can be proven.

> **If you cannot evidence it, you do not control it.**

Failure Modes

Organisations fail in predictable ways.

The Accountability Illusion

Organisations assume vendors carry accountability because they carry responsibility. Ownership becomes unclear, and when failure occurs, no one is positioned to respond effectively.

Control Without Visibility

Systems are approved based on initial validation but not monitored continuously. Behaviour changes over time while governance remains static, creating a gap between expectation and reality.

Contract-Led Governance

Organisations rely on legal agreements to manage risk without translating those agreements into operational controls. Contracts define obligations, but they do not enforce behaviour.

Unmanaged Dependency

Foundation model reliance and embedded AI within vendor platforms are not fully mapped. Organisations become exposed to dependencies they do not recognise or understand.

Absence of Evidence

Controls exist in principle but cannot be demonstrated in practice. When challenged, organisations are unable to prove that systems are operating within defined limits.

> **ACCOUNTABILITY IS ASSUMED RATHER THAN ENFORCED.**

Governing Accountability Across the Chain

Governance must follow accountability, not stop at the organisational boundary.

Traditional governance models assume that risk can be contained within the enterprise and managed through oversight of internal systems. That assumption does not hold for AI.

Accountability sits at the point of decision, but behaviour is shaped across a distributed system. Governance must therefore extend across that system, even where direct control does not.

This requires a shift in approach.

Organisations must define accountability at the level of decision. They must map dependencies, including those embedded within vendor platforms. They must establish behavioural boundaries that translate intent into enforceable limits. They must monitor continuously, not periodically. They must escalate when behaviour deviates. And they must evidence control in a way that can withstand regulatory scrutiny.

This is not about controlling every layer.

> **It is about ensuring that accountability can be demonstrated despite not controlling every layer.**

Questions for Boards

- Do we have clear accountability for every AI-enabled decision, assigned to a named owner?
- Can we demonstrate that systems operate within defined behavioural limits?
- What changes could occur without our visibility, and how would we detect them?
- Do we understand our dependency on foundation model providers?
- Would we be able to evidence control to a regulator today?

Questions for Executives

- Have we defined behavioural envelopes for all critical AI systems?
- Are monitoring and drift detection operating continuously?
- Do we have visibility of embedded AI within vendor platforms?
- Are escalation pathways defined and tested?
- Can we produce evidence of control on demand?

What Leaders Should Do Now

Accountability in AI does not emerge from structure. It must be deliberately established and continuously enforced.

Make accountability explicit.

Every AI-enabled decision must have a named owner who is responsible for outcomes, not just system operation

Define behavioural boundaries.

The behavioural envelope must be established for all systems that influence customers, employees or financial outcomes.

Strengthen visibility.

Dependencies across vendors, foundation models and embedded AI capabilities need to be mapped and understood.

Monitor continuously.

Behaviour should be observed in real time, with clear thresholds for deviation and defined escalation pathways.

Evidence control.

Organisations must be able to demonstrate, at any point in time, that systems are operating within defined limits.

Closing Insight

> **AI concentrates accountability.**

For boards, this is not an abstract consideration. It is a practical responsibility.

The organisations that will succeed are those that define accountability explicitly, enforce behavioural boundaries and require evidence of control.

The question is not whether AI is governed.

> **It is whether accountability can be demonstrated when it matters.**

References

- Moffatt v Air Canada, 2024, British Columbia Civil Resolution Tribunal.
- European Union. Product Liability Directive (EU) 2024/2853. Official Journal of the European Union, 2024.
- European Commission. Withdrawal of the AI Liability Directive. February 2025.
- Marchand v Barnhill, 212 A.3d 805. Delaware Supreme Court, 2019.
- In re McDonald's Corporation Stockholder Derivative Litigation. Delaware Court of Chancery, 2023.
- Australian Securities and Investments Commission. Regulatory guidance on misleading conduct and digital systems. ASIC, 2025.
- Australian Prudential Regulation Authority. Prudential Standard CPS 230 Operational Risk Management. APRA, 2023.
- Australian Institute of Company Directors. Director Duties and AI Governance. AICD, 2025.
- National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST, 2023.
- International Organization for Standardization. ISO/IEC 42001 Artificial Intelligence Management System. ISO, 2023.
- International Organization for Standardization. ISO/IEC 23894 Artificial Intelligence Risk Management Guidance. ISO, 2023.

Disclaimer: *This paper does not constitute legal advice. The regulatory and liability landscape for artificial intelligence is evolving rapidly. Organisations should seek guidance from qualified legal and compliance advisers to assess specific obligations and risks.*

About the Author

Manoj Tavarajoo has spent over two decades working with boards, executives and senior leaders across enterprise, digital and AI transformation. His work sits at the intersection of strategy, operating model design, governance and execution, the point where good intentions either become operating reality or quietly fail.

He is the author of *Leading the AI Transformation* and *The AI Operating Model Playbook*. This paper series extends that work into the governance layer: how AI is directed, controlled, assured and held accountable at enterprise scale.

He works through MyConsultancy, an independent advisory practice based in Australia.

 @manojtavarajoo

About MyConsultancy

MyConsultancy works with boards and executives navigating the distance between AI ambition and operating reality. The firm focuses on strategy, governance and operating model design, helping organisations build the portfolio discipline and transformation assurance needed to scale AI responsibly across complex enterprise environments.

 www.myconsultancy.com.au

MyConsultancy