W O R K W I S E  S O L U T I O N S

W H I T E  P A P E R

# Agentic AI Governance in Private Equity

*A Behavioral Framework for Autonomous Decision Systems*

**Dr. Leigh Coney**
Founder & Principal Consultant, WorkWise Solutions
Q1 2026

workwisesolutions.org
Contact@workwisesolutions.org

## Abstract

As private equity firms deploy autonomous AI agents across deal sourcing, due diligence, and portfolio monitoring, governance frameworks have failed to keep pace. Current approaches treat agentic AI governance as a technical compliance exercise—focused on data access controls, audit trails, and regulatory checklists. This paper argues that the most consequential governance failures in private equity will not be technical but behavioral: investment professionals who rubber-stamp AI recommendations, escalation pathways that go unused under deal pressure, and trust calibration errors that compound across the investment lifecycle.

Drawing on organizational psychology, decision science, and established research on human-automation interaction in high-stakes environments, this paper introduces the Behavioral Governance Framework (BGF)—a model that integrates human cognitive and social dynamics into the design of agentic AI oversight systems. The BGF addresses three critical gaps in existing governance models: the escalation design problem (why professionals fail to override autonomous systems even when they detect errors), the trust calibration problem (how fiduciary responsibility distorts rational AI reliance), and the organizational incentive problem (how firm-level pressures systematically degrade oversight quality over time).

The framework proposes a set of design principles, organizational structures, and behavioral interventions tailored to the private equity operating environment. It is intended for general partners, chief technology officers, chief compliance officers, and operating partners responsible for deploying AI across the investment lifecycle.



## Keywords

Agentic AI governance, autonomous AI agents, private equity, human-in-the-loop oversight, AI trust calibration, behavioral science, decision-making frameworks, AI adoption resistance, fiduciary duty, AI risk management, organizational psychology, investment committee AI, deal screening AI, portfolio monitoring AI, EU AI Act compliance, data sovereignty

# Table of Contents

# 1. Introduction: The Governance Gap in Private Equity AI

The private equity industry is in the early stages of a structural transformation driven by artificial intelligence. What began as narrow automation of document extraction and data normalization has evolved rapidly into the deployment of autonomous AI agents capable of multi-step reasoning, independent action, and consequential decision-making across the investment lifecycle. These agentic systems do not merely summarize information or generate recommendations on demand; they independently source deals, evaluate risks, draft investment committee memoranda, monitor portfolio company performance, and generate investor reports—often with minimal human intervention between initiation and output.

The pace of adoption is striking. Industry surveys indicate that more than 80% of PE and VC firms were actively using AI tools by late 2024, up from less than half the prior year. While the majority of current deployments involve assistive tools rather than fully autonomous agents, the trajectory toward agentic deployment is accelerating rapidly. Some of the world's largest sponsors now estimate that 30–40% of investment committee discussions focus on whether portfolio companies can deploy AI effectively or could be disrupted by AI-enabled competitors. A small but growing number of firms have introduced AI systems into their investment committee workflows—reviewing deal materials, surfacing sector-specific risks, and feeding structured recommendations into the deliberation process.

Yet the governance infrastructure surrounding these systems remains dangerously underdeveloped. The dominant approach to AI governance in financial services borrows from enterprise IT compliance: access controls, data retention policies, audit logs, model validation, and regulatory checklists. These technical safeguards are necessary, but they are designed for a world in which humans make decisions and technology supports them. Agentic AI inverts this relationship. When an autonomous system generates a deal screening score, drafts a due diligence summary, or flags a portfolio company for intervention, the human role shifts from decision-maker to reviewer. And the behavioral dynamics of review are fundamentally different from the dynamics of creation.

This paper addresses a specific gap in the governance conversation: the behavioral and organizational dimensions of agentic AI oversight in private equity. It introduces the Behavioral Governance Framework (BGF), which integrates insights from organizational psychology, decision science, and research on human-automation interaction in high-stakes environments. The BGF is designed to complement—not replace—technical governance measures. Its premise is straightforward: if you

design a governance system that assumes humans will behave rationally, consistently, and attentively when overseeing autonomous AI, that system will fail.

# 2. The Agentic AI Landscape in Private Capital

## 2.1 From Tools to Agents: A Qualitative Shift

The distinction between AI tools and AI agents is not merely semantic; it reflects a qualitative change in the relationship between technology and the professionals who use it. Traditional AI tools in private equity—document extraction engines, financial modeling accelerators, natural language search across data rooms—operate within a request-response paradigm. A human asks a question or initiates a task, and the tool produces an output. The human retains full control over the decision-making loop, and the AI's contribution is bounded by the specificity of the prompt.

Agentic AI systems break this paradigm in three important ways. First, they pursue multi-step goals autonomously. A deal screening agent, for example, does not merely extract data from a confidential information memorandum; it identifies the document type, selects an appropriate analytical framework, runs comparative analyses against historical deals, flags risk factors, synthesizes the findings into a structured memorandum, and assigns a preliminary score—all without intermediate human instructions. Second, agentic systems make decisions about their own workflow. They choose which data sources to consult, which analytical methods to apply, and how to handle ambiguity. Third, they take actions that have real-world consequences: updating databases, generating reports that flow downstream to investors, and triggering alerts that may initiate material business decisions.

## 2.2 Current Deployment Patterns

Agentic AI deployment in private equity currently clusters around four primary domains, each with distinct governance implications:

- **Deal Intelligence:** Autonomous deal sourcing engines that scan proprietary and public datasets, identify high-fit opportunities, and generate initial screening memoranda. These systems increasingly operate on a continuous basis, surfacing potential targets without being prompted by deal team members.
- **Due Diligence Acceleration:** Agents that conduct data room analysis, competitor mapping, EBITDA adjustments, and ESG risk flagging. Some systems compress what was previously a multi-week manual review into a matter of hours, generating investment committee-ready dossiers with minimal human editing.

- **Portfolio Monitoring:** Cross-asset surveillance systems that aggregate financial, operational, and market data across portfolio companies, detect performance deterioration, and generate early warning alerts. These agents operate continuously and may flag issues weeks before they surface in standard quarterly reporting.
- **Stakeholder Reporting:** Automated generation of LP reports, board packs, and regulatory filings from raw portfolio data. These agents transform unstructured operational information into institutional-grade documents with minimal human intervention.

Each of these domains presents a different governance challenge. Deal intelligence agents create risks of over-reliance on AI-generated deal flow at the expense of relationship-driven sourcing. Due diligence agents create risks of automation bias in risk assessment. Portfolio monitoring agents create risks of alert fatigue and desensitization. Reporting agents create risks of accuracy degradation when humans stop verifying AI-generated outputs. The behavioral dynamics differ in each case, and governance systems must be designed accordingly.

# 3. Why Technical Governance Is Necessary but Insufficient

## 3.1 The Current State of AI Governance Frameworks

The governance landscape for agentic AI is evolving rapidly. Singapore's Infocomm Media Development Authority published its Model AI Governance Framework for Agentic AI in early 2026, providing one of the first government-published operational blueprints for structuring agent oversight. The EU AI Act, which entered into force in August 2024 with phased implementation extending through 2027, classifies AI systems used in high-risk domains—including employment, creditworthiness, and critical infrastructure—as high-risk, subjecting them to requirements for risk management, human oversight mechanisms, transparency, and robustness controls. In practice, most agentic AI systems deployed in the PE investment lifecycle are likely to fall within these high-risk categories given their role in consequential financial decision-making. Industry organizations including OWASP, NIST, and ISO have published or are developing standards that address autonomous system risks.

These frameworks share a common architecture: they emphasize access controls, data governance, model validation, audit trails, transparency requirements, and human oversight checkpoints. They are necessary and valuable. However, they share a critical blind spot: they assume that if you create the structural opportunity for human oversight, humans will exercise it effectively. This assumption does not hold under the conditions that characterize private equity decision-making.

## 3.2 The Oversight Assumption and Its Failure Modes

Decades of research in organizational psychology and human factors engineering demonstrate that human oversight degrades predictably under specific conditions—conditions that are pervasive in the PE operating environment:

> *The fundamental error in current governance design is the assumption that providing a checkpoint is equivalent to ensuring effective oversight. A mandatory sign-off screen is not governance; it is theater, unless the behavioral conditions for genuine review are present.*

- **Time pressure:** Deal execution operates under intense competitive pressure. When a deal team is racing to submit a bid, the incentive to carefully review an AI-generated due diligence summary is overwhelmed by the incentive to move fast. Governance checkpoints become speed bumps, not guardrails.
- **Automation complacency:** As AI systems demonstrate consistent accuracy, human reviewers develop learned trust and reduce the cognitive effort they invest in verification. This is not laziness; it is a well-documented cognitive adaptation. Oversight becomes perfunctory precisely when it should remain rigorous.
- **Authority gradients:** Junior team members are typically the ones closest to AI outputs but least empowered to challenge them. When an AI agent produces a deal screening score that aligns with a senior partner's thesis, the organizational cost of raising objections may exceed the perceived benefit.
- **Diffusion of responsibility:** When multiple people are involved in an oversight process, each individual's sense of personal accountability diminishes. If an AI-generated LP report passes through three review stages, each reviewer may assume the others are conducting the rigorous check.
- **Alert fatigue:** Portfolio monitoring agents that generate continuous alerts create a signal-to-noise problem. When—as is common in early deployments—the vast majority of alerts turn out to be false positives or immaterial, reviewers begin ignoring the remainder—which may include the genuine early warning signals the system was designed to detect.

These failure modes are not theoretical. They are well-documented in industries ranging from aviation to healthcare to nuclear power. The private equity industry's adoption of agentic AI is creating identical conditions, but the governance conversation has not yet caught up.

# 4. The Behavioral Governance Framework (BGF)

## 4.1 Framework Overview

The Behavioral Governance Framework rests on a core premise: effective governance of agentic AI in high-stakes environments requires designing for how humans actually behave, not how they should behave. This means integrating behavioral science insights into the structural design of oversight systems, escalation pathways, incentive structures, and organizational culture.

The BGF consists of three pillars, each addressing a distinct category of behavioral governance failure:

| Pillar | Core Problem | Design Principle |
|---|---|---|
| I. Escalation Architecture | Professionals fail to override AI outputs even when they detect errors, because escalation is costly, slow, or socially penalized. | Make escalation the path of least resistance. Design systems where flagging disagreement is faster and easier than silent approval. |
| II. Trust Calibration | Fiduciary responsibility and professional identity distort rational AI reliance, creating both over-trust and under-trust failure modes. | Create structured mechanisms for ongoing trust recalibration based on empirical performance data, not intuition or status. |
| III. Incentive Alignment | Organizational pressures (speed, throughput, competitive positioning) systematically degrade oversight quality over time. | Align individual and team incentives with governance quality, not just output velocity. Make oversight performance visible and valued. |

## 4.2 Theoretical Foundations

The BGF draws on four established research traditions. First, the literature on automation bias and complacency, which demonstrates that human operators of automated systems systematically over-rely on automated outputs and fail to detect automation errors at rates far exceeding their failure rates in manual tasks. Second, the organizational behavior literature on authority gradients and psychological safety, which explains why hierarchical organizations systematically suppress dissent from junior members—precisely the individuals most likely to be proximate to AI outputs. Third, decision science research on trust calibration, which shows that human trust in automated systems follows predictable trajectories that can be influenced by system design. Fourth, the technology acceptance literature, which identifies performance expectancy, social influence, and facilitating conditions as

key determinants of adoption behavior—dynamics that take on distinctive characteristics when the adopters are senior professionals with established analytical identities.

A methodological note is warranted. While the behavioral dynamics described in this framework are well-established in aviation, healthcare, process control, and other high-stakes domains, empirical research specifically examining automation bias and trust calibration in private equity decision-making environments remains limited. This paper applies established behavioral science principles to the PE context by analogy, informed by the structural similarities between PE decision-making and other high-stakes, time-pressured environments where these dynamics have been extensively documented. Closing this empirical gap through PE-specific field research is an important direction for future work.

Critically, the BGF treats governance not as a compliance function but as a design discipline. The question is not “what rules should we impose?” but “how should we design systems, processes, and organizational structures so that effective oversight emerges naturally from how people work?”

# 5. Pillar I: Escalation Architecture Design

## 5.1 The Escalation Problem in PE

In a typical agentic AI deployment, the escalation pathway is designed as an exception-handling mechanism: the AI operates autonomously by default, and humans are expected to intervene when something goes wrong. This design choice—borrowed from industrial automation—is poorly suited to the PE environment for two reasons.

First, detecting that something has “gone wrong” with an AI agent’s output requires domain expertise, attention, and time. An AI-generated deal screening memo that contains a subtly flawed EBITDA adjustment is not obviously wrong; it requires an analyst with relevant sector experience to recognize the error. If that analyst is processing ten AI-generated memos per day instead of the two they would have reviewed manually, the probability of catching any individual error drops substantially.

Second, escalation in PE firms carries social and professional costs. Flagging an error in an AI output may implicitly challenge the technology investment decisions of senior partners. It may slow down a deal process that the firm has committed to pursuing. It may require the escalating individual to articulate a nuanced analytical

judgment to colleagues who lack the context to evaluate it. The path of least resistance is almost always to approve the output and move on.

## 5.2 Design Principles for Effective Escalation

The BGF proposes five design principles for escalation architecture:

### Principle 1: Asymmetric Friction

Make it harder to approve an AI output without review than to flag it for further examination. This inverts the default in most AI governance systems, where approval is frictionless (a single click) and escalation requires effort (filling out a form, writing a justification, notifying a supervisor). Practically, this means designing review interfaces where the default action is "flag for discussion" and where approval requires the reviewer to affirmatively attest to specific verification steps.

### Principle 2: Anonymous Escalation Channels

Provide mechanisms for junior team members to flag concerns about AI outputs without attribution. This reduces the authority gradient effect and removes the social cost of dissent. An anonymous escalation channel is not a substitute for a culture of psychological safety, but it provides a backstop when cultural norms are insufficient. In smaller deal teams where true anonymity is impractical—as is common in PE, where teams of three to six professionals work on a given deal—the same objective can be achieved through structured dissent protocols or mandatory devil's advocate assignments that normalize disagreement as a role rather than a personal act.

### Principle 3: Structured Disagreement Protocols

Require AI systems to present their reasoning in a format that facilitates structured disagreement. Rather than presenting a single recommendation with a confidence score, agents should present the two or three most plausible interpretations of the data and explain why they selected one over the others. This transforms the reviewer's task from "find the error" (which is cognitively demanding) to "evaluate the reasoning" (which leverages existing analytical skills).

### Principle 4: Rotational Review Assignment

Rotate the assignment of AI output review across team members to prevent familiarity-based complacency. When the same individual reviews the same agent's outputs day after day, they develop mental models of the agent's behavior that cause them to process outputs with decreasing attention. Rotation disrupts this pattern and maintains the cognitive engagement required for effective oversight. Importantly, rotation should be implemented within sectors or domain clusters

where reviewers share baseline expertise, rather than across unrelated verticals. The goal is to prevent the same individual from reviewing the same agent's outputs daily—not to remove domain expertise from the review process.

### Principle 5: Calibrated Autonomy Levels

Not all AI agent actions require the same level of oversight. A deal sourcing agent that surfaces potential targets for human evaluation requires less oversight than a due diligence agent that generates risk assessments used in investment committee deliberations. The BGF proposes a four-tier autonomy classification:

| Tier | Autonomy Level | Example | Oversight Requirement |
|---|---|---|---|
| Tier 1: Informational | Full autonomy | Market news aggregation, public data compilation | Periodic audit (monthly) |
| Tier 2: Analytical | High autonomy | Preliminary deal screening, competitor mapping | Sampling review (10-20% of outputs) |
| Tier 3: Evaluative | Moderate autonomy | Risk scoring, EBITDA adjustments, ESG flagging | Mandatory review before downstream use |
| Tier 4: Decisional | Supervised autonomy | IC memo drafting, LP report generation, portfolio alerts | Dual review with documented sign-off |

The tier classification should be determined collaboratively by the deal team, compliance function, and technology team—not by the technology team alone. This ensures that the people closest to the consequences of an error have input into the level of oversight applied.

## 6. Pillar II: Trust Calibration Systems

### 6.1 The Trust Calibration Problem

Trust in AI systems is not static; it evolves along predictable trajectories. Research in human-automation interaction identifies three common trust trajectories, all of which are observable in PE firms deploying agentic AI:

- **Overtrust trajectory:** Initial skepticism gives way to high trust as the system demonstrates competence. Over time, trust exceeds the system's actual reliability, leading to complacency and failure to detect errors. This is the most common trajectory in PE firms where AI adoption is championed by senior leadership.

- **Undertrust trajectory:** A single high-profile failure—an AI-generated memo that contains an embarrassing error, a missed risk factor that surfaces in post-mortem—causes trust to collapse disproportionately. The system is abandoned or relegated to low-value tasks, regardless of its aggregate accuracy. This is common when AI adoption is driven by technology teams without adequate stakeholder management.
- **Bifurcated trust trajectory:** Different individuals or teams develop divergent trust levels based on their personal experiences with the system, creating organizational inconsistency. Senior partners who have seen the system perform well in their sector trust it highly; partners with less exposure remain skeptical. This creates uneven governance quality across the firm.

## 6.2 The Fiduciary Complication

Trust calibration in PE is further complicated by fiduciary responsibility. Investment professionals operate under fiduciary obligations—defined by their limited partnership agreements, regulatory requirements, and professional norms—to act in the best interests of their limited partners. When an AI agent generates a risk assessment that the investment professional relies upon in making an allocation decision, questions of accountability become acute. Who bears responsibility for an error in the AI's analysis: the technology vendor, the firm's CTO, the deal team lead, or the individual who signed the investment committee memo?

This ambiguity creates perverse incentive effects. Some professionals over-rely on AI outputs precisely because doing so diffuses personal accountability: if the AI was wrong, they can point to the system's track record and their good-faith reliance on it. Others under-rely on AI because they view personal analytical judgment as a core component of their fiduciary duty, regardless of the AI's empirical accuracy. Both responses are rational given the ambiguity, and both are governance failures.

## 6.3 Design Principles for Trust Calibration

### Principle 6: Empirical Performance Dashboards

Provide every individual who interacts with AI agent outputs with a continuously updated, domain-specific performance dashboard. This dashboard should show the agent's accuracy rate for the specific types of outputs that individual reviews, broken down by sector, deal size, and complexity. The purpose is to replace intuitive trust calibration—which is subject to availability bias, recency bias, and anchoring—with data-driven calibration.

### Principle 7: Structured Accountability Mapping

Create explicit, written accountability maps for every AI agent deployed in the investment process. These maps should specify who is responsible for verifying each component of the agent's output, what specific checks they are expected to perform, and what the consequences are for both false acceptance (approving an erroneous output) and false rejection (unnecessarily escalating an accurate output). The goal is to eliminate the diffusion of responsibility that occurs when accountability is vaguely distributed.

#### Principle 8: Deliberate Failure Exposure

Periodically and without warning, introduce controlled errors into AI agent outputs in sandboxed review environments to test whether human reviewers detect them. This approach—borrowed from quality assurance practices in aviation and medical diagnostics—serves two purposes: it provides empirical data on the quality of human oversight, and it maintains reviewer vigilance by making oversight a task with genuine consequences. Critically, deliberate failure exposure must be conducted in environments where injected errors are intercepted before any downstream action regardless of whether the reviewer catches them; introducing controlled errors into live deal processes or investor communications would create unacceptable liability risk. The detection rate should be tracked and reported to governance committees, not used punitively against individual reviewers. The purpose is system improvement, not blame.

#### Principle 9: Confidence Interval Communication

Require AI agents to communicate uncertainty explicitly, not as a single confidence score but as a structured disclosure of what the agent knows, what it is uncertain about, and what it does not know. A deal screening agent should distinguish between high-confidence quantitative assessments (the target's revenue grew at 12% CAGR over five years) and low-confidence qualitative judgments (the management team appears capable but has not been tested in a downturn). This transforms AI output from an apparent statement of fact into a structured analytical input that invites professional judgment.

## 7. Pillar III: Organizational Incentive Alignment

### 7.1 The Incentive Misalignment Problem

The most carefully designed escalation architecture and trust calibration system will degrade over time if the organizational incentive structure rewards speed and throughput over governance quality. This is not a hypothetical concern; it is the central challenge of agentic AI governance in PE.

Private equity firms operate under intense competitive pressure. Deals move quickly, and the ability to evaluate opportunities faster than competitors is a genuine source of competitive advantage. When AI agents enable a deal team to screen three times as many opportunities per quarter, the organizational expectation shifts accordingly. The team is now expected to screen three times as many deals. The time saved by AI is not reallocated to deeper analysis; it is consumed by increased throughput. In this environment, governance becomes the bottleneck—the one thing standing between the firm and its new throughput target—and organizational pressure to streamline or circumvent governance intensifies.

## 7.2 Design Principles for Incentive Alignment

### Principle 10: Governance Quality Metrics

Develop and track metrics that measure the quality of human oversight, not just the speed of review. These metrics should include the rate at which reviewers identify AI errors (detection rate), the rate at which reviewers provide substantive modifications to AI outputs (engagement rate), and the time invested in review relative to the complexity of the output (depth-of-review ratio). These metrics should be reported alongside productivity metrics in performance reviews and compensation discussions.

### Principle 11: Governance Success Stories

Actively document and celebrate instances where human oversight prevented a governance failure. When a reviewer catches an error in an AI-generated EBITDA adjustment that would have materially affected an investment decision, that should be treated as a win of equal significance to a successful deal closure. Creating organizational narratives around governance success reframes oversight from a cost center to a value-creation activity.

### Principle 12: Senior Partner Accountability

Assign named senior partners as governance sponsors for each major AI deployment. These individuals should be personally accountable for the quality of governance in their domain and should report on governance metrics at the same cadence and with the same rigor as investment performance metrics. This prevents governance from being delegated entirely to compliance or technology functions, where it lacks organizational authority.

### Principle 13: Capacity Budgeting for Oversight

When deploying an AI agent that increases deal team throughput, explicitly budget capacity for the incremental oversight work that the increased throughput creates. If an AI agent enables a team to screen 200 deals per quarter instead of 60, the

governance question is not just "how do we review the AI's outputs?" but "how many additional review hours does this throughput require, and who is allocated to perform them?" Throughput increases without corresponding oversight capacity increases are governance failures in waiting.

### Principle 14: Periodic Governance Stress Tests

Conduct quarterly governance stress tests in which an external or independent party reviews a sample of AI-generated outputs that passed through the governance process and evaluates whether the human oversight was substantive or perfunctory. The results of these stress tests should be reported to the firm's management committee or advisory board. This creates organizational accountability for governance quality that exists independent of individual reviewer performance.

## 8. Implementation: A Phased Deployment Model

Implementing the BGF is not a single event but a staged process that should be integrated with the firm's broader AI deployment timeline. The following phased model provides a practical roadmap:

### Phase 1: Assessment and Baseline (Weeks 1–4)

- **Agent inventory:** Catalog all AI agents currently deployed or planned for deployment. For each agent, document its function, autonomy level, data access, output consumers, and downstream consequences of errors.
- **Behavioral baseline:** Conduct structured interviews and workflow observations to assess current oversight behaviors. How much time do reviewers spend on AI output review? How frequently do they modify or override AI outputs? What percentage of escalations reach resolution?
- **Incentive audit:** Map the current incentive structure to identify misalignments. Are deal teams rewarded for throughput? Is governance quality measured? Are there consequences for governance failures that do not result in financial losses?

### Phase 2: Architecture Design (Weeks 5–10)

- **Tier classification:** Assign each AI agent to an autonomy tier using the four-tier model described in Section 5. This classification should be a collaborative exercise involving deal teams, compliance, and technology.
- **Escalation pathway design:** Design escalation pathways for each tier, incorporating asymmetric friction, anonymous channels, and structured disagreement protocols.

- **Trust calibration infrastructure:** Build or procure performance dashboards, accountability maps, and confidence interval reporting systems.
- **Incentive restructuring:** Propose modifications to performance evaluation criteria to incorporate governance quality metrics.

### Phase 3: Pilot Deployment (Weeks 11–18)

Deploy the BGF for one or two AI agents in a single deal team or portfolio management function. Monitor escalation rates, detection rates, trust calibration accuracy, and reviewer engagement. Conduct weekly debriefs with pilot participants to identify friction points and refine the design.

### Phase 4: Firmwide Rollout (Weeks 19–30)

Extend the BGF across all deployed AI agents, incorporating lessons from the pilot. Implement governance stress tests, senior partner accountability structures, and organizational narratives around governance success. Establish quarterly governance reporting cadence.

### Phase 5: Continuous Improvement (Ongoing)

The BGF is designed to evolve. As AI agents become more capable and take on higher-stakes functions, governance requirements will intensify. The framework should be reviewed at minimum semi-annually, with updates driven by empirical data on oversight quality, regulatory developments, and changes in the AI agent portfolio.

## 9. Regulatory Context and Future Outlook

The regulatory environment for agentic AI governance is maturing rapidly and unevenly. The EU AI Act classifies AI systems used in high-risk domains as subject to binding obligations for risk management, human oversight, transparency, and robustness, with phased implementation extending through 2027. Most agentic AI systems deployed in the PE investment lifecycle will likely fall within these high-risk categories given their role in consequential financial decision-making. Singapore's Model AI Governance Framework for Agentic AI provides detailed operational guidance that, while voluntary, is expected to influence regulatory expectations across Asia-Pacific markets. In the United States, the regulatory landscape remains fragmented, with the SEC's existing examination priorities around AI-driven investment advice and the Colorado AI Act (effective June 2026) providing early indicators of a more prescriptive approach.

For PE firms, the practical implication is that governance frameworks built today should be designed for tomorrow's regulatory requirements. The BGF's emphasis on documented accountability, structured oversight, and empirical governance metrics positions firms to meet regulatory expectations that are currently emerging, rather than scrambling to retrofit compliance mechanisms when formal obligations take effect.

Looking ahead, three trends will shape the governance landscape for agentic AI in private capital:

- **Multi-agent systems:** As firms deploy multiple AI agents that interact with each other—a deal sourcing agent feeding into a due diligence agent, which feeds into a portfolio monitoring agent—governance must account for emergent behaviors that arise from agent-to-agent interactions, not just individual agent performance.
- **LP scrutiny:** Limited partners are increasingly monitoring how their GPs deploy AI. Industry surveys indicate that nearly half of LPs are closely tracking GP AI adoption practices. Firms with robust, documented governance frameworks will have a competitive advantage in fundraising.
- **Insurance and liability:** The professional liability and errors-and-omissions insurance landscape for AI-assisted investment decisions is evolving. Firms with documented behavioral governance frameworks may benefit from more favorable underwriting terms as insurers develop risk models for agentic AI.

## 10. Conclusion and Recommendations

The private equity industry's adoption of agentic AI is accelerating faster than its governance infrastructure can accommodate. Current governance approaches, while technically sound, systematically underestimate the behavioral dimensions of human oversight in high-stakes, time-pressured environments. The result is a governance gap that will widen as AI agents take on more consequential functions across the investment lifecycle.

The Behavioral Governance Framework addresses this gap by integrating organizational psychology, decision science, and research on human-automation interaction into the design of AI oversight systems. Its three pillars—escalation architecture, trust calibration, and organizational incentive alignment—provide a structured approach to the human dimensions of agentic AI governance that complements existing technical safeguards.

The recommendations for PE firms deploying agentic AI are as follows:

1. Begin with a behavioral audit. Before deploying new governance mechanisms, understand how your people actually interact with AI outputs today. Map the gap between intended oversight behaviors and actual oversight behaviors.
2. Classify AI agents by consequentiality, not capability. The governance question is not how sophisticated the AI is, but how significant the consequences are if it fails. A highly capable agent performing low-stakes tasks requires less oversight than a moderately capable agent generating investment committee recommendations.
3. Design for the path of least resistance. Every governance mechanism should be evaluated against the question: will a busy, distracted professional under deal pressure actually use this? If the answer is no, redesign it.
4. Invest in governance metrics. If you do not measure oversight quality, you cannot manage it. Detection rates, engagement rates, and depth-of-review ratios are as important as deal throughput metrics.
5. Assign senior accountability. Governance that is owned by the compliance function alone will be treated as compliance—a box-checking exercise. Governance that is owned by named senior partners and reported at the management committee level will be treated as a strategic priority.
6. Budget capacity for oversight. Throughput gains from AI agents are real, but they create proportional oversight obligations. Plan for them.
7. Prepare for regulatory convergence. The current patchwork of voluntary frameworks and emerging regulations will converge toward binding requirements. Build governance infrastructure now that exceeds today's minimum requirements so that your firm is positioned for tomorrow's.

> *The firms that will lead in the age of agentic AI are not those that deploy the most sophisticated technology, but those that build the governance infrastructure to deploy it responsibly. The competitive advantage belongs to firms that can scale AI-augmented decision-making without degrading decision quality—and that is fundamentally a behavioral challenge, not a technical one.*

# References

Bahner, J. E., Hüper, A. D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. International Journal of Human-Computer Studies, 66(9), 688–699.

Cummings, M. L. (2017). Automation bias in intelligent time critical decision support systems. In D. Harris (Ed.), Decision Making in Aviation. Routledge.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. International Journal of Human-Computer Studies, 58(6), 697–718.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. Administrative Science Quarterly, 44(2), 350–383.

European Commission. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI Act).

Private Equity International. (2025). LP Perspectives 2026 Survey. PEI Media.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121–127.

IMDA Singapore. (2026). Model AI Governance Framework for Agentic AI. Infocomm Media Development Authority.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80.

McKinsey & Company. (2026). Global Private Markets Report 2026: Private Equity — Clearer View, Tougher Terrain.

NIST. (2023). Artificial Intelligence Risk Management Framework (AI 100-1). National Institute of Standards and Technology.

OWASP. (2025). OWASP Top 10 for Large Language Model Applications. Open Web Application Security Project.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3), 381–410.

PwC. (2026). Global M&A Trends in Private Equity and Principal Investors: 2026 Outlook.

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? International Journal of Human-Computer Studies, 51(5), 991–1006.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478.

## About the Author

**Dr. Leigh Coney** is the Founder and Principal Consultant of WorkWise Solutions. With a PhD in Organizational Psychology, Dr. Coney has spent over a decade at the intersection of AI, behavioral science, and organizational design. His research focuses on decision-making frameworks in high-stakes environments, with particular attention to why sophisticated AI systems fail to achieve adoption and how governance systems can be designed to account for human cognitive and social dynamics.

Dr. Coney's work is distinguished by its integration of behavioral science with practical technology deployment. He advises private equity firms, venture capital funds, and investment banks on AI strategy, governance design, and psychology-informed change management.

This paper is part of an ongoing research series on responsible AI adoption in financial services. Previous publications include "The Skill Erosion Paradox: Preserving Analytical Capability in AI-Augmented Teams" (Q1 2026).

## About WorkWise Solutions

WorkWise Solutions builds secure, purpose-built AI systems for private equity, venture capital, and investment banking firms. The firm specializes in zero-retention AI architecture that ensures proprietary deal flow and portfolio data never train public models.

WorkWise's approach is grounded in a core insight: most AI implementations fail not because of technology but because of broken workflows and poor adoption strategies. Every engagement integrates behavioral science and organizational psychology into the technical design, ensuring that AI systems become invisible, indispensable parts of how investment teams actually work.

**Contact:**

Contact@workwisesolutions.org | workwisesolutions.org

Schedule a consultation: calendly.com/contact-atqf/30min