A 42 CONSULTING.AI WHITEPAPER

The Accountability Gap

Why AI Governance Frameworks Protect No One When Something Goes Wrong

Dr M Maruf Hossain, PhD, GAICD
Chief AI Strategist · 42 Consulting.AI
enquire@42consulting.ai · www.42consulting.ai

42 Consulting.AI
Scaling Intelligence

CONTENTS

Table of Contents

FOREWORD

From the Author.

When an AI system makes a decision that harms a customer, breaches a regulation, or moves money it should not have moved, a single question follows — and it is not a technical one. *Who is accountable for this?*

Most organisations believe their AI governance framework answers that question. In my experience, it usually does not. It describes principles, assigns committees, and sets out policies — and then, when something goes wrong, it turns out that no individual can be shown to have authorised the system to do what it did, within what limits, on what date. The framework documented intent. It did not establish accountability.

This paper is about that gap — the distance between a governance framework that appears accountable and one that actually assigns responsibility to a named person and protects them with evidence. It is the gap that determines whether, when the regulator calls, your directors and executives are defensible or exposed.

It is written for three readers. The board needs accountability and clarity. The chief executive, who needs self-protection. And the chief AI officer, who needs to know whether the architecture they have built will hold.

M Maruf Hossain

SECTION ONE

Executive Summary.

When an AI system causes harm in a regulated enterprise, accountability does not disperse into the framework that governed it. It concentrates on named individuals — and most governance frameworks leave those individuals undefended. This whitepaper establishes three propositions.

First, most AI governance frameworks cannot answer the regulator's first question. That question is not 'did you have a policy?' It is 'who authorised this system to do what it did, within what limits, and on what date?' A framework built from principles, committees, and policies describes how decisions ought to be made. It does not, by itself, produce the record of who actually authorised a specific system to take a specific action. When that record does not exist, accountability cannot be located — and a framework that cannot locate accountability protects no one.

Second, accountability is personal, and it runs up a chain. In regulated enterprises, responsibility for an AI system's conduct does not stop at the system, the vendor, or the team that deployed it. It runs up through the chief AI officer, the chief executive, and ultimately the board. Each link is a named, accountable person. The framework's job is to make that chain defensible — to ensure that authority delegated downward is matched by evidence that can be produced upward. Most frameworks delegate the authority and never build the evidence.

Third, accountability is established by architecture, not by policy. The protection a director needs is not a stronger statement of intent. It is a set of artefacts: an aggregate view of how much authority has been delegated, a record created before each system acts, and a continuous test of whether that authority still holds. These are structural, not declaratory. They are produced by a constitutional governance architecture — board-adopted, enforced by infrastructure — rather than asserted by a policy that commercial pressure can override.

This paper sets out the accountability chain, the question frameworks fail to answer, why they fail, and the architecture that closes the gap. It ends with five questions that every board should be able to answer today — because each is one a regulator could ask tomorrow.

SECTION TWO

The Accountability *Chain.*

Accountability for an AI system in a regulated enterprise is not a diffuse organisational property. It is a chain of named individuals that runs in a specific direction.

At the top sits the board, which adopts the mandate under which AI may operate and sets the ceiling on the authority it may exercise. Below the board, the chief executive is accountable for the enterprise's conduct, including the conduct of its AI systems. Below the chief executive, the chief AI officer or responsible executive oversees AI within the delegated authority. And at the base, the AI system acts — and when it acts, it binds the organisation. Authority flows down the chain. When something goes wrong, accountability flows back up to it.

Figure 1 — The accountability chain. Authority is delegated downward; accountability and the evidence to discharge it must flow back up.

The asymmetry in that diagram is the entire problem. Authority is easy to delegate: a system is deployed, a team is empowered, a capability is switched on. Accountability is hard to discharge because doing so requires *evidence* — proof that the person at each level authorised what occurred within the limits they set and could show they were overseeing it. Delegation happens by default. Evidence has to be deliberately built. Where it is not built, the chain still exists — but it is a chain of exposure rather than a chain of defence.

This is why accountability cannot be answered by pointing to a governance committee or a policy document. A committee is not an accountable person. A policy is not a record of authorisation. When a regulator traces the chain, it arrives at an individual and asks what that individual authorised and how they oversaw it. The framework either equips that person with an answer or it does not.

The chain runs in one direction and is set out below. Authority is delegated downward; the evidence to discharge accountability must travel back up.

Link in the chain	What this person authorises	What they must be able to produce
The board	The mandate under which AI may operate, and the ceiling on the authority it may exercise	The adopted mandate, the ceiling, and the record of any decision to raise it
The chief executive	The enterprise's conduct, including the conduct of its AI systems	Evidence that the mandate was operationalised, not merely adopted
The chief AI or responsible executive	Which systems operate, at what autonomy, under what oversight	A current register of delegations and the limits set on each
The system owner	The specific decisions this system makes within the limits it was given	The authorisation record created before the system acted

Table 1 — The accountability chain, and what each link must be able to produce.

The asymmetry is the entire problem. Every row in the first column is easy to fill: a system is deployed, a team is empowered, a capability is switched on. Every row in the third is hard, and it is the only column a regulator reads.

SECTION THREE

The Regulator's *First Question.*

When an AI system causes a material problem, the regulatory enquiry does not begin with the algorithm. It begins with authority. The first question is some version of: who authorised this system to do what it did, within what limits, and on what date?

This question is disarmingly simple, and most governance frameworks cannot answer it. They can produce an AI policy. They can produce minutes of a governance committee. They can produce a risk assessment completed before deployment. What they typically cannot produce is a single, specific record stating that a named person authorised *this* system to take *this* kind of action, up to *these* limits, as at *this* date — a record that existed before the system acted, not one reconstructed afterwards to explain what happened.

The distinction between a record and a reconstruction is decisive. A reconstruction assembled after an incident is, by its nature, shaped by knowledge of the outcome. It is vulnerable to the suspicion — fair or not — that it was built to justify rather than to authorise. A record created before the action carries no such vulnerability. It is evidence of authorisation precisely because it could not have been informed by what the system later did.

In the Australian context, this question has personal force. Under the Financial Accountability Regime, accountability for areas of an entity's operations is assigned to named accountable persons, and the directors' duty of care under section 180 of the Corporations Act applies to the oversight of those operations. A director who cannot point to the authorisation for an AI system operating within their area of responsibility has not discharged the duty by relying on a general policy. The regulator's first question is, in substance, a question put to a person — and the framework must have armed that person to answer it.

> *The regulator's first question is not 'did you have a policy?' It is 'who authorised this system to do what it did?' Most frameworks cannot answer it.*

SECTION FOUR

Why Frameworks *Protect No One.*

If accountability is personal and the regulator's first question is about authorisation, why do so many governance frameworks fail to protect the people they ostensibly cover? The failure is structural, and it has three causes.

The first is the confusion of policy with proof. A policy states how decisions should be made. Proof is the record that a decision was made that way. Most frameworks are rich in the former and silent on the latter. They specify that high-risk AI requires senior authorisation, without producing the artefact that shows a particular system received it. When the regulator asks for proof, the organisation offers a policy — and a policy is not proof.

The second is the absence of an aggregate view. Authority is granted system by system, each grant defensible on its own. Almost no organisation maintains a consolidated view of how much autonomous decision authority it has delegated across its entire AI estate, or of who holds the power to increase that total. An accountable person cannot be protected in respect of an exposure that the organisation cannot even measure. The board cannot be shown to have controlled what was never shown.

The third is reliance on point-in-time assurance. A system authorised at deployment is treated as authorised indefinitely, even as the conditions that justified the authorisation drift. When a problem emerges months later, the authorisation on file describes a system that no longer exists in the same form. The record, such as it is, has gone stale — and a stale authorisation defends no one. Assurance that was true once is offered as though it were true continuously, and the gap between the two is where liability lands.

Each of these is a failure of evidence, not of intent. The organisations they catch are rarely careless; they have policies, committees, and good faith. What they lack is the architecture that converts intent into producible proof. That architecture is the subject of the next three sections.

SECTION FIVE

Governance That *Cannot Be Overridden.*

The first element of an architecture that protects people is that it cannot be quietly set aside. A governance arrangement that yields to commercial pressure does not protect the individuals in the accountability chain — it exposes them, because it creates the appearance of control without its substance.

The distinction is between a policy and a constitution. A policy is adopted by management and can be adjusted by management — under deadline pressure, a competitor's launch, or a revenue target, the policy bends, and there is no structural record that it did. A constitutional arrangement is different: the authority under which AI may operate is set by a board-adopted ceiling that management cannot unilaterally raise. To exceed it requires returning to the board. The constraint is not a statement of intent that can be reinterpreted; it is a structural limit that leaves a trace whenever it is approached or changed.

This matters for accountability because it changes what each person in the chain can demonstrate. Under a policy regime, an executive who exceeded prudent authority can always be said to have acted within a policy that was, conveniently, flexible. Under a constitutional regime, the ceiling is fixed and visible: either authority stays within it, or there is a broad record of it being raised. The accountable person is protected by the same structure that constrains them, because the structure produces the evidence of where authority stood at every moment.

Constitutional governance is therefore not a matter of stricter rules. It is a matter of where the rules live. Rules that live in a policy document live at the discretion of those they govern. Rules that live in a board-adopted mandate, enforced by the infrastructure that operates the AI, live beyond that discretion — and only rules of the second kind can defend the people who must answer for them.

SECTION SIX

Knowing What *You Have Delegated.*

An accountable person cannot answer for an exposure the organisation has not measured. The second element of a protective architecture is therefore an aggregate; current view of how much machine authority has been delegated across the enterprise — and who controls its total.

The mechanism for this is what the accompanying preprint series terms an Autonomy Budget (Hossain, 2026a). Rather than approving each AI system in isolation, the organisation quantifies the decision authority each system holds, aggregates it into a portfolio-level total, and sets a board-approved ceiling on that total. Individual deployments are then drawn against the budget, and the board can see — at any moment — how much of the enterprise's autonomous authority is committed, to what, and how close the total sits to the ceiling it approved.

The decision authority of a given system is not a matter of impression. It can be scored across dimensions that determine how much consequence the system can bind — the financial authority it exercises, its reach into customers and operations, and the velocity at which it acts — with loadings for factors such as irreversibility and multi-agent interaction. The scoring is the basis on which authority is drawn against the budget and on which the board's ceiling is set and monitored.

The protective effect is direct. When a regulator asks how much autonomous authority the enterprise has delegated, the organisation has an answer rather than a search. When it asks who could have increased it, the answer is the board, by resolution — not an unrecorded accumulation of individually reasonable decisions. The aggregate view converts a question that would otherwise expose every accountable person into one that the architecture answers on their behalf.

SECTION SEVEN

The Record That *Answers the Question.*

The single artefact that most directly closes the accountability gap is a record, created before an AI system acts, of what it was authorised to do, by whom, and within what limits. It is the direct answer to the regulator's first question — and it is the protection that no policy can provide.

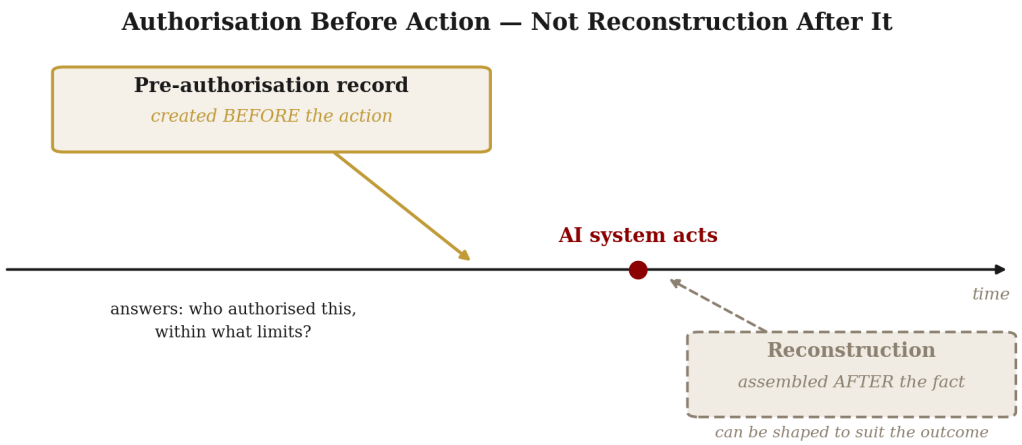

Figure 2 — A pre-authorisation record exists before the action and answers who authorised it; a reconstruction is assembled afterwards and can be shaped to the outcome.

The principle is that authorisation precedes action, and the record of authorisation precedes it too. Under a Write-Before-Execute standard (Hossain, 2026c), the authorising record is committed *before* the system is permitted to proceed — it is the act that grants permission, not a log written after the fact. If the record cannot be written, the action does not occur. This inverts the usual relationship between action and audit trail: the trail is not a description of what happened; it is the precondition of its happening.

This is what makes the record defensible in a way a reconstruction never can be. A reconstruction assembled after an incident is shaped by hindsight and open to the charge that it was built to justify the outcome. A record that demonstrably existed before the action cannot have been. For the accountable person, the difference is the difference between offering evidence and offering an explanation — and a regulator weighs the two very differently.

Authorisation is also not permanent. The conditions that made a delegation appropriate when it was granted will change, and a record of original authorisation does not, by itself, show that the authority remains appropriate now. A complete architecture therefore tests admissibility continuously (Hossain, 2026b), monitoring whether a system's operating conditions still match the profile under which it was authorised, and requiring reassessment when they diverge. The pre-authorisation record answers what was authorised; continuous testing answers whether it still holds. An accountable person needs both answers, because a regulator will ask both questions.

SECTION EIGHT

Five Questions *for Your Board.*

The test of whether a governance framework protects the people accountable under it is not its comprehensiveness. It is whether it can produce evidence. Each of the following five questions is one a regulator could ask after an incident — and each should be answerable today, with an artefact rather than an assurance.

1. **Who authorised it?** For any AI system in operation, can you produce a record — created before the system acted — of who authorised it, to do what, within what limits, and on what date?
2. **How much have we delegated?** Can the board state the total autonomous decision authority delegated across all AI systems, and name who is authorised to increase it?
3. **Could anyone override the limit?** Is the ceiling on AI authority a board-adopted constraint that management cannot unilaterally raise — or a policy that bends under pressure without leaving a trace?
4. **Does the authorisation still hold?** Can you show that each system's authority remains valid under current conditions, rather than relying on a single assessment at deployment?
5. **Who answers for it?** If a regulator traced the accountability chain for an AI decision to a named person, would that person be armed with evidence — or left to reconstruct a defence?

> *A framework that cannot answer these five questions does not protect the people accountable under it. It documents them.*

ABOUT

42 Consulting.AI.

42 Consulting.AI builds constitutional AI governance for regulated enterprises. Its core work, the MANDATE Suite, is a comprehensive governance architecture comprising 64 normative instruments across five frameworks — governing the delegation of machine decision authority, the ethical and rights boundaries on its use, the management system that maintains it, and the regulatory extensions for EU and group structures. The Suite is designed for organisations in financial services, insurance, superannuation, healthcare, government, and other regulated sectors operating autonomous AI systems at scale.

The architecture is supported by an academic preprint series published on Zenodo, which establishes the theoretical basis for each governance mechanism, and by the forthcoming book *Governance-First AI*. The Suite's distinguishing conviction is that AI governance at the depth regulated enterprises require cannot be delivered by a compliance framework alone — it requires a constitutional architecture, adopted by the board, enforced by the infrastructure, and continuously tested against current conditions.

Dr M Maruf Hossain, PhD, GAICD · Chief AI Strategist
enquire@42consulting.ai · www.42consulting.ai

The preprint series is available at: zenodo.org/communities/mandate/records

REFERENCES

References *and Sources.*

The regulatory instruments referenced in this paper were verified at the primary source. The governance mechanisms are set out in full in the MANDATE preprint series, published through Zenodo.

REGULATORY AND INSTITUTIONAL SOURCES

APRA & ASIC (2024). *Financial Accountability Regime: Information for Accountable Entities.* Commenced 15 March 2024 (banking); 15 March 2025 (insurance and superannuation).

ASIC (2024). *Report 798: Beware the Gap — Governance Arrangements in the Face of AI Innovation.* Australian Securities and Investments Commission, 29 October 2024.

Commonwealth of Australia. *Corporations Act 2001 (Cth), section 180 — Care and diligence.*

MANDATE PREPRINT SERIES

Hossain, M. M. (2026a). *The Autonomy Budget: A Portfolio-Level Framework for Governing Delegated Machine Authority in Regulated Enterprises.* Zenodo. https://doi.org/10.5281/zenodo.20480491

Hossain, M. M. (2026b). *Continuous Admissibility: A Temporal Governance Framework for Delegated Machine Authority in Regulated Enterprises.* Zenodo. https://doi.org/10.5281/zenodo.20580716

Hossain, M. M. (2026c). *Write-Before-Execute: A Governance Standard for Evidentially Defensible AI Decision Logging in Regulated Enterprises.* Zenodo. https://doi.org/10.5281/zenodo.20603712

Consulting.AI

Scaling Intelligence

enquire@42consulting.ai · www.42consulting.ai