Governing Agentic AI

A Board Pack

Frédéric Verhelst · Version 1.1 · 16 June 2026

Your AI agents no longer just answer questions; they take actions, moving money, changing settings, sending messages, on their own and faster than anyone can review. The organisation is fully liable for what they do, yet the controls most companies rely on were built for people working at human pace and cannot keep up. Accuracy is not the issue: these systems work by prediction, not by fixed rules, and even an agent working from perfect data can still act outside what it was ever allowed to do. What closes the gap is a set of controls that enforce the rules automatically before an agent acts, can stop an agent the instant it goes wrong, and are held by an executive with the authority to halt a launch the business is pushing. This pack is the board's view of those controls, and how to tell whether yours exist.

Inside: why this is a board-level liability, the controls that prevent it and who owns each, whether you own or rent them, and the questions to put to management today. The supporting papers carry the detail.

The board problem

Major advisory firms name the Chief Data Officer as the executive most prepared for AI transformation, and in the same breath project that most CDOs not seen as essential to AI success will lose their C-level position within two years. Both readings are correct. The traditional CDO mandate, built for analytics and dashboards, is misaligned with the risks of agentic AI.

The reason is tempo. The traditional mandate assumed human operators resolved semantic ambiguity across business units through consensus. Agents act at machine speed and cannot wait for consensus. At scale, ambiguity stops being coordination overhead and becomes irreversible error.

The liability reality

The organisation is strictly liable for what its agents do, independent of how the underlying model reasoned. This is the organisational equivalent of Keeper's Liability, operated at machine tempo. Insurability requires two conditions: a determinable debtor and a calculable damage envelope. Underwriters are already introducing sub-limits and exclusions for AI-related losses, because correlated failures executing at machine speed are uninsurable.

Traditional governance rests on three lines of defence: operational management, independent risk and compliance, and internal audit. All three were designed for human tempo and review after the fact. Agentic AI exposes the absence of a fourth line, one that fires before an agent acts and can halt it while it acts. Without that line, the organisation holds absolute liability for systems it cannot reliably constrain.

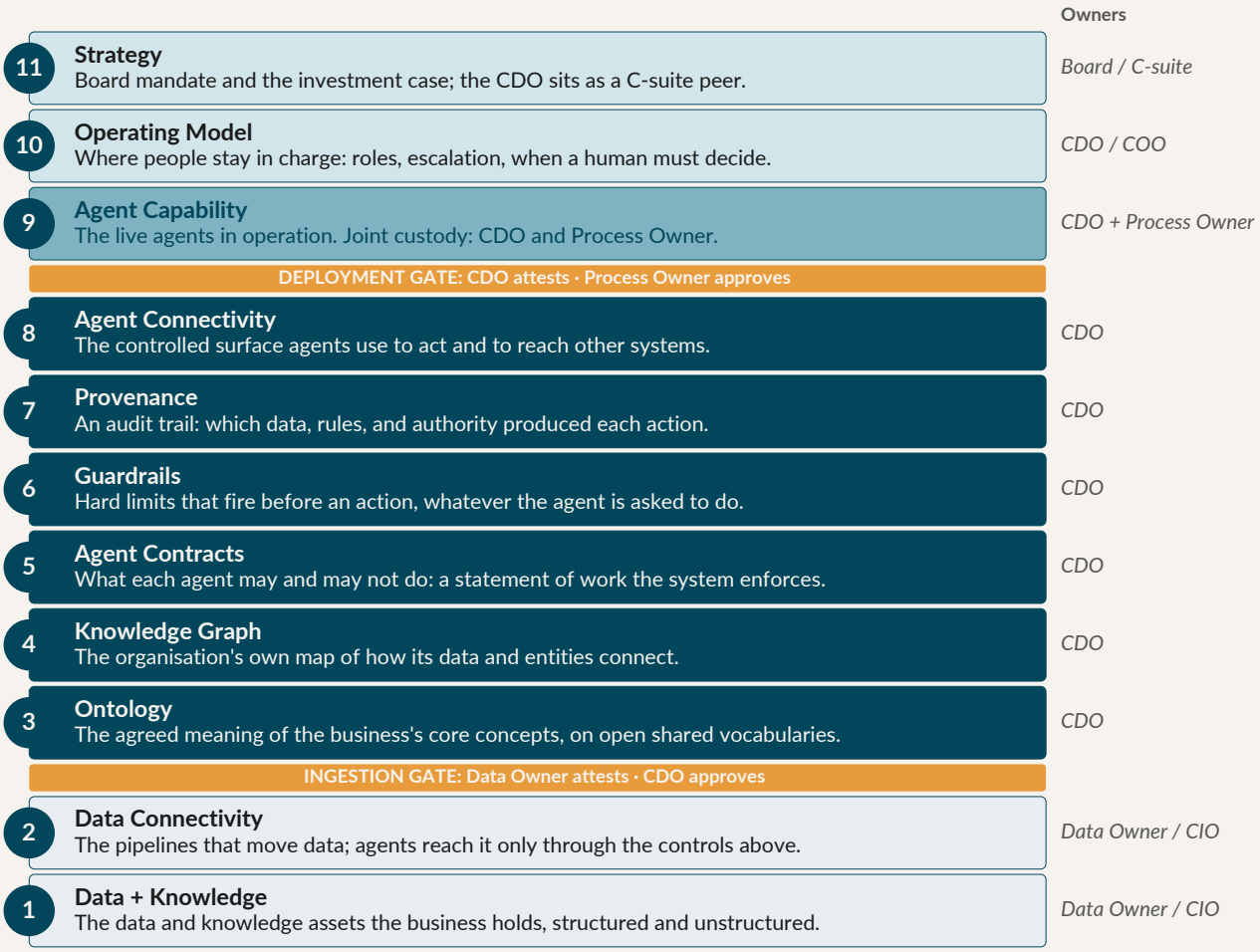
The standard-of-care question

After any agent-driven loss, one question arrives first: did the board provide reasonable governance against a known risk? Deterministic governance that constrains agents before they act, and records every decision, is the structural answer. A policy document is not.

The Capability Stack in one view

The fourth line is not a strategy document. It is a deterministic, machine-readable governance structure: the Agentic AI Capability Stack. It names the eleven capabilities an enterprise must own to deploy agentic AI it can govern, and who owns each, on open standards the organisation controls rather than a vendor's. Two gates decide what enters the meaning capability and what reaches production.

A board is often reassured that its AI is grounded, or that it runs on a knowledge graph. Grounding governs what an agent knows; it does not govern what an agent is permitted to do. A perfectly grounded agent, working from impeccable data, can still take an action it was never authorised to take. That is why the Stack carries capabilities above the data: the ontology and knowledge graph (Capabilities 3 and 4) make the agent well informed, while the contracts, guardrails, and provenance (5 to 7) and the two gates govern what it is allowed to do, before it acts.



The Agentic AI Capability Stack™: eleven capabilities and two gates. Capabilities 3 to 8 are owned by the CDO and built on open standards the organisation controls, not a vendor's format; capability 9 is joint custody with the Process Owner.

The control points

The Stack enforces governance through three named control points: two preventive gates and one runtime halt. Each is an authority a named executive holds, not a policy a document asserts. The gates fire before anything irreversible happens; the halt governs the agent while it is live. The Ingestion Gate stands between data and meaning, so bad data never becomes a trusted fact. The Deployment Gate stands between governance and runtime, where the executive who approves custody also accepts the liability.

INGESTION GATE	DEPLOYMENT GATE
between data and meaning	*between governance and runtime*
Data Owner attests	**CDO attests**
data meets agent-readiness	governance content complete
CDO approves	**Process Owner approves**
admission to meaning	operational custody

RUNTIME HALT · Capability 9, after the gates, while the agent acts

CDO or Process Owner halts unilaterally · stop is instant · reinstatement requires both signatures

Two principles make the control points enforceable

Separation of duties. At each gate, attesting is deliberately separated from approving, and the CDO's role flips between them. It is the same discipline that lets a CFO certify the numbers while the business authorises the spend.

Fail-safe asymmetry. At the runtime halt, either custodian can stop a live agent unilaterally, while reinstatement requires both signatures. This answers the only question that matters in an incident: not who approved the agent months ago, but who can stop it now.

Ownership and accountability

The reporting line

The CDO reports to the CEO. The standards-setting authority cuts across the Data Owners' domains, the CIO's substrate, the COO's operating model, and the Process Owners' deployments; only the CEO has authority over all of those at once. A mandate positioned below the C-suite cannot hold a deployment gate against a peer or superior pushing for rapid adoption. There is no Chief AI Officer above it. That role exists where the CDO mandate stopped at pipelines and dashboards.

The financial-governance parallel

The mandate has a precedent boards already trust. The CFO owns the chart of accounts, the general ledger, the internal controls that fire before unauthorised activity executes, and the audit trail regulators can examine. The CDO mandate carries the same four functions for agents: the meaning of its core concepts, the graph of how its data connects, the constraints that fire before an action, and the audit trail. The parallel is design, not metaphor.

The Process Owner

The mandate names a counterpart rarely formalised today. Where the Data Owner is accountable for what data means, the Process Owner is accountable for the business process the agent serves. After the Deployment Gate fires, the agent enters joint custody: the CDO continues to enforce its contract, guardrails, and provenance; the Process Owner holds operational liability for the outcomes. Each retains unilateral halt authority; reinstatement requires both signatures.

Who answers for what

Function	Accountable for	Relationship to the CDO
Board	Mandate charter, fiduciary oversight, risk appetite	Contributes the governance case; charters gate authority and sets the risk appetite the deployment thresholds operationalise.
Process Owner	The operational process and the outcomes the agent produces	Joint custody of Capability 9; unilateral halt authority within the process.
CIO	Substrate: databases, pipelines, security	The CDO sets provenance and verifiability standards.
COO	Operating model: human-agent roles, escalation	Co-owns the Operating Model governance dimension with the CDO.
CRO	Enterprise risk-tolerance frameworks	Co-owns the tolerance frameworks the CDO encodes as guardrails and deployment thresholds.

Legal, Risk, and the CISO set the enterprise standards the CDO encodes and are consulted at the gates rather than shown as separate columns.

Sovereignty and the vendor decision

The means to govern agents is routinely outsourced in the rush to deploy. That is a fiduciary miscalculation. Relying on a single vendor for the model, the deployment engineering, and the runtime transfers the operational graph outside the organisation. If the organisation does not own the meaning capability that defines its business logic, the meaning its agents reason from is rented.

Three sovereignties operate as a chain. Semantic sovereignty is ownership of the ontology and knowledge graph. Enforcement sovereignty is the ability to update constraints on the organisation's schedule, not the vendor's release cycle. Execution sovereignty is independent evidence that the runtime enforced those constraints. These three links rest on a fourth the chain does not name: jurisdictional sovereignty asks whether a third party holds the off-switch, since an export-control or sanctions order can disable a model for every customer overnight, and backups do not restore an agent that can no longer reason. Lose a link and the chain collapses to whoever owns it; lose the ground and all three go dark at once. The mandate cannot be bought. It must be held, on open standards that keep the ontology, the constraints, and the provenance portable to another jurisdiction.

Six questions for the board to ask today

An organisation that cannot answer these is operating autonomous systems without a calculable damage envelope. The liability is absolute; the mandate must match it.

#	Six questions for the executive team
Q1	Who is chartered to block an unsafe agent at the deployment gate, before it reaches production?
Q2	Once an agent is live, who can halt it within seconds, without convening a committee?
Q3	Can we update agent constraints on our schedule, not the vendor's release cycle?
Q4	Can we reconstruct what an agent was authorised to do within 24 hours of an incident?
Q5	Do we have independent evidence agents ran as specified, or only the vendor's word?
Q6	Who is accountable when an agent produces a consequential error?

The three operating modes

The mandate scales with the scope of autonomous operation, across three modes. The two structural gates apply at every mode; what changes is how much technical governance the Process Owner delegates to the CDO, and where operational liability sits. Most organisations begin at Mode 1, a few governed agents in a single business unit with the CDO holding both gates, and mature toward broader delegation as the foundational ontology and the control record prove out.

Mode	Scope, and where operational accountability sits
1 · Governed Infrastructure	A few agents inside a business unit. The CDO defines standard governance contracts and holds both gates; the Process Owner selects from pre-approved templates and retains operational accountability for outcomes.
2 · Delegated Governance	One fully agentic workflow, end to end. The Process Owner formally delegates technical governance of the agentic portion to the CDO by charter or board resolution, and retains operational liability for outcomes.
3 · Operational Accountability	A fully agentic, demanned division. The CDO inherits the Process Owner role; both gates still fire, and the audit acts remain distinct.

About this pack

This is a compilation. It draws the essential element from each artifact in the governance suite: the Agentic AI Capability Stack for the capabilities and gates, the Board Briefing for the liability case, and the CDO Office Mandate for ownership and the operating modes. The full artifacts, together with the Vendor Coverage Diagnostic, the Agentic AI Readiness Assessment, and the Use Case Risk Classification Framework, are in the library. Licensed CC BY-ND 4.0.

About the author

Frédéric Verhelst, PhD (Applied Physics, TU Delft), works with boards and executive teams on the governance of trustworthy agentic AI. Twenty-five years across data, AI, and regulated industrial operations: external executive advisor on the Tyra Redevelopment FID (a Mærsk Oil joint venture with Shell and Chevron, later operated by TotalEnergies), then Head of Data Office at TotalEnergies EP Danmark, most recently bringing agentic AI into safety-critical maritime operations. Currently available for non-executive director (NED), board-advisory, and CDO appointments focused on corporate AI governance.

