

PUBLISHED BY



**PRIVATE DIRECTORS
ASSOCIATION®**
Creating Value Through Board Excellence

IN COLLABORATION WITH



Center for
AI Oversight

— A STRATEGIC GUIDE FOR PRIVATE COMPANY BOARDS

AI Oversight

*The Private Director's
Body of Knowledge*

Published by the **Private Directors Association**, developed in
collaboration with the **Center for AI Oversight**.

FIRST EDITION, V.1.0 | JULY 2026

Contents

Foreword

FRONT MATTER

Executive Summary

FRONT MATTER

I

The Board Briefing

- 01 The Fiduciary Pivot
- 02 How to Use This Body of Knowledge

II

The Oversight Blueprint

- 03 Ensure the Program Has Authority and Accountability
- 04 Define How Much Risk You Are Willing to Accept

III

Asset Integrity and Operations

- 05 Verify That AI Systems Are Trustworthy, Secure, and Compliant
- 06 Ensure AI Investments Are Strategically Sound and Vendor Risk Is Managed

IV

Escalation and Disclosure

- 07 Ensure Critical Risks and Emerging Opportunities Reach the Board Through Defined Channels



The Director's Toolkit

Tool A. Board-Level AI Oversight Questions

TOOLKIT

Tool B. Governance Maturity Diagnostic

TOOLKIT

Tool C. Board Mandate Matrix

TOOLKIT

Tool D. Vendor AI Due Diligence Questionnaire

TOOLKIT

Appendix A. AI Fundamentals for Directors

APPENDIX

Glossary

REFERENCE

Endnotes

BACK MATTER

Appendix B. The AI Oversight Program

APPENDIX

Appendix C. Third-Party Contract Considerations

APPENDIX

Back Matter

BACK MATTER

FOREWORD

Artificial intelligence is reshaping how private companies create value, manage risk, and compete. The implications for boards of directors are immediate: AI is no longer a technology initiative to be delegated. It is an enterprise-wide strategic reality that demands informed oversight.

This Body of Knowledge was developed through a collaboration between the Private Directors Association and the Center for AI Oversight. It draws on the Center's governance methodology and the collective experience of directors, risk professionals, and legal practitioners who have spent years at the intersection of emerging technology and fiduciary responsibility.

The purpose of this document is practical. It provides private company directors with a framework for ensuring that their organizations have a functioning AI oversight program, one that is proportionate to risk, grounded in legal defensibility, and capable of evolving as the technology and regulatory landscape change.

This is not a compliance manual. It is a strategic guide for boards that want to move confidently and quickly, with the controls in place to do so responsibly.

Lead Author: Brian Allen, Executive Director, Center for AI Oversight.

Developed with the input of a working group of private company directors convened by the Private Directors Association. The full list of contributors appears in the Acknowledgements.

EXECUTIVE SUMMARY

The central thesis of this Body of Knowledge is that effective AI oversight is not a committee. It is a program. A program is a formal management system through which the company directs, manages, and monitors its AI activities. The committee is the oversight mechanism; the program is what the committee oversees. The charter, policies, risk methodology, reporting protocols, escalation triggers, and monitoring mechanisms are the program.

Governance is the accelerator, not the constraint. The competitive landscape is not waiting for boards to deliberate. Companies that can make risk-informed AI decisions quickly will outperform those that cannot. The question courts, investors, and insurers will increasingly ask is not “Do you have a committee?” but “Do you have a functioning system, and is it being actively monitored?”

Three patterns predict where private companies will fail. The first is *Missed Opportunity*: the board, lacking a framework to evaluate AI initiatives quickly, defers promising investments and cedes competitive ground. The second is *Reckless Haste*: the board pressures management to move quickly without the governance infrastructure to manage the risks being introduced. The third is *Paralyzed Inaction*: the board defers every significant AI decision until someone else (regulators, industry bodies, advisors, or peers) defines the path first. It mistakes the absence of an external answer for the absence of a decision to make. All three share a common root cause, which is the absence of a functioning oversight program.

The legal foundation is direct. The Caremark line of Delaware cases, extended through *Marchand v. Barnhill* in 2019 to a private, family-controlled company, establishes that directors have an affirmative duty to ensure that information and reporting systems exist for mission-critical risks. The standard has two prongs: directors must ensure that a system exists, and they must actively monitor it. As AI becomes central to how private companies operate, compete, and create value, it is becoming a mission-critical risk. The absence of any oversight program is itself the failure.

This standard reaches private companies even where Caremark does not formally bind them. Investors conducting due diligence, lenders evaluating credit risk, insurers underwriting D&O coverage, and acquirers assessing valuation will all apply the same standard. The practical question is not whether Caremark technically applies, but whether the board's oversight would withstand scrutiny from any of these audiences.

What This Body of Knowledge Covers

The document is organized into five Parts, a Toolkit of artifacts designed for board use, and three appendices that provide reference material. The framework is designed to scale: a five-person company and a five-hundred-person company need different levels of formality, but both need a functioning system.

Part I. The Board Briefing

Establishes the case for action. Addresses the strategic imperative of AI as a structural shift in how value is created, the structural mismatch between AI adoption and quarterly board cycles, the distinction between companies that build AI and the majority that buy it, and the legal standard of care that courts are applying to board-level oversight.

Part II. The Oversight Blueprint

Defines the structural requirements of a functioning oversight program. Covers the Governance Charter as the program's constitutional document, the policy framework that establishes the tolerances within which management operates, board-level risk appetite expressed in concrete financial terms, the Crown Jewels Assessment that focuses governance attention on what matters most, and the discipline of auditing the program itself rather than just the AI.

Part III. Asset Integrity and Operations

Turns to what directors need to verify about the AI systems and vendors their companies rely on. Covers data ownership and provenance, the AI Registry as the mandatory inventory, human oversight standards and Agency Limits for autonomous systems, AI-specific security risks distinct from traditional cybersecurity, vendor concentration risk and contract protections, the financial discipline of AI unit economics, and operational resilience when critical AI dependencies fail.

Part IV. Escalation and Disclosure

Addresses the mechanisms that ensure critical risks and emerging opportunities reach the board through defined channels rather than through management's discretion. Covers escalation protocols triggered by defined criteria, materiality thresholds calibrated to the company's existing disclosure obligations to investors, lenders, regulators, and insurers, and crisis readiness exercises that test the response before an actual incident occurs.

Part V. The Director's Toolkit

Four one-page artifacts designed to be brought into a board meeting or handed to management with a clear expectation attached. Tool A is a curated set of board-level oversight questions. Tool B is a descriptive governance maturity diagnostic. Tool C is a RACI-style board mandate matrix. Tool D is a structured vendor AI due diligence questionnaire.

Appendices

Three reference appendices support the main document. Appendix A provides a brief orientation to the categories of AI, written for directors without a technical background. Appendix B presents one illustrative model of how the principles can be operationalized: the AI Oversight Program developed by the Center for AI Oversight. Appendix C provides a structured set of provisions for AI vendor contracts. A glossary defines key terms used throughout.

How to Use This Body of Knowledge

The framework is rigorous, not rigid. We call this the *Accordion Principle*: the program expands or contracts based on the organization's AI risk exposure, size, and complexity. What remains constant are the oversight functions. What varies is the formality, depth, and frequency with which they are executed.

Three operating modes describe most private companies. Founder-led companies, where authority is concentrated and AI exposure is limited, may need only a documented policy, an inventory of AI tools, and a quarterly conversation at the board level. Growth-stage companies, where AI use is expanding and the board is separating from day-to-day management, need a governance charter, defined risk appetite, and standing board reporting. Institutional companies, where the AI portfolio is complex and regulatory exposure is significant, need a full program with committee oversight, independent assurance, and formal escalation protocols.

Most organizations will not fit neatly into a single mode. A company may have Institutional-level AI exposure but Founder-led governance structure. That mismatch is itself a finding, and closing it is the purpose of the work the BoK describes.

What the Board Should Do Next

Three actions, in order, will start closing the gap between most private companies and the standard this document describes.

First, complete the Governance Maturity Diagnostic in Tool B. It takes fifteen minutes and locates the organization honestly. The mismatches across dimensions are the most important finding, not the absolute mode.

Second, identify who at the management level is accountable for the AI oversight program. If the answer is unclear, the program does not have authority. The Board Mandate Matrix in Tool C resolves the ambiguity.

Third, put AI oversight on the next board meeting agenda as a substantive item, not an information update. Use the questions in Tool A to structure the discussion. Document the discussion in the minutes. That single action begins building the record that Caremark requires.

One illustrative model of how these principles can be operationalized, the AI Oversight Program developed by the Center for AI Oversight, is presented in Appendix B. The Program is maintained at cfaio.org and is updated regularly to reflect changes in regulation, technical standards, and case law.

Part I

THE BOARD BRIEFING

Context, Liability, and the Case for Action

SECTION 1

The Fiduciary Pivot

1.1 The Strategic Imperative: Why This Cannot Wait

Artificial intelligence is not the next software upgrade. It is a structural shift in how value is created, captured, and competed for.

Previous technology waves digitized existing business processes. They made established work faster and cheaper. AI does something categorically different: it operates on cognition itself. It augments human judgment, automates cognitive work and, unlike any prior technology, improves at doing both. Analysis, pattern recognition, decision support, and increasingly, autonomous action are all within its reach. Across industries, this shift will compress margins, lower barriers to entry, and redefine customer expectations faster than traditional governance cycles anticipate.

What does this mean for private companies? Two things simultaneously. On one side, the risk of strategic irrelevance: competitors that re-architect their operating models around AI will create structural advantages that legacy operators cannot match through incremental

improvement. On the other, a genuine opportunity: companies that move decisively, with the right governance in place, can enter new markets, reshape their cost structures, and strengthen competitive positions in ways that were not previously available.

Directors should expect management to address four questions with specificity:

How does AI affect the company's core cost structure? Where are labor-intensive processes that AI could automate or augment, and what is the timeline and investment (both monetary and people resources) required?

Where could AI disintermediate the company's current offerings? Are adjacent competitors or new entrants using AI to deliver comparable value at lower cost or higher speed?

How are direct competitors using AI to change industry economics? What intelligence does the company have about competitors' AI adoption, and is it sufficient?

Which core processes require reinvention rather than incremental optimization? Is management distinguishing between processes that benefit from AI augmentation and those that require fundamental redesign?

A fifth question is retrospective, and in the current environment it may be the most urgent of all: does the board have visibility into the AI footprint that already exists in the business? AI capabilities are being embedded in enterprise systems faster than governance processes were designed to track. By the time a formal AI initiative reaches the board agenda, a substantial AI footprint may already exist beneath the reporting threshold. The board's question is not what each of those decisions was. That is management's work. *The board's question is whether a program exists that can surface them.*

Governance without strategic repositioning is defensive. Strategic repositioning without governance is reckless. Boards must insist on both.

The role of the board is to oversee, not to operate, the company's AI strategy. Management owns the execution. The board's responsibility is to ensure that management has a credible AI strategy, that the strategy is informed by an honest assessment of both risks and opportunities, and that the governance program is equipped to oversee its execution.

1.2 The Velocity Gap

The mismatch between the pace of AI adoption and the cadence of board oversight is not a temporary condition that more frequent meetings will resolve. AI capability develops on a

curve that keeps accelerating. Board oversight, even at its best, is linear. The gap widens on its own.

The deeper problem is that acceleration changes the character of the decisions themselves. A board that established adequate AI oversight two years ago may be structurally behind today without having made any errors. Cadence adjustments cannot close that gap. What closes it is a change in the nature of oversight: pre-mapped decision frameworks that allow the board to respond with appropriate speed and judgment when a decision arrives, rather than waiting for the next scheduled meeting.

The danger is not that directors must become technologists. They need not. The danger is that the cadence and depth of board-level AI oversight remains calibrated to a slower era, creating a structural gap between what the company faces and what the board sees. That gap is not only a risk exposure. It is a missed opportunity.

When oversight cannot keep pace, three predictable patterns emerge.

The first is Missed Opportunity: the board, lacking a framework to evaluate AI initiatives quickly and confidently, defaults to caution. Promising investments are deferred. Competitive advantages that require speed go uncaptured. Strategic windows close. The company does not fail spectacularly; it simply falls behind, one deferred decision at a time.

The second is Reckless Haste: the board hears that competitors are adopting AI and pressures management to move quickly, without ensuring that the governance infrastructure exists to manage the risks being introduced. Speed without controls is not strategy. It is exposure.

The third is Paralyzed Inaction: the board defers every significant AI decision until someone else (regulators, industry bodies, advisors, or peers) defines the path first, ceding competitive ground to companies willing to move with appropriate controls in place. It mistakes the absence of an external answer for the absence of a decision to make. Uncertainty becomes an excuse for inaction, and inaction compounds the strategic irrelevance described in the preceding section.

All three patterns share a common root cause: the absence of a functioning oversight program. A board with a program can make risk-informed decisions at speed, capturing opportunity while managing exposure. A board without one is guessing.

Governance is the accelerator. It surfaces problems before they become crises and opportunities before competitors capture them. It is what allows the board to authorize

management to move quickly, because the controls, reporting mechanisms, and escalation triggers are in place to support confident decision-making in both directions.

CASE IN POINT

OpenAI, November 2023

On Friday, November 17, 2023, the OpenAI board fired Sam Altman as CEO over a video call. Within seventy-two hours, more than seven hundred employees had signed a letter threatening to resign, Microsoft had announced it would hire Altman to lead a new AI research team, and the company that had introduced ChatGPT and reached a ninety billion dollar valuation was hours from collapse. By November 21, Altman was back as CEO and five of the six original board members had been replaced.¹

The lesson: even the board of the most consequential AI company in the world can be caught flat-footed by the speed of its own organization. A board without a functioning program cannot decide at the speed AI requires. The risk is not theoretical and the velocity is not waiting.

1.3 Builders and Buyers: Two Different Oversight Agendas

Not all companies face the same AI governance challenge. The distinction between companies that build AI and companies that buy it shapes the character of the oversight program, though not its scope.

Builders are companies developing proprietary AI models, training systems on their own data, or creating AI-powered products and services. They face the full spectrum of operational complexity: model development lifecycle management, training data provenance and bias, validation and testing protocols, explainability requirements, and the emerging obligations around autonomous agent deployment.

Buyers are the majority. These are companies purchasing AI capabilities embedded in third-party software, cloud services, and vendor platforms. They are not training models. They are using tools built by others. For most private companies, this is the reality: the AI that powers their operations was built by someone else.

For Buyers in particular, AI capability is often systemic rather than discrete. Vendor-embedded AI may already be present in the business before the board has had the opportunity to review it, and the dependency it creates may not be visible until something goes wrong.

The governance challenge for Buyers is different in kind, not just in degree. Buyer boards are not overseeing model development. They are overseeing vendor selection, contract protections, data exposure, concentration risk, and business continuity. The question is not “Is our model biased?” but “Is our vendor’s model biased, and are we exposed if it is?”

What does not change is the scope of the board’s oversight responsibility. Every section of this BoK applies to both Builders and Buyers. Where the distinction matters is in operational concentration: Builders will find that their management teams spend disproportionate time on the development lifecycle and security topics in Sections 5 and 6, and the board should expect correspondingly detailed reporting from those areas. Buyers will find that vendor risk, supply chain governance, and business continuity consume the bulk of management’s attention, and board reporting should reflect that reality. In both cases, the board’s job is the same: ensure the program exists, ensure it covers the right risks, and ensure it is being actively monitored.

Most private companies are primarily Buyers, and some are hybrids. Directors who understand where their company sits on this spectrum will be better equipped to ask the right questions and allocate their oversight attention accordingly.

1.4 The Legal Foundation: Caremark and the Duty of Oversight

The legal case for board-level AI oversight is grounded in three decades of Delaware case law that has progressively expanded the personal liability exposure of directors who fail to ensure that functioning oversight systems exist for mission-critical risks.

In *re Caremark International Inc. Derivative Litigation* (1996) established the foundational principle: directors have an affirmative duty to ensure the corporation maintains adequate information and reporting systems. The court recognized that the absence of such systems can, standing alone, establish the bad faith necessary for director liability. This was not about directors making wrong decisions. It was about directors failing to create the conditions under which informed decisions could be made.

For private company directors, the most instructive case is *Marchand v. Barnhill* (2019), involving Blue Bell Creameries. Blue Bell was a private, family-controlled company. Its board had no committee responsible for food safety, no regular reporting protocols for safety

compliance, and no system to escalate food safety issues to the board. When a listeria contamination crisis destroyed the company's operations and independence, the Delaware Supreme Court held that the board's complete failure to establish any reporting system for a mission-critical risk was sufficient to state a claim for breach of the duty of oversight.

The lesson is direct: for a company whose value depends on a specific operational risk, the board must ensure that an information and reporting system for that risk exists. If AI is becoming central to how a private company operates, competes, and creates value, it is becoming a mission-critical risk. The absence of any oversight program is itself the failure.

Subsequent cases have reinforced and extended this standard. *Stone v. Ritter* (2006) clarified that a sustained failure of oversight constitutes a breach of the duty of loyalty, meaning personal liability cannot be eliminated by charter provisions. *In re Clovis Oncology* (2019) showed that a governance structure existing on paper but not actually used provides no protection. *In re Boeing* (2021) demonstrated that when red flags reach management but are not escalated to the board, the board's defense collapses. *In re McDonald's* (2023) expanded the doctrine to officer-level liability. The full case-by-case analysis is presented in Appendix B.

For purposes of this Body of Knowledge, the practical takeaway is this: the Caremark framework establishes two prongs of director liability. Prong 1 asks whether the board ensured that a reporting and information system exists. Prong 2 asks whether the board actively monitored that system. A board that cannot demonstrate both is exposed.

This standard is not limited to Delaware corporations or to large enterprises. While the Caremark line originates in Delaware, it has become the de facto benchmark against which any board's oversight will be measured, by courts in other jurisdictions, by investors conducting due diligence, by insurers underwriting Directors & Officers (D&O) coverage, and by acquirers evaluating governance risk. Blue Bell itself was a family-controlled private company. The standard applies wherever a board of directors exists.

What proportionate compliance looks like will vary. For a small company with limited AI exposure, the "system" may be a documented policy, an inventory of AI tools in use, and a standing agenda item at quarterly board meetings. For a larger enterprise with significant AI dependencies, it will be substantially more formal. The *Accordion Principle* described in Section 2 provides the framework for scaling. The legal requirement is not that every company build the same program. It is that every board ensures a functioning system exists and that it is actively monitored.

Directors do not need to become AI experts. They need to ensure that a functioning system exists and that they are actively monitoring it. That is the standard, and it is the standard this BoK is designed to help private boards meet.

DIRECTOR'S FOCUS

What to take away

Section 1 establishes why AI oversight is now a fiduciary responsibility, not a technology initiative. The question is no longer whether to engage, but whether the engagement is structured well enough to meet the standard courts, investors, and acquirers are applying to it. Private companies do not have the benefit of delay. They face a widening acceleration gap, an increasingly systemic AI footprint in their own businesses, and a legal framework that treats the absence of a reporting system as a breach in itself.

Questions the director should be asking:

- ◆ Does the board have a credible view of the AI strategy, the AI footprint, and the gap between what the company faces and what the board currently sees?
- ◆ Are we a Builder, a Buyer, or a hybrid, and is our oversight attention allocated accordingly?
- ◆ If a Caremark-style inquiry happened tomorrow, could we produce evidence that a functioning oversight system exists and that we are actively monitoring it?

SECTION 2

How to Use This Body of Knowledge

2.1 Rigorous, Not Rigid: The Accordion Principle

Every company in this document's audience is different. A five-person firm using a single AI-powered CRM is not the same as a 500-person PE-backed platform deploying AI across multiple business lines. The governance program must reflect that reality.

We call this the Accordion Principle: the program expands or contracts based on the organization's AI risk exposure, size, and complexity. What remains constant are the oversight functions. What varies is the formality, depth, and frequency with which they are executed.

How the Accordion Principle applies depends on where the organization sits today. For smaller and founder-led companies, the concern is practicality. A company with five employees and one AI vendor does not need a dedicated committee, a formal charter, and a reporting dashboard. It needs a documented policy, an inventory of AI tools in use, and a quarterly conversation at the board level about whether the risks have changed. That is a functioning system, and it satisfies the legal standard described in Section 1.4. For growing companies, the concern is timing. AI vendors are multiplying, employees are adopting tools on their own, and the board is beginning to separate from day-to-day management. The question is when to formalize, and the answer is when the gap between the company's AI exposure and its governance structure becomes a risk in itself. For private equity firms and directors serving on multiple boards, the concern is consistency. A PE executive sitting on four portfolio company boards carries fiduciary duties at each one, and a scalable approach requires a consistent baseline of expectations across the portfolio with formality calibrated to each company's size and risk profile.

The form of the program may vary. The oversight functions should be present. The legal standard does not require every company to build the same program. It requires every board to ensure a functioning system exists and is actively monitored. How much formality that system requires depends on how much risk the company faces.

2.2 The Maturity Map

Where does your organization stand today? The matrix below provides a self-assessment framework. It maps five dimensions of AI oversight against three *operating modes*: Founder-led, Growth, and Institutional.

These are descriptive categories, not a maturity progression. A Founder-led company operating at the appropriate mode for its risk profile is not “behind” an Institutional company. The question is fit: does the level of oversight formality match the level of AI risk exposure?

Company Profile	Owner-operator or small board. One or two AI tools deliberately adopted. Ambient AI embedded in everyday software, largely uninventoried.	Expanding AI use. Multiple vendors. Emerging need for formal structure. Board separating from management.	Complex AI portfolio. Significant vendor dependencies. Regulatory exposure. Formal governance required.
Oversight Program	Documented AI policy. Inventory of AI tools. Owner reviews quarterly.	Governance charter. Designated oversight responsibility. Defined risk appetite. Standing board agenda item.	Full program with committee or task force. Independent assurance. Formal escalation protocols. Regular reporting cycle.
Risk Management	<i>Crown Jewels</i> identified. Basic vendor due diligence. Acceptable use policy.	Risk appetite defined in financial terms. AI registry maintained. Vendor contracts reviewed for AI provisions.	Enterprise risk methodology applied to AI. Concentration risk monitored. Scenario testing. Insurance reviewed for AI coverage.
Board Reporting & Documentation	Informal. Owner-operator awareness. Annual review at minimum. Key decisions documented.	Standing agenda item. Quarterly AI risk and opportunity summary from management. Board minutes reflect AI oversight discussions.	Dedicated reporting dashboard. Incident tracking. Trend analysis. Management attestation. Formal record of board deliberations, decisions, and follow-up actions.
Director AI Literacy	Understands what AI tools the company uses and	Understands Builder/Buyer distinction. Can	Can challenge management representations.

	why. Can articulate basic risk exposure.	evaluate management's AI strategy. Recognizes key risk categories.	Understands regulatory landscape. Evaluates program effectiveness, not just activity.

Most organizations will not fit neatly into a single column. A company may have Institutional-level vendor dependencies but Founder-level governance structure. That mismatch is itself a finding, and closing it is the purpose of the sections that follow.

If you are uncertain where your organization belongs, start with two questions: How many AI systems or AI-enabled vendors does the company rely on? And what would happen to the business if the most critical one failed tomorrow? The answers will locate you on the map faster than any formal assessment. The column matters less than the questions it prompts. The point is the program, not the placement.

2.3 Director AI Literacy: What the Board Needs to Know

Directors are not expected to become technologists. They are expected to ask informed questions and evaluate management's answers.

What does that require? At minimum, directors should understand three things. First, what AI can and cannot do: the difference between a predictive model and a generative model, between a tool that recommends and an agent that acts. Precision is less important than directional understanding. Second, where value and risk concentrate: which AI systems touch the company's Crown Jewels, which vendors have access to sensitive data, and where a failure would have material consequences. Third, how to evaluate management's representations: when management says "we have AI governance," what questions distinguish a functioning program from a slide deck?

The literacy expectation scales with the operating mode. A Founder-led board needs to understand what AI tools are in use and why. A Growth-stage board needs to evaluate whether management's AI strategy is credible and whether the risk categories are understood. An Institutional board needs to challenge management's representations, understand the regulatory landscape, and evaluate program effectiveness rather than just program activity.

The Board-Level AI Oversight Questions in the Toolkit (Tool A) are designed to help directors at every level ask the right questions. They are not a test of technical knowledge. They are a test of whether the oversight program is functioning.

Part II

THE OVERSIGHT BLUEPRINT

Structuring Authority and Defining Risk

SECTION 3

Ensure the Program Has Authority and Accountability

A governance program that lacks the authority to require information, establish the balance of risk and reward within which the organization operates, and ensure critical risks and opportunities are reported to the board cannot fulfill its purpose. The structure matters less than whether the program has a documented mandate, clear accountability, and the capacity to function independently of management's willingness to cooperate.

This section addresses the structural requirements that give an AI oversight program the authority to function and the accountability to be verified.

3.1 The Structure: Beyond the “Committee Fallacy”

The most common governance failure is not the absence of a committee. It is the assumption that forming a committee constitutes governance.

A committee is a meeting. A program is a management system. The committee provides oversight of the program; the program provides the charter, policies, risk methodology, reporting protocols, and escalation triggers that make oversight possible. Without a program to oversee, the committee has nothing to evaluate and no basis for informed judgment and defensible oversight.

Where should AI oversight live? The answer depends on the organization's size and structure. For Founder-led companies, it may reside with the owner and one trusted advisor. For Growth-stage companies, it may be assigned to an existing committee (Audit or Risk) with an explicit charter amendment. For Institutional enterprises, it may warrant a dedicated AI oversight task force or a standing subcommittee. Organizations at any stage should also consider engaging outside expertise to supplement the board's knowledge. External advisors do not replace the board's fiduciary obligation, but they can materially strengthen the quality of oversight, particularly during the early stages of program development.

What matters more than the structure is fiduciary clarity: specific individuals must be assigned defined responsibilities within the AI oversight program. Identifying a responsible party is only the beginning. The program must also define the authorities, decision rights, resource commitments, reporting requirements, and escalation mechanisms necessary to ensure that identified risks are appropriately evaluated and addressed. If the board cannot answer "Who identifies issues?", "Who decides what action is required?", and "Who has the authority and resources to implement that action?", the program does not have authority. It has ambiguity. Effective governance requires a management system that converts awareness into action and action into accountability. Without those connections, oversight becomes procedural rather than operational, undermining the defensibility of the entire oversight system.

Committee charters must explicitly mention AI risk. If the charter is silent on AI, the oversight is presumed absent. Amending the charter is not a formality. It is the act that creates the legal record of the board's intent to oversee.

3.2 The Governance Charter: The License to Operate

The governance charter is the program's constitutional document. Without it, authority is assumed rather than granted, and the program lacks the defensibility that a documented mandate provides. The charter should establish:

Scope: what AI activities fall within the program’s purview, whether the program covers the entire organization or specific areas of critical AI dependency, and what thresholds trigger formal review. The charter defines where the line is and ensures that critical activities cannot bypass the program.

Authority: the program’s mandate to require information from management, to establish the risk and reward tolerances within which AI decisions are made, and to escalate concerns directly to the board. This includes strategic authority: the program should surface AI-related opportunities, not just risks.

Accountability: who owns the program at the management level, who reports to the board, and at what cadence. Reporting lines must be explicit. The board should know who will stand in front of them and answer for the program’s effectiveness. In private companies where the management layer is often thin, the same individual may carry responsibility for the program, its execution, and the first-line reporting about it. The charter should make clear whether that concentration is acceptable at the company’s current stage, and if so, what compensating oversight applies.

Risk and Reward Balance: the expectation that the program evaluates AI initiatives through both a risk lens and an opportunity lens. An AI opportunity typically looks like one of two things: a use case that enables the company to do something it could not do before (a new product line, a new service tier, a new customer segment that was previously uneconomic to serve), or a use case that changes the unit economics of something the company is already doing (lower cost, faster cycle time, or higher margin on an existing offering). Governance that focuses exclusively on downside protection will slow the organization without capturing either.

Cross-functional coordination: AI governance is not an IT function. It requires coordination across Legal, Technology, Human Resources, Operations, and Finance. Depending on the organization’s size, the charter should authorize or encourage this collaboration. AI risk does not respect organizational boundaries, and the governance structure should not either.

3.3 The Policy Framework: Establishing Tolerances

The board’s role is not to approve individual AI projects. That is management’s responsibility. The board’s role is to establish a program that defines the tolerances within which AI decisions can be made, the criteria for when decisions must be escalated, and the risk and reward balance that guides management’s execution.

The policy framework operationalizes these tolerances. It establishes clear boundaries that allow management to move quickly within defined parameters while ensuring that activities outside those parameters are escalated to the appropriate decision-maker in a programmatic way.

At minimum, the framework should include an acceptable use policy governing how employees interact with AI tools. This addresses confidential data (what can and cannot be entered into AI systems), client deliverables (when AI-generated content must be disclosed), attribution (how AI assistance is documented), and the use of AI-generated content in business decisions that affect customers, employees, or financial outcomes.

The framework should also establish clear escalation criteria: when an AI initiative moves from within-tolerance to requiring elevated review. Triggers include AI activities that touch customer data, affect regulated processes, involve autonomous decision-making, create a new vendor dependency, or interact with any of the company's Crown Jewels identified in Section 4.2.

The objective is a governance model that enables speed within defined boundaries. Management operates within the tolerances the program establishes. When an initiative falls outside those tolerances, the program defines who makes the decision and how quickly. This is not a constraint on innovation. It is a *decision architecture* that allows the organization to move with confidence.

3.4 Auditing the Program, Not Just the AI

A governance structure on paper that nobody follows provides no protection. A program that is never evaluated for effectiveness is a program that cannot be relied upon.

There is a critical distinction between auditing AI systems and auditing the governance program. Auditing AI systems is a technical discipline: testing models for bias, validating outputs, evaluating security controls. Auditing the governance program is an assurance discipline: verifying that the oversight structure is functioning, that tolerances are being observed, that reporting is reaching the board, and that the board is acting on what it receives.

Both are necessary. The lesson from cybersecurity is instructive. For years, organizations audited technical controls, vulnerability scans, penetration tests, compliance checklists, while neglecting the governance question: had the business first defined its tolerance for risk and reward, and then directed security teams to execute within those boundaries? These are business decisions first. When breaches occurred, the gap was rarely in the

technical controls. It was in the oversight system that never established whether those controls were adequate for the actual risk profile.

AI oversight must not repeat that pattern. The board should require periodic, independent evaluation of whether the governance program itself is working. Is the *AI Registry* current? Are tolerances being observed? Is management reporting reaching the board on schedule and with sufficient detail? Are escalation triggers functioning? These are governance assurance questions, and they are distinct from technical AI audits.

DIRECTOR'S FOCUS

What to take away

Section 3 is about whether the program has the authority to function. The board's job is to ensure that authority has been documented, that someone is accountable for program effectiveness, that the policy framework defines tolerances rather than approvals, and that the program itself is periodically evaluated for whether it actually works. Without these, a committee is a meeting and a charter is a document.

Questions the director should be asking:

- ◆ Does the governance charter specifically mention AI risk, and can it be produced on demand?
- ◆ Who is the named individual accountable for the program's effectiveness, and does every director know who that person is?
- ◆ When was the last independent evaluation of whether the program is functioning as designed?

SECTION 4

Define How Much Risk You Are Willing to Accept

Most boards can articulate that AI presents risk. Few have defined how much risk they are willing to accept, or how much opportunity they are willing to pursue. Without those definitions, management has no boundary within which to operate, and the board has no standard against which to evaluate.

4.1 Risk Appetite and Tolerance: The Multidimensional Yardstick

Risk appetite is the total amount of risk the organization is willing to accept in pursuit of its strategic objectives. *Risk tolerance* is the acceptable variation around specific risk categories. Both should be expressed in terms the board already uses, and both must account for the full spectrum of risk dimensions, not just financial impact.

Financial impact is the most prevalent consideration and often the most intuitive: impact on EBITDA, cash flow, enterprise valuation, and key operating metrics. Moving from vague “High/Medium/Low” heat maps to concrete financial thresholds forces specificity. But financial loss is one dimension among several that the board should define:

Operational risk: what is the tolerance for disruption to business processes that depend on AI?

Trust and reputation: what is the acceptable exposure to incidents that could damage customer trust, brand value, or competitive position?

Legal and regulatory: what is the tolerance for AI activities that could trigger enforcement actions, litigation, or regulatory scrutiny?

Each of these categories can translate to financial loss, but evaluating them only through a financial lens may understate the speed and severity of the impact.

The board’s risk appetite statement should address all of these dimensions in terms specific enough to guide management. Vague aspirations will not work. Concrete tolerances do: a defined ceiling on revenue exposure to any single AI dependency, a requirement that customer–data systems pass vendor due diligence before deployment, a standard that customer–facing AI requires defined human oversight criteria. The exact thresholds will vary by company. The discipline of expressing them explicitly is universal.

Materiality in a private company is shaped by who the company answers to. A PE-backed company’s materiality thresholds are influenced by LP reporting obligations and sponsor expectations that sit alongside the company’s own operational standards. A founder-led company’s enterprise valuation is often the founder’s personal wealth, which can compress

the appetite for reputational or regulatory risk in ways that would not apply to a similarly sized public company. The board's risk appetite statement should reflect these realities rather than treat the company as operating in a generic market.

Risk transfer belongs in this conversation as one consideration among many. D&O coverage, cyber insurance with AI-specific endorsements, and technology errors and omissions policies are increasingly relevant. A documented, functioning oversight program directly affects insurability and premium pricing. The board should understand what the company's current policies cover, what they exclude, and whether AI-related risks have been explicitly addressed.

Strategic risk deserves separate consideration. AI creates not only risks associated with deployment, but risks associated with inaction. The board should explicitly consider its tolerance for competitive disadvantage arising from delayed adoption, missed productivity opportunities, talent displacement, or business model disruption. In some circumstances, the greater risk may not be implementing AI too quickly, but implementing it too slowly.

The board should also consider external factors that can rapidly alter the organization's risk profile, including geopolitical developments, changes in regulatory regimes, supply chain dependencies, concentration among AI vendors, cybersecurity threats, and other systemic events. These factors often manifest through operational, legal, financial, or reputational impacts, but their potential to change assumptions abruptly warrants explicit consideration within the risk appetite discussion.

4.2 The “Crown Jewels” Assessment

Not all AI activities carry equal risk. A Crown Jewels assessment maps the company's most valuable assets, its intellectual property, customer data, brand reputation, key processes, and competitive advantages, against the AI systems that interact with them.

The purpose is focus. The board does not oversee individual AI tools. That is management's responsibility. The board oversees the program, and the program should prioritize based on where AI intersects with the assets that drive the company's valuation and exit multiple. The program does not need to treat every AI activity equally. It needs to be clear about which ones matter most and ensure those are governed with proportionate rigor.

The Crown Jewels that depend on AI are the “mission critical” risks referenced in the Caremark framework. They are the assets and processes where the absence of oversight creates the most significant liability exposure. The board should ensure that the program has identified what they are, which AI systems interact with them, and whether the controls

are proportionate to the risk. If the program cannot produce this mapping, the Crown Jewels assessment has not been done.

4.3 Risk Intelligence: Signal vs. Noise

The AI landscape generates more noise than signal. New models, new regulations, new vendor announcements, new enforcement actions, new competitive moves. Boards that try to monitor it directly face information overload. Boards that delegate it entirely face dangerous blind spots.

Risk Intelligence is the defined process that solves this. Its job is to filter the external landscape and surface what actually requires board awareness: regulatory changes affecting the company's sector, enforcement actions signaling shifting standards, competitor moves that change the strategic calculus, technology shifts that alter the risk surface, and new use cases from frontier AI models that could create applicable risks or strategic possibilities.

The governance charter should explicitly assign ownership of this function to a named individual, typically the Chief Risk Officer, General Counsel, or designated AI program lead. That person is accountable for maintaining the signal filter, continuously calibrating its inputs, and applying pre-defined escalation criteria that are documented, testable, and periodically reviewed. These criteria should include illustrative thresholds and examples. For instance: model capability changes that materially alter decision autonomy; regulatory actions affecting peers or critical vendors; incidents involving data leakage, bias, or safety failures; or vendor changes that impact system dependencies or control. A regulatory enforcement action against a peer company may warrant out-of-cycle communication. A vendor product update may not. The distinction should be pre-decided, not improvised.

Accountability in this context extends beyond reporting. The designated owner is responsible for ensuring that each escalated item is explicitly classified into one of three paths: (1) monitor, (2) management action, or (3) board-level decision.

Where escalation triggers board-level awareness, the charter should also define the corresponding expectation of response (acknowledgment, directive, or formal decision) and the timeframe in which that response is required.

Critically, the governance model must distinguish between responsibility for reporting risk and authority to act on it. If the designated owner does not control remediation resources or budget, the charter must identify the accountable executive who does, and define the handoff, decision rights, and escalation path if action is delayed or declined.

This ensures that no material signal results in passive awareness alone; each must produce a documented outcome, even if that outcome is a conscious decision not to act.

Risk Intelligence designed this way is also the mechanism by which the board demonstrates active monitoring of the oversight program. It is not a passive receipt of management summaries. It is a functioning system with defined ownership, defined triggers, and a documented record of what surfaced, when, and what the board did with it.

4.4 The International and Regulatory Landscape

International exposure is now a first-order governance topic in its own right. Many private companies face AI-related regulatory obligations well beyond their domestic jurisdiction, and the applicable landscape is both expanding and fragmenting.

Three dimensions are worth monitoring. Jurisdictional reach, meaning the territories where the company operates, sells, or employs. Cross-border data flows, meaning where customer, employee, and operational data is collected, stored, processed, and moved. Supply chain exposure, meaning where the company's AI vendors, their sub-processors, and the upstream model providers those vendors depend on are located. A company with no foreign customers can still inherit regulatory exposure through a vendor hosted on infrastructure in another country.

Beyond AI-specific law, tangential regimes often apply. Data protection, privacy, cybersecurity, and sector-specific consumer protection rules frequently reach AI activities even when the underlying statute predates AI.

The specific regimes vary, but the oversight themes regulators are converging on are consistent. Named accountability at the board or senior management level. Complete inventories of AI systems in use. Risk-proportionate governance, calibrated to the materiality of what the AI is doing. Third-party and concentration risk management, including for upstream providers. Meaningful human oversight by people with the competence and authority to intervene. A private company that has these in place is substantially aligned with what regulators are formalizing internationally, regardless of which specific regime applies.²

The Risk Intelligence function described in Section 4.3 is the operational mechanism for keeping the board informed as this landscape evolves. The board's role is not to track every regulatory development. It is to ensure that someone is tracking the developments that matter, and that a defined trigger exists to bring them forward when they do.

What to take away

Section 4 is about boundaries. The board sets tolerances that define where management can move on its own authority and where decisions must come up. Those tolerances have to cover the full spectrum of risk, not just financial loss. They have to be focused on the assets that matter most. And the program has to include a functioning mechanism for surfacing material shifts in both the strategic and regulatory landscape before the company is surprised by them.

Questions the director should be asking:

- ◆ What is our risk appetite for AI, expressed in terms specific enough that management knows when to escalate and we know when to intervene?
- ◆ Which of the company's Crown Jewels depend on AI, and is the oversight proportionate to what's at stake if any of them fail?
- ◆ Who owns Risk Intelligence in our program, and what are the pre-defined triggers that bring something to the board rather than leaving it in management's inbox?
- ◆ Where does our international exposure sit, and who is accountable for keeping the board informed about regulatory developments that affect it?

Part III

ASSET INTEGRITY AND OPERATIONS

What Directors Need to Verify About AI Systems and Vendors

SECTION 5

Verify That AI Systems Are Trustworthy, Secure, and Compliant

Asset Integrity means ensuring the AI the company relies on is a valuable asset rather than a latent liability. For Buyers, this is predominantly about verifying vendor claims and managing data exposure. For Builders, it extends to the full development lifecycle. In both cases, the board's role is to ensure the program is verifying, not to perform the verification.

5.1 Data Governance: Provenance and Exit Readiness

Who owns the data? Who owns the model? If the company trained an AI system on proprietary data, is that training data documented and the IP defensible? If the company is feeding proprietary data into a vendor's platform, what rights does the company retain?

These are not technical questions. They are valuation questions. Strict tracking of *data lineage* and *data provenance*, respectively ensures the company actually owns what it claims to own. Acquirers and investors conducting due diligence will evaluate data ownership and AI-related IP. Undocumented provenance creates valuation haircuts. Clean IP is not a compliance exercise. It is exit readiness.

The M&A exposure goes further than undocumented data. Acquirers conducting AI diligence will look for the governance infrastructure itself: is there an inventory of AI systems in use, have vendor contracts been reviewed for AI-specific provisions, does the company have documented risk tolerances and escalation protocols, and has anything material been escalated to the board that should have been? The absence of these artifacts is itself a finding, and findings become price negotiations. A private company planning toward a transaction benefits from approaching its AI governance the way it approaches financial audit readiness: well before the letter of intent, not during diligence.

Change-of-control exposure in vendor contracts deserves specific attention. Many AI vendor agreements contain provisions that can be triggered by an ownership change: automatic termination rights, renegotiation clauses, data-handling obligations, or model-retraining requirements. A company that depends on AI tooling it cannot guarantee will survive a transaction has a dependency that directly affects deal structure and valuation. The board should ensure the program has identified these provisions in the company's most material AI vendor relationships, and that the exposure is understood before a transaction is contemplated rather than discovered inside one.

For Buyers, the critical questions are: what data are you exposing to vendors, what rights do you retain, and what happens to your data if the vendor relationship ends? Vendor agreements should explicitly address data ownership, data portability, data deletion upon termination, and whether the vendor can use your data to train or improve its models for other customers, and whether that is an opt-in or opt-out function. If these terms are absent or ambiguous, the board should know.

5.2 The AI Registry: The Mandatory Inventory

You cannot govern what you cannot see. The AI Registry is a centralized inventory of every AI system and AI-enabled tool in use across the organization, whether purchased, built, or adopted informally by employees.

Shadow AI, unauthorized or untracked AI tools adopted by employees without governance review, is a primary threat to private companies. An employee using an AI tool to process

client data without approval creates risk that the board may not know exists until it materializes as a breach, a regulatory inquiry, or a due diligence finding. The pace at which new AI tools become available makes Shadow AI an ongoing challenge, not a one-time cleanup.

The Registry is not a directive issued by the board. It is what the program requires to provide the board with the visibility necessary for defensible oversight. Without a comprehensive inventory, the program cannot scope its coverage and the board cannot evaluate risk exposure. At minimum, the Registry should capture what AI systems are in use, who owns each system, what data each system accesses, which vendor provides it, and whether it has been reviewed under the policy framework described in Section 3.3. The Registry is also the instrument that answers the question every board should have already asked: what AI is already operating in this business that has not been formally reviewed?

5.3 Human Oversight, Fairness, and Labor Impact

When an AI system makes a consequential mistake, the question will not simply be whether a human was present. It will be whether the organization had established the criteria for when human involvement was required, whether those criteria were grounded in policy and practice, and whether the humans involved were trained and equipped to exercise meaningful judgment.

Human-in-the-Loop (HITL) and *Human-on-the-Loop (HOTL)* protocols remain essential controls. HITL requires human approval before action is taken. HOTL requires human monitoring with the ability to intervene. In practice, human oversight takes several forms depending on the stakes and cadence of the decision: pre-approval for high-consequence actions before they execute, confirmation steps that interrupt routine workflows when an AI output crosses a defined threshold, and periodic audit sampling of decisions that were made autonomously. The right form depends on the risk classification of the system, not on a single standard applied everywhere. The board's role is not to manage these protocols. It is to ensure the program has been established when human oversight is required: what constitutes a critical decision, what confidence level in the AI output warrants review, and whether the people performing oversight are prepared to do so.

Human oversight should not be evaluated solely by the presence of a person in the decision process. Research across aviation, healthcare, industrial automation, and other safety-critical domains demonstrates that human proficiency can degrade when highly reliable automated systems perform most tasks successfully. Over time, supervisors may become

less attentive, less practiced in exercising judgment, and less capable of recognizing or correcting failures when intervention is required.

Accordingly, organizations should periodically assess not only whether human oversight exists, but whether it remains effective. Oversight personnel should receive ongoing training, maintain sufficient familiarity with the underlying processes, and demonstrate the ability to identify, challenge, and override AI-generated outputs when appropriate. The objective is not merely to keep a human in the loop, but to ensure that the human remains capable of exercising independent and informed judgment.

AI systems that affect decisions about individuals, whether in employment, financial services, healthcare, customer interactions, or any other context where automated outputs have material consequences for people, carry an additional obligation: ensuring they do not produce discriminatory or harmful outcomes. Enforcement at the state level is expanding, and the liability exposure from AI-driven discrimination is significant. The board should ensure the program addresses this risk across all applicable areas of the business.

AI-driven workforce change raises a different board-level question than fairness in individual decisions. When a company adopts AI that displaces or restructures roles at scale, the board is not being asked to evaluate whether the change should happen. That is management's call. The board's role is to ensure management has done the full risk analysis, not just the business case.

The categories are familiar ones applied to a new context. Is the AI capability mature enough to reliably do the work being reassigned, or is the projected benefit dependent on performance that has not yet been demonstrated? Will institutional knowledge walk out the door before the replacement capability is proven? Does the change process comply with notice, severance, and anti-discrimination obligations across every jurisdiction where the company employs people? Can the remaining workforce function effectively after the change, and can the company's standing with customers, regulators, and talent markets absorb it? For private companies, these dynamics often land harder than they would at a larger public competitor. The company's workforce, customers, and community are typically closer to each other, and reputational effects travel faster through that proximity than through a quarterly earnings call.

A workforce transition built on projected cost savings alone is an incomplete plan. The board should expect the same rigor it would apply to any other strategic decision of comparable scale.

CASE IN POINT

iTutorGroup, August 2023

iTutorGroup used AI-driven recruitment software programmed to automatically reject female applicants aged fifty-five and older and male applicants aged sixty and older. More than two hundred qualified U.S.-based applicants were screened out on the basis of age. The Equal Employment Opportunity Commission sued under the Age Discrimination in Employment Act. On August 9, 2023, iTutorGroup entered a consent decree, paying \$365,000 in damages and back pay and adopting new anti-discrimination policies and training requirements.³ The EEOC has identified algorithmic discrimination as a continuing enforcement priority.

The lesson: federal enforcement against algorithmic discrimination is active, not theoretical. The defense that “we used the vendor’s tool” is not protective. Whoever deploys the system is accountable for its outputs and for the demonstrable bias those outputs encode.

5.4 The Agentic Frontier

Traditional AI oversight assumes a human is in the loop. The model produces an output, a person reviews it, a person acts. Agentic AI breaks that assumption. Autonomous agents can place orders, send communications, modify records, approve transactions, and initiate workflows that cut across Finance, Legal, HR, Operations, and Technology without waiting for a human to connect the dots.

The central governance problem is accountability. An agent cannot be accountable. Accountability is a human attribute, and when an agent acts, some identified person has to own the outcome. Without that assignment, the organization has deployed decision-making capacity faster than it has deployed responsibility for it.

The board’s role is not to approve individual agents. It is to ensure the program has established the conditions under which agents can be deployed at all. Those conditions fall into four areas.

Decision rights. Every agent should fall into one of three categories. Some agents are authorized to act autonomously within defined boundaries. Some are authorized to recommend but not to act. Some must escalate to a human for a decision. The program sets the category based on the materiality and reversibility of the action. A customer service agent issuing a refund below a defined threshold may be appropriate for autonomous action. An agent approving a supplier payment or modifying a contract is not. These tiers represent a program-level decision, not a case-by-case negotiation.

Named ownership. Shared or ambiguous ownership is not workable for agentic systems. Every deployed agent should have a named business owner accountable for outcomes, and a named technical owner accountable for how the system behaves. In the pre-agentic era, organizational silos served as informal checkpoints. Agents cross those silos by design. When they do, accountability has to be assigned explicitly, because organizational structure no longer assigns it on its own.

Hard boundaries. *Agency Limits* and *Circuit Breakers* are the constraints that define what an agent cannot do regardless of its reasoning. Categories of decisions outside its authority. Financial thresholds it cannot exceed. Data it cannot access. Automatic triggers that halt activity when boundaries are approached. These have to be enforced technically, not documented in a policy. The board should expect the program to verify that they function as designed.

Auditability. Every consequential action an agent takes has to be loggable and reconstructable after the fact. An organization that cannot show what an agent did, why it did it, and on whose authority cannot defend that decision to a regulator, a customer, a litigant, or an acquirer. Auditability is the precondition for every other form of oversight.

Enterprise-wide policy is the right scope for agentic AI regardless of operating mode. An agent that cuts across Finance, HR, and Operations does not respect the boundaries a Founder-led board might draw around its most visible AI systems. The Accordion Principle applies to how formally the program is documented, not to whether these four conditions have to be present. For agentic systems, they do.

CASE IN POINT

Air Canada, February 2024

A passenger asked Air Canada's website chatbot about bereavement fares after a death in the family. The chatbot told him he could apply for a discount within ninety days of travel. Air Canada's actual policy required pre-approval. When the passenger sought the refund, Air Canada refused, arguing in tribunal proceedings that the chatbot was a separate legal entity responsible for its own actions. The British Columbia Civil Resolution Tribunal rejected the argument, finding that the chatbot was part of Air Canada's website and that Air Canada owed a duty of care to ensure its representations were accurate. The airline was ordered to pay the refund.⁴

The lesson: agency without ownership is not a defense. Whoever deploys the agent is on the hook for what the agent says and what the agent does. The dollar amount in this case was modest; the precedent is not. Courts and tribunals across multiple jurisdictions are now applying ordinary principles of contract, agency, and negligent misrepresentation to AI outputs, with the company on the hook.

5.5 AI Security: Protecting the Asset

AI introduces security risks distinct from traditional IT. Prompt injection, model manipulation, data extraction through carefully crafted queries, and adversarial inputs that cause models to produce harmful outputs are all active threat categories. These risks exist on top of, not instead of, traditional cybersecurity threats.

For Buyer companies, which represent the majority of private boards, the critical security question is: can your vendor's AI be manipulated to expose the proprietary data you have entrusted to it? A vendor's AI security posture is your AI security posture if your data is in their system.

Five questions every director should ask about AI security posture:

- ◆ Has management identified the AI-specific attack vectors relevant to our systems and vendors?
- ◆ Are AI systems subject to the same security testing cadence as other critical infrastructure?
- ◆ Do vendor contracts address AI-specific security requirements, not just general cybersecurity terms?
- ◆ Is there a defined incident response protocol for AI security events, distinct from the general cybersecurity incident response plan?
- ◆ Has any AI system been tested for susceptibility to data extraction or prompt manipulation?

Third-party AI risk management deserves particular attention for Buyer organizations. The program should address the vendor's own AI governance and security practices, data handling and retention policies, contractual protections for data ownership and portability, indemnification for AI-generated outputs and autonomous actions, incident notification commitments, and the vendor's approach to model updates that could affect operations. A more detailed set of third-party contract considerations is provided in Appendix C.

What to take away

Section 5 is about whether the AI the company relies on is a trustworthy asset or a latent liability. For Buyers, that question is primarily about vendor behavior, data exposure, and what the program can verify. For Builders, it extends across the full development lifecycle. For every organization, it now extends to autonomous agents whose accountability cannot be inferred from organizational structure and whose actions cannot be overseen the way earlier generations of AI were overseen.

Questions the director should be asking:

- ◆ Does a comprehensive AI Registry exist, is it current, and does it include the systems employees adopted without going through governance?
- ◆ For autonomous agents operating in our business, do we have a named business owner and technical owner for each one, and can every consequential action be reconstructed after the fact?
- ◆ Where our data is in a vendor's system, do we understand what security testing has been done, what the vendor's own AI governance looks like, and what happens to our data if the relationship ends?

SECTION 6

Ensure AI Investments Are Strategically Sound and Vendor Risk Is Managed

For the majority of private boards operating as Buyers, this is where operational risk concentrates most heavily. The AI systems the company depends on were built by someone else, and the governance challenge is ensuring those dependencies are managed, not just purchased.

6.1 The Buyer's Dilemma: Vendor Concentration Risk

Dependency is not the same as a vendor relationship. A vendor relationship is a commercial arrangement. A dependency is a structural vulnerability: when a single provider's failure can halt business operations and impact the company's ability to serve customers, meet obligations, or maintain competitive position.

The board should understand where AI vendor concentration exists and how many critical processes depend on a single provider. The AI Registry (Section 5.2) should make these dependencies visible. If three different business functions all depend on the same AI vendor, the company has a concentration risk that may not be apparent when each function is evaluated independently.

The concentration question extends beyond the vendors the company contracts with directly. Many AI vendors are not the creators of the underlying models they provide. An application vendor may depend on one or more third-party foundation model providers, cloud platforms, data providers, or other critical infrastructure, and those dependencies are often obscured by contractual arrangements and invisible to customers.

As a result, a company may believe it has diversified its AI ecosystem when multiple vendors ultimately rely on the same underlying provider. A model provider's outage, policy change, pricing adjustment, geographic restriction, safety modification, or product retirement can affect numerous business processes simultaneously, even when the company holds contracts with several different vendors. Some vendors maintain a primary model and a fail-over model to preserve service delivery through an upstream outage; many do not.

Boards should therefore expect the program to seek visibility not only into direct vendor relationships, but into material upstream dependencies. The question is not simply "Who is our vendor?" but "What underlying technologies, providers, and decision-makers could materially affect our operations?"

A related private-company pattern deserves attention. In thinner management structures, the relationship with a critical AI vendor may be carried by one or two specific people: the person who selected the vendor, negotiated the contract, and has the direct working relationship with the vendor's account team. That concentration of institutional knowledge is itself a dependency. If the person leaves or the relationship changes, the company's practical control over the vendor relationship can degrade quickly. The board should expect the program to identify where vendor relationships are person-dependent and what the continuity plan looks like.

Vendor contracts deserve specific scrutiny for AI provisions. Standard technology contracts often fail to address AI-specific risks. The board should verify that contracts address:

- ◆ Data ownership and portability
- ◆ Indemnification for AI-generated outputs, including for autonomous agent actions
- ◆ Service level agreements that reflect AI-specific failure modes
- ◆ The vendor's right to modify or retrain models that affect your operations
- ◆ Liability allocation for decisions made or influenced by the vendor's AI

The detailed contract considerations referenced in Appendix C apply here as well.

Vendor lock-in, the inability to migrate away from a provider without prohibitive cost or disruption, is a strategic risk that deserves board-level visibility. If the cost of switching vendors exceeds the cost of staying with a problematic one, the company's negotiating position is compromised and its operational resilience is weakened.

6.2 The Unit Economics of AI: The Profitability Test

AI initiatives should be held to the same financial discipline as any other investment. The board should require management to: articulate business goals; outline resources needed to achieve success; and, demonstrate measurable, quantitative business impact, not just capability deployment.

A framework for evaluating AI investments should track ROI and Total Cost of Ownership (TCO), including licensing, integration, change management and training, ongoing support, and the less visible costs of data preparation, workflow redesign, and change management. Many AI projects look compelling in a vendor pitch but underperform when fully loaded costs are accounted for. The board's role is not to evaluate the technology. It is to require that management can demonstrate that AI spending is generating measurable return.

Equally important is the forward-looking investment question. In a landscape where competitors are actively investing in AI capabilities, standing still is itself a strategic risk. The board should evaluate not only whether current AI investments are generating returns, but whether the company is investing sufficiently to remain competitive. The question is not just "Is our AI investment profitable?" It is also "Are we investing enough to maintain our position, and are we making the right bets?" For a private company operating on direct cash discipline rather than public-market capital access, these questions carry a different weight than they do for a publicly-traded competitor. The answer cannot rely on the option

to raise equity on short notice. It has to be built into the company's operating plan. The board's role is not to perform the analysis. It is to require that management can demonstrate it, and to ensure the assumptions behind the numbers are challenged with the same rigor as any other significant investment decision.

6.3 Operational Resilience: What Happens When It Stops Working

If the company's most critical AI vendor experiences an outage that exceeds the downtime the board has defined as acceptable, can the business continue to operate? If the answer is no, or if the answer is unknown, that is itself a governance finding.

The board should require Business Continuity Planning (BCP) and Disaster Recovery (DR) for mission-critical AI dependencies, including:

- ◆ Defined fallback procedures for operating without the AI system
- ◆ Acceptable downtime thresholds established by the board, not management
- ◆ Tested recovery protocols validated through exercises
- ◆ Identification of which business processes have no manual alternative if the AI system fails
- ◆ Recovery time objectives for restoring AI-dependent operations

Operational resilience for AI dependencies is a senior management question, and depending on the severity and scope of the dependency, a board-level question. The board should know which AI systems are critical, what happens when they fail, and whether the contingency has been tested.

What to take away

Section 6 is where operational risk concentrates most heavily for Buyer boards. The AI the company depends on was built by someone else. The question is not whether to trust the vendor. It is whether the company understands its dependencies, has priced them into its financial discipline, and has a plan for what happens when a critical AI system stops working.

Questions the director should be asking:

- ◆ Where do we have vendor concentration we may not see, and are the contracts with those vendors addressing AI-specific risks?
- ◆ Are our AI investments generating measurable return, and are we investing enough in the right places to maintain competitive position?
- ◆ If our most critical AI vendor went down tomorrow, what happens, and has that scenario actually been tested?

Part IV

ESCALATION AND DISCLOSURE

Ensuring the Board Is Never the Last to Know

SECTION 7

Ensure Critical Risks and Emerging Opportunities Reach the Board Through Defined Channels

The most common oversight failure pattern is not the absence of a governance program. It is information known to management that never reaches the board, whether that information is a risk that has materialized, a tolerance that has been breached, or an opportunity that requires board-level decision-making.

Escalation is not just about incidents. It is how the program ensures the right information reaches the right decision-maker at the right time. A tolerance is not a “no.” It is a decision point. When an AI activity approaches or exceeds a defined tolerance, the protocol determines who evaluates, who decides, and how quickly. The decision may be to accept the risk, pursue the opportunity, or adjust the tolerance itself.

7.1 The Defined Escalation Protocols

The program must define what triggers escalation, who must be notified, by when, and through what channel. Escalation cannot depend on management's judgment about what the board "needs to know." It must be triggered by defined criteria.

Escalation triggers should include, but are not limited to:

- ◆ A risk that falls outside a defined tolerance, whether financial, operational, reputational, or regulatory
- ◆ A data breach involving AI systems or AI vendors
- ◆ An AI system producing outputs that result in material loss or regulatory exposure
- ◆ Discovery that a significant AI system was deployed without governance review
- ◆ A vendor failure affecting mission-critical operations
- ◆ Identification of a strategic opportunity related to AI that requires board-level evaluation or investment approval

Opportunities belong in the escalation framework alongside risks. A competitor's AI adoption that threatens the company's market position is both a risk and a signal. A new AI use case that could create competitive advantage but requires investment beyond management's authority is an opportunity that the board needs to see. The escalation protocol should surface both.

Triggers should be specific, measurable, and documented. The program should define thresholds that automatically require notification at the appropriate level, whether that is senior management, a committee, or the full board, based on severity and scope. Defined escalation protocols protect management as much as the board. They ensure that no individual is later exposed to having sat on material information when a clear protocol existed for surfacing it. The Boeing lesson from the Caremark line applies both ways: when management knows about a problem and the board does not, the board's legal defense collapses, and the individuals within management who knew become personally exposed as well.

Illustrative Escalation Events

The program's escalation triggers should include both risk events, where an activity is approaching or has exceeded a defined tolerance, and incident events, where something has already occurred. The following are illustrative of both categories. Each represents a type of event where the question of whether the board was notified, and how quickly, became material after the fact.

TOLERANCE-BASED TRIGGERS:

- ◆ An AI-dependent business process approaches the concentration threshold the board has defined for single-vendor exposure.
- ◆ An AI investment or deployment approaches the financial, operational, or reputational boundaries set in the program.
- ◆ A new use case, vendor integration, or autonomous capability falls outside the scope the policy framework was built to cover.
- ◆ A strategic opportunity emerges that would require investment, partnership, or capability expansion beyond management's defined authority.

INCIDENT-BASED TRIGGERS:

- ◆ A customer-facing AI produces outputs at scale that are biased, inaccurate, or misleading in ways that create contractual or regulatory exposure.
- ◆ A vendor pushes a model update that materially changes system performance, behavior, or cost without advance notice.
- ◆ An autonomous agent is discovered to have been operating outside its defined agency limits.
- ◆ A prompt-injection, model-manipulation, or data-extraction event exposes confidential data through an AI system or vendor.
- ◆ A significant AI system is discovered to have been deployed without formal governance review, and the scope of what it touches is not yet known.

The program should have pre-defined which of these are notified at the management level, which at the committee level, and which require immediate

full-board communication. The criteria should not depend on management's judgment about what the board wants to hear.

7.2 Materiality and Disclosure

Private companies do not have SEC filing obligations, but they face functionally equivalent disclosure requirements from the parties that fund, insure, and ultimately acquire them.

Disclosure obligations may arise from:

- ◆ **LP and investor agreements:** reporting requirements triggered by material events
- ◆ **Credit covenants:** notification requirements for events affecting financial condition or operational risk
- ◆ **Regulatory requirements:** sector-specific obligations under applicable law
- ◆ **Insurance policy terms:** notice requirements that, if unmet, may void coverage

The program should define objective materiality thresholds for AI-related events and map them against the company's existing disclosure obligations. A material AI event that triggers an insurance notification but is not reported may void coverage precisely when coverage is needed most. The board should understand which AI events could trigger disclosure requirements and ensure the escalation protocols in Section 7.1 are calibrated to those thresholds.

7.3 Crisis Readiness: The “Break Glass” Plan

Plans that have never been tested are assumptions, not plans.

The board should require that management has practiced its response to an AI failure event before one occurs. Tabletop exercises simulating realistic scenarios, a model producing harmful outputs, a data breach through an AI vendor, a regulatory inquiry triggered by an AI-driven decision, test whether the escalation protocols work, whether the right people are notified, and whether the response is fast enough.

Effective crisis exercises should surface answers to questions that planning alone cannot resolve:

- ◆ Who has the authority to shut down an AI system in an emergency?
- ◆ How quickly can the company revert to manual processes for critical functions?

- ◆ Who communicates with customers, regulators, the media, and other stakeholders, including LPs, sponsors, lenders, and other parties whose agreements may require notification?
 - ◆ Has legal counsel been briefed on AI-specific liability exposure?
 - ◆ Are escalation protocols functioning as designed under time pressure?
- These are questions best answered in a simulation, not during an actual crisis. The board should require evidence that the plan has been tested, not merely drafted.

The kill switch deserves specific attention because it is the question boards most often assume has been answered and most often find has not been. Circuit Breakers, discussed in Section 5.4, are preventive: the hard-coded constraints that keep an autonomous system from crossing a boundary it should not cross. A kill switch is reactive: the ability to stop a system that is already causing harm. The board should expect the program to have verified that for every critical AI system and every autonomous agent, a specific person has the authority to shut it down, the technical means to do so actually exists and has been tested, and the downstream consequences of that shutdown are understood. A kill switch that has never been tested is an assumption. A kill switch that requires coordinating with a vendor under the terms of a support contract is not a kill switch. A kill switch that creates more damage than it prevents is not a usable control. The board does not need to know which button to press. It needs to know that the program has confirmed the button exists, the right person can press it, and the company has thought through what happens after.

What to take away

Section 7 is about whether the board is structurally positioned to act on what it knows. Escalation protocols surface both risks that have breached tolerance and opportunities that require board-level decision-making, on defined triggers rather than management discretion. Materiality thresholds align the program with the disclosure obligations the company has already taken on through its investors, lenders, and insurers. Crisis readiness ensures that when an AI failure happens, the response is something the company has practiced rather than something it is inventing under pressure.

Questions the director should be asking:

- ◆ What are the defined triggers that require immediate notification of the board, and when was the last time one of them actually fired?
- ◆ Where do our LP, credit, and insurance agreements create disclosure obligations that could be triggered by an AI event, and are our escalation thresholds calibrated to them?
- ◆ For every mission-critical AI system and autonomous agent, does a tested kill switch exist, and does a named person have the authority and the means to use it?

The Standard

The body of this document has made a single argument in several forms. AI oversight is not a committee. It is a program. A program is a functioning system that defines who is accountable, what the tolerances are, when decisions must be escalated, and how the board knows whether any of it is actually working. The committee is the oversight mechanism. The program is what the committee oversees.

The private director's responsibility is to ensure that the program exists and functions. Not to build it, not to run it, and not to become an AI expert. The board is accountable for the conditions under which informed decisions can be made. That is the standard Caremark established in a different industry three decades ago, and it is the standard that courts, investors, insurers, and acquirers will apply to AI oversight in whatever jurisdiction the company operates.

Most private companies do not have the program they need. Some have a committee without a program. Some have a policy without a charter. Some have a charter without accountability. A smaller number have nothing at all. For any of these, the next board meeting where AI comes up is the moment to start closing the gap. The questions in this document are a starting point. The Toolkit that follows provides the artifacts. The appendices provide the source material. The work itself belongs to the board.

AI is moving faster than the governance structures built to oversee it. That is not an argument for waiting until the landscape settles. It is an argument for building the oversight program a private board can stand behind today, and committing to the discipline of keeping it current as the technology, the law, and the competitive landscape keep moving.

Part V

THE DIRECTOR'S TOOLKIT

Actionable Scripts and Tools

This Toolkit collects artifacts designed to be used immediately. Each tool is one page. Each is intended to be brought into a board meeting or handed to management with a clear expectation attached.

Contents

Part V. The Director's Toolkit

Tool A. Board-Level AI Oversight Questions. Fifteen questions, grouped by section, that distinguish informed oversight from performative governance. Designed to be sequenced across multiple board meetings.

Tool B. Governance Maturity Diagnostic. A descriptive self-assessment mirroring the Maturity Map in Section 2.2. Lets a board locate the organization across five dimensions (Company Profile, Oversight Program, Risk Management, Reporting, Director Literacy) and identify where the dimensions are out of alignment.

Tool C. Board Mandate Matrix. A RACI-style table assigning AI oversight roles across the Board, Management, and the Program Owner. Resolves ambiguity before ambiguity becomes a governance failure.

Tool D. Vendor AI Due Diligence Questionnaire. A structured questionnaire for the selection stage, organized across governance, data, model behavior, security, resilience, and contractual protections. Sits upstream of the contract terms in Appendix C.

Appendices

Appendix A. AI Fundamentals for Directors. A brief orientation to the categories of AI most commonly encountered in business today, written for directors without a technical background. Covers rules-based systems, machine learning, deep learning, generative AI and large language models, multimodal AI, agentic AI, and a note on AGI.

Appendix B. The AI Oversight Program: An Illustrative Model. One illustrative model of how the principles in this Body of Knowledge can be operationalized, developed by the Center for AI Oversight and presented at the Pillar and Domain level, with a concordance table mapping each section of the BoK to the corresponding Pillar and Domain.

Appendix C. Third-Party Contract Considerations. A structured set of provisions for AI vendor agreements, covering data ownership, model behavior, AI-specific security, liability allocation, change of control, sub-processors, audit rights, and regulatory alignment. The contract-stage companion to Tool D.

Back Matter

Glossary. Defined terms used in this Body of Knowledge, with citations to authoritative external sources (NIST AI RMF, OWASP Top 10 for LLM Applications, ISO 31000, Caremark line cases) where applicable. Terms coined in the BoK itself are defined in the section where they first appear and are not repeated.

Notice and Disclaimer. Standard publication notice covering the limits of the document: not legal or professional advice, no attorney-client relationship, citations accurate as of publication date, IP terms for Appendix B, and the contributors' disclaimer of liability for reliance.

Tool A. Board-Level AI Oversight Questions

Fifteen questions that distinguish informed oversight from performative governance

These questions are designed for the next board meeting. They are not a test of technical knowledge. They are a test of whether the oversight program is functioning. A board that cannot answer them should treat the gap as the most important finding from the conversation.

The questions are grouped by section so directors can sequence them across multiple meetings. Three to five at a time is enough to surface whether the program is real.

SECTION 1. THE FIDUCIARY PIVOT

- 1 Does the board have a credible view of the company's AI strategy, the AI footprint that already exists in the business, and the gap between what the company faces and what the board currently sees?
- 2 Are we a Builder, a Buyer, or a hybrid, and is our oversight attention allocated accordingly?
- 3 If a Caremark-style inquiry happened tomorrow, could we produce evidence that a functioning oversight system exists and that we are actively monitoring it?

SECTION 3. PROGRAM AUTHORITY AND ACCOUNTABILITY

- 1 Does the governance charter specifically mention AI risk, and can it be produced on demand?
- 2 Who is the named individual accountable for the program's effectiveness, and does every director know who that person is?
- 3 When was the last independent evaluation of whether the program is functioning as designed, distinct from any technical audit of AI systems?

SECTION 4. RISK APPETITE AND TOLERANCES

- 1 What is our risk appetite for AI, expressed in terms specific enough that management knows when to escalate and we know when to intervene?

- 2 Which of the company's Crown Jewels depend on AI, and is the oversight proportionate to what's at stake if any of them fail?
- 3 Who owns Risk Intelligence in our program, and what are the pre-defined triggers that bring something to the board rather than leaving it in management's inbox?

SECTION 5. ASSET INTEGRITY

- 1 Does a comprehensive AI Registry exist, is it current, and does it include the systems employees adopted without going through governance?
- 2 For autonomous agents operating in our business, do we have a named business owner and a named technical owner for each one, and can every consequential action be reconstructed after the fact?
- 3 Where our data sits in a vendor's system, do we understand what security testing has been done, what the vendor's own AI governance looks like, and what happens to our data if the relationship ends?

SECTION 6. VENDOR RISK AND STRATEGIC SOUNDNESS

- 1 Where do we have vendor concentration we may not see, and do the contracts with those vendors address AI-specific risks?
- 2 Are our AI investments generating measurable return, and are we investing enough in the right places to maintain competitive position?

SECTION 7. ESCALATION AND CRISIS READINESS

- 1 For every mission-critical AI system and autonomous agent, does a tested kill switch exist, and does a named person have the authority and the means to use it?

Tool B. Governance Maturity Diagnostic

A self-assessment to locate the organization on the Maturity Map

This diagnostic mirrors the Maturity Map in Section 2.2. It is designed to be completed by the board, or by the board and management together, in roughly fifteen minutes. For each dimension, mark the column that best describes the organization today. There are no right answers. The purpose is to locate the organization honestly, not to produce a score.

Reading the result: a consistent pattern across columns indicates the organization is operating at one mode. A mixed pattern, where some dimensions sit Institutional and others sit Founder-led, is the more common case and is itself the most important finding. The dimensions that lag behind the others are where governance attention belongs.

This is a descriptive instrument, not a scoring system. For a deeper, scored assessment with benchmarking against peer organizations, boards should consider engaging an outside advisor. Independence matters: the assessment should be conducted by someone with no role in operating the program being evaluated, so that the findings carry the weight of an outside view rather than the self-report of the people whose work is under review.

Company Profile	Owner-operator or small board. One or two AI tools deliberately adopted. Ambient AI embedded in everyday software, largely uninventoried.	Expanding AI use across multiple vendors. Board separating from day-to-day management.	Complex AI portfolio. Significant vendor dependencies. Regulatory exposure.
Oversight Program	Documented AI policy. Inventory of AI tools. Owner reviews quarterly.	Governance charter. Designated oversight responsibility. Defined risk appetite. Standing board agenda item.	Full program with committee or task force. Independent assurance. Formal escalation protocols.
Risk Management	Crown Jewels identified. Basic vendor due diligence. Acceptable use policy in place.	Risk appetite defined in financial terms. AI Registry maintained. Vendor contracts reviewed for AI provisions.	Enterprise risk methodology applied to AI. Concentration risk monitored. Scenario testing.
Board Reporting and Documentation	Informal. Owner-operator awareness. Annual review at minimum. Key decisions documented.	Standing agenda item. Quarterly AI risk and opportunity summary from management. Board minutes reflect AI oversight discussions.	Dedicated reporting dashboard. Incident tracking. Trend analysis. Management attestation. Formal record of board deliberations and follow-up actions.
Director AI Literacy	Understands what AI tools the company uses and why. Can articulate	Understands Builder and Buyer distinction. Can evaluate	Can challenge management representations. Understands

Current State	Current State	Current State	Current State
	basic risk exposure.	management's AI strategy. Recognizes key risk categories.	regulatory landscape. Evaluates program effectiveness, not just activity.

After completing the diagnostic, ask two follow-up questions. First, where is the gap between the current state and where the organization should be given its actual AI exposure? Second, what would close that gap in the next two board cycles? The answers point to where the program needs to develop next, and they belong in the board's documented oversight record.

Tool C. Board Mandate Matrix

A RACI-style template assigning AI oversight roles

The Mandate Matrix clarifies who approves, who executes, and who monitors each function of the AI oversight program. It is designed to be completed by the board and management together. The output is a one-page document that resolves ambiguity before ambiguity becomes a governance failure.

The matrix uses three roles. The Board ensures the function exists and is functioning, and approves where the function requires it. Management owns execution within the boundaries the program establishes. The Program Owner is the named individual (per Section 3.1) accountable for the program's effectiveness; in smaller organizations this role may be carried by an existing executive.

Conventions: A means accountable (the role that decides or approves). R means responsible (the role that does the work). M means monitors (the role that verifies the work is being done). One A per row. R and M may be shared. If a row has no A, the function is unowned.

Approve governance charter and charter amendments	A	R	R
Set risk appetite and tolerance across financial, operational, reputational, and regulatory dimensions	A	R	R
Identify Crown Jewels and map AI dependencies against them	M	A	R
Maintain the AI Registry, including Shadow AI	M	R	A
Approve policy framework and escalation thresholds	A	R	R
Operate within tolerances and execute the policy framework	M	A	M
Own Risk Intelligence and apply pre-defined escalation criteria	M	R	A
Decide deployment of autonomous agents above defined materiality	M	A	R
Assign named business and technical owners for each autonomous agent	M	A	R
Select vendors for mission-critical AI dependencies	M	A	R
Review AI-specific provisions in vendor contracts	M	R	A
Test Business Continuity and kill switches for critical AI systems	M	A	R

AI Oversight Task	Board of Directors	AI Oversight Task Force	AI Oversight Committee
Receive standing AI oversight reports on defined cadence	A	R	R
Commission independent evaluation of program effectiveness	A	R	M
Conduct AI crisis tabletop exercises	M	A	R

A Accountable: decides or approves **R** Responsible: does the work

M Monitors: verifies it is done

A: Accountable · R: Responsible · M: Monitors

This matrix is a starting point. Boards should adapt it to the organization's operating mode. For Founder-led companies, the Program Owner role may be carried by the owner and one trusted advisor. For Institutional enterprises, it may be a dedicated AI oversight task force. What does not change is that every row has a named A.

Tool D. Vendor AI Due Diligence Questionnaire

A structured questionnaire for evaluating third-party AI vendors

For Buyer companies, the vendor's AI security posture is the company's AI security posture if company data is in the vendor's system. This questionnaire is designed for use during vendor selection and at contract renewal. The board is not expected to administer it. The board is expected to ensure the program does.

A vendor that cannot or will not answer these questions is a finding in itself. The detailed contract considerations referenced in Appendix C apply once a vendor is selected; this tool sits upstream of the contract.

GOVERNANCE AND ACCOUNTABILITY

- 1 Does the vendor have a documented AI governance program, and can it be produced?
- 2 Who at the vendor is named accountable for AI risk, and what is their reporting line?
- 3 Has the vendor completed an independent assessment of its AI governance, and what were the findings?

DATA HANDLING AND PROVENANCE

- 1 What customer data does the vendor's AI ingest, store, process, or retain?
- 2 Can the vendor use our data to train, fine-tune, or improve its models for other customers? If so, on what terms can we opt out?
- 3 Where is our data stored and processed, including the jurisdictions of any sub-processors?
- 4 On termination, how is our data returned or deleted, on what timeline, and how is the deletion verified?

MODEL GOVERNANCE AND BEHAVIOR

- 1 How are the AI models tested for bias, accuracy, and security before deployment, and on what cadence after?

- 2 How and when does the vendor update or retrain models, and what notice will we receive of changes that could affect our operations?
- 3 Does the vendor offer explainability or audit trails for consequential outputs, and to what level of detail?
- 4 For autonomous capabilities, what controls limit what the system can do without human review?

SECURITY

- 1 Has the vendor's AI been tested for prompt injection, model manipulation, and data extraction, and can results be shared?
- 2 How is access to our data segregated from access to other customers' data?
- 3 What is the incident notification timeline for AI-specific security events, distinct from general cybersecurity events?

RESILIENCE AND CONTINUITY

- 1 What is the vendor's stated uptime commitment, and what is the historical performance against it?
- 2 What is our exit plan, including data portability, transition support, and the realistic timeline to switch providers?
- 3 Does the vendor's contract contain change-of-control provisions that could be triggered by a transaction on either side?

CONTRACTUAL PROTECTIONS

- 1 Will the vendor indemnify for AI-generated outputs and for actions taken by autonomous agents operating in our environment?
- 2 Are service level agreements specific to AI failure modes, not just general availability?
- 3 Does the contract address sub-processor disclosure and our right to audit?

Appendix A. AI Fundamentals for Directors

A brief orientation, written for the director who does not need to become a technologist

AI is not new. The field dates to the 1950s, and the underlying ideas (machines that perform tasks normally requiring human intelligence) have moved through several generations of capability and several cycles of optimism and disappointment. What has changed in the last few years is not the existence of AI, but the capability and accessibility of a particular branch of it. Directors should understand the categories well enough to ask informed questions about which the company is using, where each one creates value, and where each one creates risk.

This appendix gives the orientation. It does not teach the technology. A director who reads it should be able to follow the rest of the BoK without feeling lost. None of what follows requires technical background.

Rules-Based Systems and Expert Systems

The original form of AI, and still in use. These are systems built on explicit rules written by people: "if X, then Y." They do not learn from data. Tax software, basic chatbots, and most rule-driven automation fall into this category. They are predictable, auditable, and limited to what their rules express.

Governance implication: low risk in most uses, because the behavior is what the rules say. The risk is that the rules become outdated, or that the system is described as "AI" in a way that overstates what it does.

Machine Learning

Systems that learn patterns from data rather than from explicit rules. A spam filter that improves as it sees more email, a credit model that predicts default risk, a fraud detection system that flags unusual transactions, a recommendation engine that learns what a customer likes. These are the workhorse AI systems most companies have been using for a decade or more.

Governance implication: bias and fairness, model drift, and the question of whether the training data is representative of the population the model is applied to. Most state-level AI regulation is aimed primarily at this category.

Deep Learning

A subset of machine learning that uses layered neural networks to learn patterns too complex for simpler statistical models to capture. Deep learning powers image recognition, speech recognition, language translation, and the underlying technology behind modern generative AI. The distinction between machine learning and deep learning matters less to a board than the use case the system serves.

Governance implication: deep learning systems are often less explainable than classical machine learning. When the board asks "why did the system produce that output," the answer may be unavailable in the form a regulator or a customer would accept.

Generative AI and Large Language Models

Generative AI refers to systems that produce new content rather than classifying or predicting from existing data. Large language models (LLMs) are the dominant form of generative AI today: systems trained on enormous volumes of text that produce coherent, contextually appropriate written output. ChatGPT (OpenAI), Claude (Anthropic), Gemini (Google), and the AI capabilities embedded in productivity software, customer service tools, marketing platforms, and developer environments are all examples.

Governance implication: this is the category most companies are now adopting at scale, often through vendor-embedded capabilities. The risks include hallucination (the model produces confident output that is factually wrong), intellectual property exposure (training data may include copyrighted material; outputs may resemble protected works), data leakage (proprietary information entered as input may be retained or used to train future models), and attribution (when AI-generated content is used in business decisions or customer-facing material).

Multimodal AI

Systems that process more than one type of input or output: text, images, audio, video, code, and structured data. Multimodal AI is now the default capability in major commercial models. A board should assume that any modern AI vendor offering text capability also offers image and code capability.

Governance implication: the expansion of modes expands the attack surface and the exposure surface. Image-based prompt injection, audio deepfakes, and visual document manipulation are categories that did not exist when the company's data handling policies were written.

Agentic AI

AI systems that take actions, not just produce outputs. An agent can place an order, send a message, modify a record, schedule a meeting, approve a transaction, or initiate a workflow without waiting for a human to do it. The agent is given a goal, makes decisions about how to pursue it, and acts. Agentic AI is the current frontier of commercial AI deployment.

Governance implication: this is the category Section 5.4 of the BoK addresses in detail. The shift from "recommend" to "act" requires the four conditions discussed there: decision rights, named ownership, hard boundaries, and auditability. Most companies do not yet have these conditions in place; many are deploying agents anyway.

Artificial General Intelligence (AGI)

A hypothetical AI system with the general cognitive capabilities of a human across most domains. AGI does not currently exist in any commercial system. Whether it is achievable, on what timeline, and what it would mean if it arrived, are contested questions even among AI researchers. AGI is not a useful framework for board-level oversight decisions today. The category exists in this glossary only because directors are likely to encounter the term and should know what it does and does not refer to.

Governance implication: none direct. The risk a board should care about is current-generation AI behaving in ways the company did not anticipate, not a future general intelligence behaving in ways no one anticipates.

How These Categories Combine

In practice, a single commercial AI product often spans several of these categories. A modern customer-service platform may use classical machine learning to route tickets, generative AI to draft responses, multimodal AI to process images attached to tickets, and agentic AI to take action on routine cases without human review. The board's job is not to disaggregate the technology. It is to understand which capabilities are at work in which parts of the business, where each creates value, and where each creates exposure that the governance program needs to address.

Glossary

Defined terms used in this Body of Knowledge

This glossary covers the industry and legal terms that appear in this document. Terms coined in the BoK itself (the Velocity Gap, the Accordion Principle, the Committee Fallacy, Crown Jewels, Builders and Buyers, operating modes, decision architecture) are defined in the section where they first appear and are not repeated here.

Citations are kept brief. Where a term has an authoritative external definition, the source is noted parenthetically. Case citations for the Case in Point examples appear in the Endnotes.

AI TECHNOLOGY TERMS

Agent / Agentic AI. An AI system designed to take actions in pursuit of goals, rather than only producing outputs for human review. Agents can place orders, send communications, modify records, and initiate workflows without waiting for human action. See Section 5.4. (OWASP, LLM06:2025, *Excessive Agency*.)

Artificial Intelligence (AI). An engineered or machine-based system that, for a given set of objectives, generates outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy. (NIST AI RMF 1.0, 2023, adapted from OECD and ISO/IEC 22989:2022.)

Artificial General Intelligence (AGI). A hypothetical AI system with broad cognitive capabilities comparable to a human across most domains. AGI does not exist in any current commercial system. The term is contested among researchers and is not a useful frame for board-level oversight today.

Deep Learning. A subset of machine learning that uses layered neural networks to learn complex patterns from data. Deep learning underlies most modern image recognition, speech recognition, and the language models behind generative AI.

Generative AI. AI systems that produce new content such as text, images, audio, video, or code, rather than only classifying or predicting from existing data. Most consumer-facing AI tools introduced since 2022 are generative. (*NIST AI 600-1, Generative AI Profile, 2024.*)

Hallucination. Confident, fluent output from a generative AI system that is factually incorrect or fabricated. Hallucinations are a property of how large language models produce text, not a malfunction that can be fully eliminated. (*OWASP, LLM09:2025, Misinformation.*)

Large Language Model (LLM). A generative AI model trained on large volumes of text, capable of producing contextually appropriate written output. LLMs are the underlying technology for modern AI assistants, document drafting tools, and most current customer-facing generative AI applications.

Machine Learning. An approach to AI in which systems learn patterns from data rather than from explicit rules written by people. Most predictive, classification, and recommendation systems in commercial use are forms of machine learning.

Model Drift. Degradation in the performance of an AI model over time as the data it encounters in production diverges from the data it was trained on. Model drift is a category of risk that requires ongoing monitoring, not one-time validation. (*NIST AI RMF 1.0, Manage function.*)

Multimodal AI. AI systems that process more than one type of input or output, such as text combined with images, audio, video, or code. Multimodal capability is now standard in major commercial AI models.

AI SECURITY TERMS

Data Extraction. An attack in which an adversary uses carefully constructed queries to cause an AI system to reveal training data, system prompts, or proprietary information embedded in its context. (*OWASP, LLM02:2025, Sensitive Information Disclosure.*)

Model Manipulation. Adversarial techniques that cause an AI model to behave in ways its developer did not intend, including data poisoning during training and adversarial inputs

during use. (OWASP, LLM04:2025, *Data and Model Poisoning*.)

Prompt Injection. An attack in which user input alters the behavior of a large language model in unintended ways. Direct prompt injection comes from the user; indirect prompt injection comes from content the model reads, such as a document, web page, or image, that contains hidden instructions. (OWASP, LLM01:2025, *Prompt Injection*.)

OVERSIGHT AND RISK TERMS

Agency Limits and Circuit Breakers. Hard boundaries on what autonomous AI systems are authorized to do, enforced technically rather than only documented in policy. Agency Limits define the scope of action; Circuit Breakers are automatic triggers that halt activity when boundaries are approached. See Section 5.4.

AI Lifecycle. The stages through which an AI system passes, from plan and design, through data and model development, to deployment, operation, monitoring, and decommissioning. Risk management practices are typically organized around the lifecycle. (NIST AI RMF 1.0; OECD Framework for the Classification of AI Systems.)

AI Registry. A centralized inventory of every AI system and AI-enabled tool in use across the organization, whether purchased, built, or adopted informally by employees. The Registry is the foundation of any defensible oversight program. See Section 5.2.

Human-in-the-Loop (HITL). An oversight model in which human approval is required before an AI system takes a consequential action. (NIST AI RMF 1.0, *Govern and Manage functions*.)

Human-on-the-Loop (HOTL). An oversight model in which a human monitors the operation of an AI system and retains the ability to intervene, but the system acts without prior approval for each action. (NIST AI RMF 1.0, *Govern and Manage functions*.)

Risk Appetite. The total amount of risk an organization is willing to accept in pursuit of its strategic objectives. See Section 4.1. (ISO 31000:2018; adapted in NIST AI RMF 1.0.)

Risk Tolerance. The acceptable variation around specific risk categories within the overall risk appetite. Risk tolerance is context-specific and changes over time as conditions evolve. (NIST AI RMF 1.0; ISO Guide 73.)

Shadow AI. Unauthorized or untracked AI tools adopted by employees without governance review. Shadow AI is a primary source of unmeasured risk in private companies. See Section 5.2.

LEGAL AND GOVERNANCE TERMS

Caremark Doctrine. The legal standard, originating in Delaware, that holds directors to an affirmative duty to ensure the corporation maintains adequate information and reporting systems for mission-critical risks. The duty has two prongs: ensuring a system exists, and actively monitoring it. See Section 1.4. (*In re Caremark International Inc. Derivative Litigation*, 698 A.2d 959 (Del. Ch. 1996).)

Duty of Oversight. The board's affirmative obligation to ensure the corporation has information and reporting systems sufficient to bring material risks to the board's attention. Under Delaware law, sustained failure of oversight is treated as a breach of the duty of loyalty. (*Stone v. Ritter*, 911 A.2d 362 (Del. 2006).)

Materiality. The threshold at which an event becomes significant enough to require disclosure or action. For private companies, materiality is typically defined by the agreements that govern the company's relationships with investors, lenders, regulators, and insurers, rather than by SEC standards. See Section 7.2.

Mission-Critical Risk. A risk category central enough to the company's operations or value that the board's failure to establish oversight for it may itself constitute a breach of duty. The standard was articulated for a private, family-controlled company in *Marchand v. Barnhill*. (*Marchand v. Barnhill*, 212 A.3d 805 (Del. 2019).)

Two-Prong Test. The framework, codified in *Stone v. Ritter*, under which a Caremark claim may proceed: (1) the directors utterly failed to implement any reporting or information

system; or (2) having implemented such a system, consciously failed to monitor or oversee its operations. See Section 1.4. (*Stone v. Ritter*, 911 A.2d 362 (Del. 2006).)

REFERENCE

Endnotes

Sources for the numbered notes in the text

1. The account of the OpenAI board events is drawn from OpenAI's public announcements and contemporaneous reporting, November 17 to 21, 2023.
2. The convergence is visible across the EU AI Act, the UK FCA and PRA regime, Canada's OSFI Guideline E-23, Singapore's MAS AI Risk Management Guidelines, Australia's APRA and ASIC supervisory expectations, and IOSCO's 2025 work on artificial intelligence in capital markets. Agentic AI is the open question across all of them; no jurisdiction has settled rules yet.
3. *EEOC v. iTutorGroup, Inc.*, No. 1:22-cv-02565 (E.D.N.Y.), consent decree filed August 9, 2023; see the EEOC's announcement of the settlement.
4. *Moffatt v. Air Canada*, 2024 BCCRT 149 (British Columbia Civil Resolution Tribunal, February 14, 2024).

Appendix B. The AI Oversight Program

An Illustrative Model from the Center for AI Oversight

This appendix presents one illustrative model of how the principles in this Body of Knowledge can be operationalized: the AI Oversight Program developed by the Center for AI Oversight. The Program is structured as five Pillars and fourteen Domains, and is designed as the interoperability layer between external regulatory standards, fiduciary duties, and the organization's internal AI activities. Organizations may adopt alternative architectures appropriate for their size, sector, and AI maturity. The critical question is whether the oversight functions described in this Body of Knowledge are present, not what label they carry.

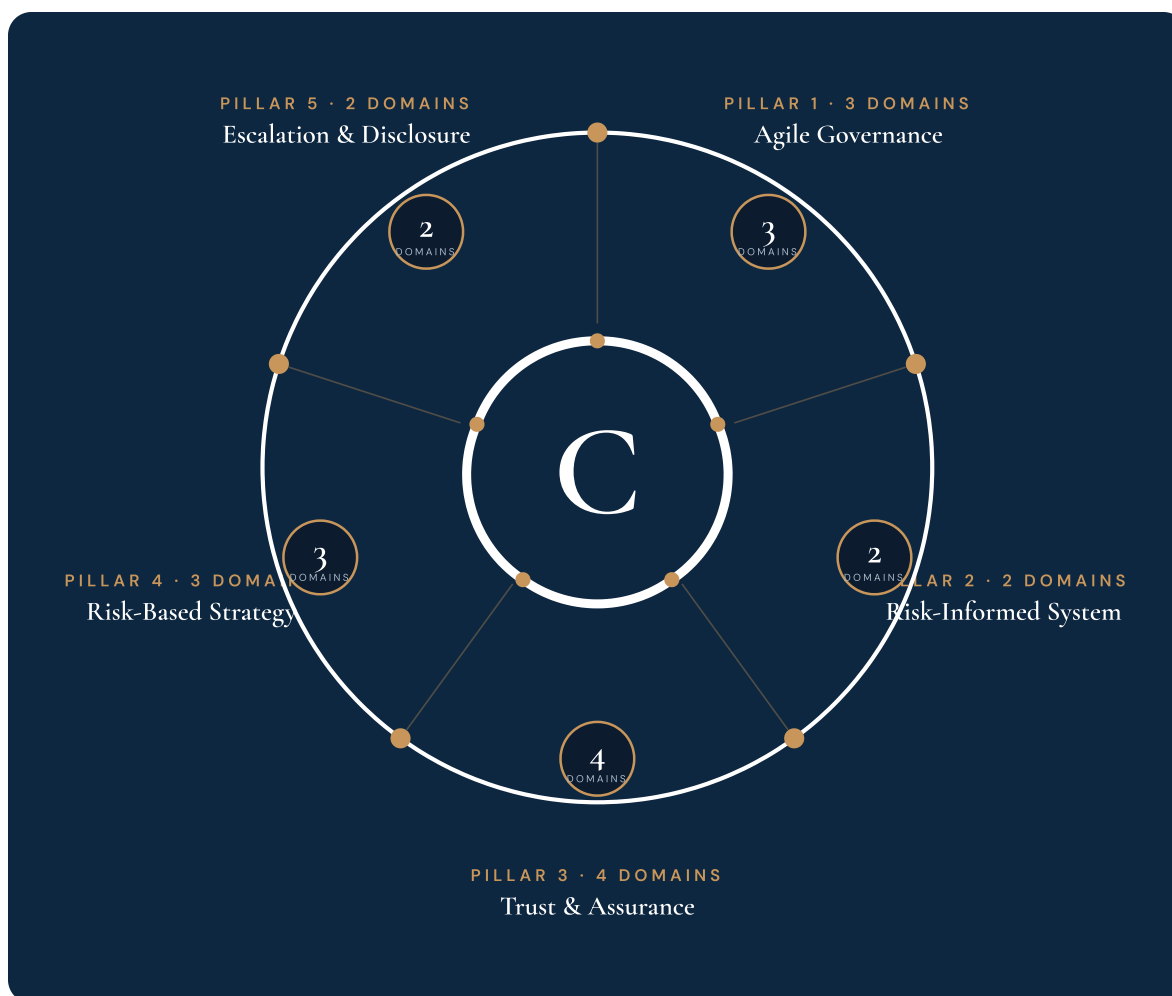


Figure B.1. The AI Oversight Program: five Pillars and fourteen Domains, with the additional layer of seventy-six controls maintained by the Center for AI Oversight.

Intellectual Property and Permitted Use

The AI Oversight Program is proprietary intellectual property of the Center for AI Oversight. This appendix is included in the Body of Knowledge with the Center's permission to support director education and PDA membership. Permitted uses: reading, internal board discussion, and citation in academic or professional writing with attribution. Prohibited without prior written permission: incorporation into commercial consulting deliverables, redistribution as a standalone framework, creation of derivative governance frameworks, and use in paid training materials. Commercial use and licensing inquiries should be directed to info@cfaio.org.

The Five Pillars

The Program is structured across five Pillars. Each Pillar groups one or more governance Domains around a common purpose.

1	Agile Governance	The Constitution	Establish the foundational management system providing strategic direction, authority, and adaptive policies.
2	Risk-Informed System	The Guardrails	Build the organization's sensory and analytical capabilities for AI risk.
3	AI Trust and Assurance	The Evidence	Define what AI systems must demonstrate. Trust is earned through evidence, not asserted.
4	Risk-Based Strategy and Execution	The Strategic Alignment	Link governance to business objectives so AI oversight enables better investment decisions.
5	Risk Escalation and Disclosure	The Voice	Ensure Decision Integrity through disciplined information flow to leadership.

The Fourteen Domains

Each Pillar is composed of one or more Domains. Each Domain has a single core function that distinguishes it from the others.

1 Agile Governance

PILLAR 1

- D1 AI Governance Program and Policy Framework.** Creates the legal and operational mandate for the Program itself.

- D2 AI Governance Structure, Oversight, and Resources.** Assigns accountability and resources the Program with necessary authority.

- D3 Governance Program Assurance and Continuous Learning.** Validates that the Program is operating as intended and evolving with the risk landscape.

2 Risk-Informed System

PILLAR 2

- D4 AI Risk Methodology, Scope, and Tolerance.** Defines the yardstick for risk and the boundaries of the Program's reach.

- D5 Risk Intelligence and Threat Landscape.** Provides forward-looking intelligence that informs leadership decisions.

- D6 AI Model Risk and Agentic Lifecycle Oversight.** Defines the governance standards for the lifecycle and autonomy of AI models and agents.
- D7 AI Data Governance Oversight.** Defines the governance standards for the integrity, legality, and quality of data within AI systems.
- D8 AI Transparency, Explainability, and Human Oversight.** Defines the standards under which AI decisions must be understandable, controlled, and demonstrably fair.
- D9 AI Security and Resilience Assurance.** Defines the standards under which AI systems are secured, and requires that assurance evidence is produced and reported.

- D10 Risk-Informed Strategy, Resources, and Organizational Readiness.** Aligns strategic ambition with risk reality, resources, and the organizational conditions necessary for governance to function in practice.
- D11 AI Value Realization and Operational Resilience Oversight.** Measures business success and ensures operational continuity.
- D12 Third-Party AI and Supply Chain Governance.** Defines the governance standards for risks inherited from the external AI supply chain.

D13 AI Risk Escalation and Disclosure Protocols. Ensures critical risks bypass silos and reach decision-makers immediately.

D14 Validation of Escalation and Governance Effectiveness. Tests the emergency brakes and audits the integrity of the decision-making process.

Concordance: BoK Sections to Program Domains

The following table maps each section of this Body of Knowledge to its primary alignment within the AI Oversight Program. Some sections touch multiple Domains; the table identifies the primary alignment. Domains are abbreviated as D1 through D14 to fit the table; the full Domain names appear in the list above.

1.1 Strategic Imperative	4. Risk-Based Strategy	D10
1.2 The Velocity Gap	(spans all five)	Framing
1.3 Builders and Buyers	4. Risk-Based Strategy	D12
1.4 Caremark and Duty of Oversight	1. Agile Governance	D2
2. How to Use This BoK	1. Agile Governance	D1, D3
3.1 Structure Beyond the Committee Fallacy	1. Agile Governance	D2
3.2 Governance Charter	1. Agile Governance	D1
3.3 Policy Framework	1. Agile Governance	D1
3.4 Auditing the Program	1. Agile Governance	D3
4.1 Risk Appetite and Tolerance	2. Risk-Informed System	D4
4.2 Crown Jewels Assessment	2. Risk-Informed System	D4
4.3 Risk Intelligence	2. Risk-Informed System	D5
4.4 International and Regulatory Landscape	2. Risk-Informed System	D5
5.1 Data Governance	3. AI Trust and Assurance	D7
5.2 The AI Registry	3. AI Trust and Assurance	D6, D7

5.3 Human Oversight, Fairness, and Labor Impact	3. AI Trust and Assurance	D8, D10
5.4 The Agentic Frontier	3. AI Trust and Assurance	D6
5.5 AI Security	3. AI Trust and Assurance	D9
6.1 Vendor Concentration	4. Risk-Based Strategy	D12
6.2 Unit Economics of AI	4. Risk-Based Strategy	D11
6.3 Operational Resilience	4. Risk-Based Strategy	D11
7.1 Defined Escalation Protocols	5. Escalation and Disclosure	D13
7.2 Materiality and Disclosure	5. Escalation and Disclosure	D13
7.3 Crisis Readiness	5. Escalation and Disclosure	D14
Appendix C: Third-Party Contract Considerations	4. Risk-Based Strategy	D12

The AI Oversight Program is maintained by the Center for AI Oversight and is updated regularly to reflect changes in regulation, technical standards, and case law. The current version, supporting materials, board-level insights, and additional resources are available at cfaio.org.

Appendix C. Third-Party Contract Considerations

A structured set of provisions for AI vendor agreements

This appendix collects the contract-stage provisions a Buyer company should secure in its agreements with AI vendors. It is the contract-stage companion to Tool D, the vendor due diligence questionnaire. Tool D sits upstream of the contract and helps the program decide whether to enter into the relationship. This appendix sits at the contract itself, and at every renewal.

The audience is the board and the program owner, not outside counsel. The appendix tells the board what to insist on. The drafting belongs to counsel, working from the company's standard forms and the specific deal in front of them.

Not every provision will be available in every deal. Vendor leverage varies, and smaller buyers often face take-it-or-leave-it terms. Where a provision cannot be secured, the gap is a finding. The program should know which gaps exist, which carry residual risk, and whether the risk has been accepted at the appropriate level under the policy framework in Section 3.3.

I. Data Ownership, Use, and Portability

Data is where most AI vendor disputes start. Ambiguous terms on ownership, training use, and portability create exposure that is hard to recover once the relationship is underway. These provisions should be explicit in every AI vendor agreement, regardless of vendor size.

- ◆ Clear statement that the company retains ownership of all data submitted to the vendor's AI, all outputs generated for the company's use, and any derivatives of either.
- ◆ Prohibition on the vendor using the company's data to train, fine-tune, or otherwise improve models that serve other customers, unless the company has affirmatively opted in on terms the program has approved.

- ◆ Data portability rights enabling the company to extract its data and any associated metadata in a usable format, at any time and on termination.
 - ◆ Defined deletion obligations on termination, including timeline, scope (production, backups, derivatives), and a verification mechanism the company can rely on.
 - ◆ Restrictions on the vendor's right to retain anonymized, aggregated, or de-identified data derived from the company's inputs, if any such retention is permitted at all.
-

2. Model Updates, Behavior, and Change Notice

AI vendors update models continuously. A model that performs one way at contract signing may perform differently three months later, sometimes materially so. Standard technology contracts do not contemplate this. AI contracts must.

- ◆ Advance notice of material model updates, retraining events, or behavioral changes that could affect the company's operations, with defined notice periods proportionate to materiality.
 - ◆ The company's right to evaluate material updates before they are deployed against its workload, or at minimum to roll back to a prior version for a defined period if performance regresses.
 - ◆ Performance commitments specific to AI failure modes (accuracy drift, output quality, hallucination rates where applicable), not just general availability.
 - ◆ For autonomous capabilities, contractual limits on what the vendor's AI can do without human review in the company's environment, mapped to the agency limits the program has defined under Section 5.4.
 - ◆ Explainability or audit-trail provisions sufficient to reconstruct consequential decisions after the fact.
-

3. AI-Specific Security

AI introduces attack surfaces traditional cybersecurity terms were not written to cover. Prompt injection, model manipulation, and data extraction through crafted queries are

categories of risk that need their own contractual treatment, on top of standard security provisions.

- ◆ Representations that the vendor has tested its AI for prompt injection, model manipulation, and data extraction risks, with the cadence of testing specified.
 - ◆ Logical and operational segregation of the company's data from the data of other customers, with controls described in sufficient detail to support audit.
 - ◆ Incident notification obligations specific to AI security events (data exposure through manipulated queries, model behavior compromised by adversarial inputs), distinct from general cybersecurity incident terms and on a defined timeline.
 - ◆ The vendor's commitment to notify the company of security issues discovered in upstream model providers or sub-processors that could affect the company's data.
-

4. Liability Allocation and Indemnification

Standard liability caps and indemnification language often fail to allocate the specific risks AI creates: outputs that cause harm, autonomous actions that produce unintended consequences, intellectual property exposure from training data, and regulatory liability from biased decisions. Each of these needs to be addressed directly.

- ◆ Indemnification for third-party intellectual property claims arising from outputs produced by the vendor's AI, including claims based on training data the vendor used.
- ◆ Indemnification for actions taken by autonomous agents operating in the company's environment under the vendor's product, including financial transactions, communications, and modifications to systems of record.
- ◆ Liability allocation that does not place the full burden of AI-driven errors on the company simply because the company deployed the vendor's tool.
- ◆ Carve-outs from liability caps for AI-specific failures that produce regulatory exposure, data loss, or material business interruption, calibrated to what the company's other agreements (D&O, cyber, technology E&O) actually cover.
- ◆ Where the vendor's AI processes data about individuals, indemnification or specific allocation for claims arising from discriminatory or non-compliant outputs.

5. Change of Control, Exit, and Continuity

AI vendor relationships often outlast the personnel and ownership structures that signed them. Provisions that survive transactions on either side, and that preserve the company's ability to exit on workable terms, are essential to operational resilience and to M&A readiness.

- ◆ Identification of any change-of-control provisions (the vendor's or the company's) that could trigger termination, repricing, renegotiation, or data-handling obligations, and the company's position on each.
- ◆ Transition assistance obligations on termination, including the vendor's commitment to support migration to a successor provider on defined terms and at defined cost.
- ◆ Realistic exit timelines reflecting the actual complexity of migrating AI workloads, not the standard technology-contract default.
- ◆ Continuity provisions covering vendor financial distress, acquisition, product discontinuation, and material changes to the vendor's AI roadmap.
- ◆ For mission-critical dependencies, escrow or equivalent arrangements covering the AI capability itself or the data necessary to reconstitute it elsewhere, to the extent commercially available.

6. Sub-Processors and Supply Chain Transparency

Most AI vendors depend on upstream model providers, cloud infrastructure, and other third parties. The company's exposure flows through that supply chain even when its direct contract is only with the vendor. Section 4.4 discusses this from an international and regulatory standpoint; the contract terms below give the program visibility into it.

- ◆ Disclosure of all sub-processors, including upstream model providers, that have access to the company's data or that materially affect the vendor's AI.
- ◆ The vendor's commitment to flow down relevant contractual protections (security, confidentiality, data handling) to its sub-processors.
- ◆ Advance notice and the company's right to object to material changes in sub-processors that would affect data handling, jurisdiction of processing, or upstream

model provenance.

- ◆ The vendor's representation that it has the legal right to provide the AI capability in the jurisdictions where the company operates.
-

7. Audit Rights and Assurance Reports

The program cannot verify what it cannot see. Audit rights and access to assurance reports are how the company satisfies its oversight obligations without operating the vendor's controls itself.

- ◆ Annual access to SOC 2, ISO 27001, or equivalent assurance reports, with scope sufficient to cover the AI capability and not just the vendor's general infrastructure.
 - ◆ Where applicable, access to assurance reports specific to the vendor's AI governance practices, including bias testing, model risk management, and incident history.
 - ◆ Direct audit rights for the company, exercisable on reasonable notice and at reasonable frequency, particularly for mission-critical dependencies.
 - ◆ The vendor's commitment to cooperate with the company's regulators, auditors, and acquirers in connection with diligence on the AI relationship.
-

8. Regulatory and Compliance Alignment

The international and regulatory landscape addressed in Section 4.4 creates obligations that often reach contractual relationships even when the company's direct exposure is domestic. AI-specific regulation is evolving on a faster timeline than most contract renewal cycles, and contracts should anticipate that movement.

- ◆ The vendor's representation that its AI capability complies with applicable law in the jurisdictions where the company uses it, with an obligation to maintain that compliance as the law evolves.
- ◆ Notification obligations when the vendor becomes aware of regulatory inquiries, enforcement actions, or material developments that could affect the company's use of the AI.

- ◆ Cooperation in the company's regulatory disclosures, where AI-related incidents or capabilities create reporting obligations under LP, credit, insurance, or regulatory regimes (Section 7.2).
 - ◆ Where the company operates internationally, alignment with the vendor's compliance obligations in those jurisdictions, including data residency and cross-border transfer terms.
-

How to Use This Appendix

This appendix is a checklist of considerations, not a contract template. The program owner should use it in three contexts: at initial vendor selection (in combination with Tool D), at every contract renewal, and during M&A diligence on either side of a transaction.

Where a provision cannot be obtained, the program should record the gap, the reason, and whether the residual risk has been accepted under the policy framework. Boards conducting Section 5 reviews should expect to see those gaps surfaced, not buried.

Counsel will adapt these considerations to the company's standard forms, the specific vendor, and the applicable jurisdiction. The board's role is to ensure the program has secured what it could, has documented what it could not, and has accepted residual risk at the appropriate level.

Back Matter

About the Contributors, Notice and Disclaimer, and Acknowledgements

This section closes the Body of Knowledge with three back-matter pieces: a brief description of the two contributing organizations, the notice and disclaimer covering the limits of this document, and the acknowledgements.

About the Contributors

The two organizations behind this Body of Knowledge

The Private Directors Association

The Private Directors Association (PDA) is the nation's leading professional association of directors who serve on the boards of privately held companies, family businesses, and not-for-profit organizations. The PDA advances the practice of governance in the private company sector through chapters, programs, and educational content tailored to directors who hold fiduciary responsibility outside the public company framework. For more information, visit privatedirectorsassociation.org.

The Center for AI Oversight

The Center for AI Oversight develops governance methodology, oversight programs, and director education focused on AI risk and opportunity at the board level. The Center works with private and public company boards, investors, and regulators to translate the AI landscape into oversight practices that are legally defensible, operationally practical, and proportionate to the organization's actual exposure. The Center contributed the governance methodology that underlies this Body of Knowledge and developed the illustrative AI Oversight Program in Appendix B. For more information, visit cfaio.org.

Lead Author

Brian Allen serves as Executive Director of the Center for AI Oversight. He is the architect of the AI Oversight Program presented in Appendix B and served as lead author of this Body of Knowledge.

Notice and Disclaimer

The limits of this document

This Body of Knowledge is published by the Private Directors Association in collaboration with the Center for AI Oversight. It is intended as a strategic guide for private company directors. It is not, and should not be construed as, legal, accounting, tax, investment, or professional advice.

Directors, officers, and other readers who face specific questions about their fiduciary duties, regulatory obligations, or the application of AI to their organization's particular circumstances should consult qualified counsel and other appropriate professional advisors. Nothing in this document creates an attorney–client relationship between the reader and any contributor.

References to case law, statutes, regulations, technical standards, and industry frameworks are accurate to the best of the authors' knowledge as of the publication date. Law, regulation, and the AI landscape continue to evolve. Readers should verify current authority before relying on any specific reference for action.

The illustrative AI Oversight Program presented in Appendix B is proprietary intellectual property of the Center for AI Oversight and is reproduced in this Body of Knowledge with the Center's permission. The terms of use are specified in Appendix B.

Errors and omissions are the responsibility of the authors. The Private Directors Association and the Center for AI Oversight disclaim any liability for actions taken or not taken in reliance on the content of this document. Comments, corrections, and suggestions for future editions are welcome and should be directed to the PDA.

Acknowledgements

The people who made this Body of Knowledge possible

The Private Directors Association is the publisher of this Body of Knowledge and the institutional home for the work. Institutional support, editorial direction, and the audience perspective that kept the document grounded in what private company directors actually need came from:

Suzanne Siradas

Director, Education and National Events

The Center for AI Oversight contributed the governance methodology that underlies the framework, the case law analysis that informs Section 1.4 and the legal terms in the Glossary, and the illustrative AI Oversight Program presented in Appendix B. Lead author:

Brian Allen

Executive Director

The PDA Working Group reviewed successive drafts and provided the feedback that shaped the document at every level, from individual word choices to substantive content additions on accountability, agentic AI, and the international landscape. The authors are grateful to:

Bich-Thuy Le

Carla Schlemminger

Fayeron Morrison

Fred Franks

Gary Cao

Jerry Alsay

Joanna Ridgway

Paula Fontana

Philip Topham

Shep Pryor

Sylvia Carrizosa

Vanessa Maren

Comments and suggestions for future editions are welcome and should be directed to the PDA.