WORKWISE SOLUTIONS

WHITE PAPER

Agentic AI in Private Equity

Multi-Agent Orchestration for End-to-End Deal Workflows

Dr. Leigh Coney

Founder & Principal Consultant, WorkWise Solutions

Q1 2026

workwisesolutions.org

Contact@workwisesolutions.org

Abstract

The term “agentic AI” has become the most overused phrase in enterprise software marketing. Every vendor now claims autonomous agents, yet most offerings amount to linear prompt chains wrapped in a loop. This paper separates the engineering reality from the vendor hype, presenting an architectural framework for deploying multi-agent systems across private equity deal workflows - from sourcing through investment committee preparation - with the verification patterns, failure modes, and human escalation protocols that production deployment actually requires.

Drawing on Anthropic’s published research demonstrating 90.2% outperformance of multi-agent versus single-model approaches, the NoLiMa benchmark findings on context window degradation, and established principles from distributed systems engineering, this paper introduces the Multi-Agent Orchestration Framework (MAOF). The MAOF provides PE firms with a practical architectural pattern for decomposing deal workflows into specialized agent roles with defined handoff protocols, confidence-based human escalation, and immutable audit trails. The framework addresses the specific challenges of PE environments: heterogeneous data formats, fiduciary accountability requirements, deal confidentiality constraints, and fixed investment committee deadlines.

This paper is intended for general partners, chief technology officers, operating partners, and technology advisors responsible for evaluating and deploying AI systems in investment workflows. It assumes familiarity with PE deal processes but not with AI systems architecture.

Keywords: Agentic AI, multi-agent systems, private equity, deal workflows, AI architecture, due diligence automation, investment committee, AI governance, context window, PE technology, AI orchestration, deal screening, portfolio monitoring

Table of Contents

1. Introduction: The Gap Between Demo and Production

Something unusual is happening in private equity technology. Every software vendor selling into PE firms now claims to offer “agentic AI” - autonomous systems that can source deals, screen opportunities, accelerate due diligence, and prepare investment committee memoranda with minimal human intervention. The marketing decks are impressive. The demos are polished. And the gap between what is demonstrated and what actually works in production, with confidential deal data, under fiduciary obligations, against fixed IC deadlines, is enormous.

This gap is not primarily a problem of AI capability. The underlying language models are genuinely powerful. As Andrej Karpathy observed in his widely discussed 2025 year-in-review, the industry has realized less than ten percent of the potential of current AI systems, even at their present capability level. The problem is architectural. Most systems marketed as “agentic” are, in engineering terms, linear prompt chains with a loop - a single large language model receiving a sequence of instructions, executing them in order, and returning results. This architecture works acceptably for simple tasks with clean inputs. It fails systematically when confronted with the complexity, heterogeneity, and stakes of PE deal workflows.

The failure modes are predictable. A single model processing an entire CIM loses coherence as its context window fills - the NoLiMa benchmark from Adobe Research and LMU Munich demonstrated that at 32,000 tokens, eleven out of twelve tested models dropped below fifty percent of their baseline performance (Modarressi et al., 2025). A monolithic agent architecture cannot maintain deal confidentiality when data from multiple opportunities flows through shared processing infrastructure. A system without structured verification protocols produces outputs that look authoritative but contain hallucinated data points - and in PE, a hallucinated EBITDA adjustment is not an inconvenience but a fiduciary failure.

This paper presents an alternative: the Multi-Agent Orchestration Framework (MAOF), an architectural pattern for deploying AI across PE deal workflows using distributed, specialized agents with defined handoff protocols, confidence-based human escalation, and immutable audit trails. The thesis is direct: agentic AI for PE requires distributed systems engineering discipline, not prompt engineering tricks. The architecture determines whether you get a demo or a deployment.

The scope of this paper covers deal workflows specifically - sourcing, screening, due diligence, and IC preparation. It does not address back-office automation, investor reporting, or fund administration, each of which presents its own architectural requirements. For governance frameworks applicable across the full deal lifecycle, see Coney (2026a); for measurement of AI’s contribution to investment performance, see Coney (2026b).

2. What “Agentic” Actually Means in a PE Context

2.1 The Autonomy Spectrum

The term “agentic” is used so loosely in vendor marketing that it has become nearly meaningless. To have a productive conversation about AI in PE deal workflows, we need a shared vocabulary for describing what level of autonomy a system actually operates at. The following four-level taxonomy, adapted for PE-specific contexts, provides this vocabulary.

Level 1: Assisted. The human initiates every step. The AI suggests completions, surfaces relevant data, or proposes text, but the human drives the workflow entirely. Examples include AI-powered search within a data room, draft paragraph suggestions in a memo, or automated formatting of financial tables. Most PE firms currently using AI operate at this level.

Level 2: Semi-Autonomous. The AI executes defined tasks end-to-end, but a human approves each output before it moves downstream. The system can extract financial data from a CIM, flag potential risk factors, or generate a preliminary deal summary, but a human reviews and validates before the output informs any decision. This level requires structured output schemas and clear quality criteria.

Level 3: Supervised Autonomous. The AI executes multi-step workflows with human oversight limited to exceptions. The system processes a complete CIM, produces a structured screening assessment, and routes the output directly to the deal team - unless it encounters low-confidence extractions, contradictory data, or inputs outside its training distribution, in which case it escalates to a human with a specific explanation of what it could not resolve. This level requires the confidence-based escalation and self-correction mechanisms described in Section 3.

Level 4: Fully Autonomous. The AI handles complete workflows from input to deliverable, with human oversight limited to final review of outputs. In theory, a Level 4 system could receive a CIM, produce a fully formed screening memo, and present it to the deal team without intermediate human intervention. In practice, Level 4 is not yet appropriate for fiduciary contexts in PE, though it may be viable for narrow, low-stakes subtasks within a larger supervised workflow.

> *Most PE firms are at Level 1. Marketing materials promise Level 4. Production-ready systems that balance speed, accuracy, and fiduciary accountability operate at Level 2–3. The gap between where firms are and where vendors claim they can be is where investment dollars are most commonly wasted.*

2.2 Why PE Deal Workflows Are Uniquely Challenging for Agents

AI agent architectures that perform well in other domains - customer service, content generation, software development - encounter specific failure modes when applied to PE deal workflows. Understanding these failure modes is essential for designing architectures that work in practice rather than in demonstration.

Heterogeneous data. A typical PE deal involves confidential information memoranda in PDF format (often scanned, inconsistently structured, and varying in quality from professionally typeset to hand-assembled); financial models in Excel with complex formula chains, hidden

sheets, and inconsistent naming conventions; legal documents in Word with nested definitions and cross-references; and virtual data rooms containing hundreds of files with no standardized organization. An agent architecture must handle all of these formats, recognize their relationships, and reconcile contradictory information across documents - challenges that compound as deal complexity increases.

High stakes with asymmetric consequences. In a customer service context, an AI error produces a frustrated customer and a support ticket. In PE due diligence, a missed covenant restriction or misread EBITDA adjustment flows into the investment thesis, the valuation model, and ultimately the bid price. The cost of a single undetected analytical error can exceed the total efficiency savings from AI across an entire fund. This asymmetry demands architectural patterns that prioritize verification over throughput - a design principle that runs counter to how most AI systems are optimized.

Confidentiality requirements. PE deals involve proprietary information protected by non-disclosure agreements and fiduciary obligations. Agent architectures that route data through shared cloud infrastructure, use multi-tenant processing endpoints, or retain information between sessions violate these requirements. Each agent in a multi-agent system must be independently secured, with ephemeral working memory and zero-retention data handling. This is not a feature request - it is a deployment prerequisite that eliminates many commercially available architectures from consideration.

Fixed deadlines with no tolerance for debugging. Investment committee meetings are scheduled weeks in advance. LOI deadlines are set by sellers. When an agent system fails - and all complex software systems fail - the failure cannot be debugged over days or weeks. The architecture must degrade gracefully: if an agent cannot complete its task, the system must escalate to a human with sufficient context and time for that human to complete the work manually. Systems that fail silently or fail catastrophically are unsuitable for PE deployment regardless of their capabilities when functioning correctly.

3. The Multi-Agent Orchestration Framework (MAOF)

3.1 Framework Overview

The MAOF is an architectural pattern for decomposing PE deal workflows into specialized agent roles with defined handoff protocols, structured verification checkpoints, and systematic human escalation. It is not a product, a specific technology stack, or a vendor solution - it is a set of design principles that can be implemented across multiple platforms and model providers. The framework draws on established patterns from distributed systems engineering, adapted for the specific requirements of PE investment workflows.

The core insight motivating the MAOF is empirical. Anthropic’s published research on their multi-agent research system demonstrated that a multi-agent architecture, using Claude Opus 4 as a lead agent coordinating Claude Sonnet 4 subagents, outperformed a single-agent Claude Opus 4 by 90.2 percent on research evaluation tasks (Anthropic, 2025). The performance advantage was most pronounced on breadth-first queries requiring simultaneous exploration of

multiple independent directions - precisely the pattern that characterizes PE due diligence, where financial, legal, operational, and market analyses must proceed in parallel. Their analysis further revealed that token usage alone explained eighty percent of performance variance, with tool call frequency and model choice accounting for additional factors. This finding has direct architectural implications: multi-agent systems outperform single agents primarily because they distribute work across separate context windows, adding reasoning capacity that a single context window cannot provide.

The MAOF contrasts with two common alternatives. The first is the single-agent approach, in which one large language model processes an entire deal workflow within a single context window. This approach fails at scale because context windows degrade: the NoLiMa benchmark showed that model performance drops precipitously as context length increases, with most models losing more than half their baseline accuracy at 32,000 tokens when tasks require non-literal reasoning (Modarressi et al., 2025). Chroma Research's 2025 investigation of context degradation confirmed this pattern across multiple model families and task types (Hong, Troynikov & Huber, 2025). A single agent processing a 200-page CIM, supporting financial model, and market analysis simultaneously is operating well beyond the context lengths where models maintain reliable performance.

The second alternative is the pipeline of independent tools - a series of disconnected AI-powered features (a document reader, a financial extractor, a summarizer) that do not communicate with each other. This approach fails because PE analysis is inherently integrative: a risk factor identified in the legal review may change the interpretation of a financial metric, which may affect the thesis alignment assessment. Disconnected tools cannot perform this cross-domain reasoning.

The MAOF addresses both failure modes through three core components: task decomposition with specialized agents, confidence-based human escalation, and self-correction loops with audit trails.

3.2 Component 1: Task Decomposition with Specialized Agents

The first principle of the MAOF is that each stage of a deal workflow should be handled by a dedicated agent with a narrow context, a specific output schema, and defined success criteria. Rather than loading an entire CIM and all supporting documents into a single context window, the workflow is decomposed into discrete tasks, each assigned to an agent that receives only the data it needs.

Consider a CIM screening workflow. Under the MAOF, this workflow is decomposed into five specialized agents:

Agent A: Document Ingestion and Structure Mapping. This agent receives the raw CIM and produces a structured map of its contents - identifying sections, tables, exhibits, and the relationships between them. Its output is a document schema, not an analysis. It succeeds when the schema accurately reflects the CIM's structure and all sections are identified and classified.

Agent B: Financial Data Extraction. This agent receives the sections identified by Agent A as containing financial information and extracts structured data: revenue figures, EBITDA, margins, growth rates, capital expenditure, working capital, and debt structures. Its output is a standardized financial data set in a defined schema. It succeeds when extracted figures match their source documents.

Agent C: Risk Flagging. This agent receives the full document schema from Agent A and performs a systematic risk assessment, flagging ESG concerns, customer and supplier concentration, regulatory exposure, litigation history, and operational risks. Its output is a structured risk register with severity classifications and source references.

Agent D: Thesis Alignment Scoring. This agent receives the outputs from Agents B and C and scores the opportunity against the fund's investment criteria - sector focus, size parameters, geographic preferences, growth profile, margin thresholds, and any fund-specific requirements. Its output is a structured alignment score with detailed justification for each criterion.

Agent E: Memo Synthesis. This agent receives the outputs from all upstream agents and produces a preliminary screening memorandum. It does not perform independent analysis - it synthesizes and structures the outputs from Agents A through D into a readable format, flagging any inconsistencies between upstream outputs.

Each agent operates within a clean, manageable context window. Agent B does not need to process legal disclosures; Agent C does not need financial models. This decomposition directly addresses the context degradation problem: no single agent is asked to maintain coherence across a volume of text that exceeds its reliable operating range. It also enables parallel execution - Agents B and C can operate simultaneously once Agent A has completed its structure mapping, compressing total processing time.

3.3 Component 2: Confidence-Based Human Escalation

Every agent output in the MAOF carries a confidence score. This is not optional safety theater appended to satisfy compliance requirements - it is the mechanism that makes Level 3 autonomy viable in fiduciary contexts.

The MAOF defines three confidence bands:

High confidence (above 0.85). The agent's output is presented directly to the analyst or downstream agent. The source locations are recorded in the audit trail, and the output is available for review, but no mandatory human verification step is imposed before the output moves downstream.

Medium confidence (0.60 to 0.85). The agent's output is flagged with specific source locations for analyst verification. The output is presented alongside the source material, with the uncertain elements highlighted. The analyst is asked to verify the specific flagged elements, not to review the entire output. This targeted verification preserves the time savings of automation while focusing human attention on the elements most likely to contain errors.

Low confidence (below 0.60). The task is escalated to a human with a structured explanation of what the agent attempted, what it could not resolve, and where in the source documents the relevant information appears. The escalation is not a vague “please review” - it is a specific request for human judgment on a defined question. This design ensures that escalated tasks can be resolved efficiently by analysts who receive context rather than raw documents.

The confidence thresholds above are proposed starting points; optimal thresholds should be calibrated to each firm’s error tolerance and the specific characteristics of each agent’s task. The critical design principle is that every output carries a confidence assessment, and that confidence determines the degree of human oversight - not a blanket policy of reviewing everything or reviewing nothing. For a detailed treatment of how confidence thresholds interact with automation complacency and verification atrophy, see Coney (2025).

3.4 Component 3: Self-Correction Loops with Audit Trails

The third component of the MAOF addresses a fundamental limitation of AI systems: they can be confidently wrong. An agent that extracts an EBITDA figure from a CIM may produce a precise number with high confidence that is, in fact, drawn from the wrong table or misinterpreted due to a footnote adjustment.

The MAOF addresses this through two mechanisms. First, agents verify their own outputs against source documents before passing results downstream. Agent B, for example, does not merely extract financial figures - it cross-references extracted values against the tables, charts, and text from which they were drawn, flagging discrepancies for re-evaluation or escalation. Second, downstream agents can query upstream agents for source confirmation. When Agent D (thesis alignment scoring) evaluates a margin threshold, it can request that Agent B confirm the specific source and context of the margin figure, rather than trusting the extracted value at face value.

Every decision, extraction, and score is logged with full provenance - which agent produced it, from which source document section, at what confidence level, and whether it was modified by human review. This creates an immutable audit trail that enables post-hoc review of any screening assessment or IC memo back to its source documents. The value of this trail extends beyond compliance: it enables systematic improvement of the agent system by identifying which agents produce the most errors, which document types cause the most difficulty, and which confidence thresholds need recalibration.

This stands in contrast to “black box” agent systems where the user receives a final output with no visibility into how conclusions were reached. In a fiduciary context, the ability to trace every analytical conclusion to its source is not a nice-to-have - it is a governance requirement. The WorkWise Verification Framework, detailed in prior publications (Coney, 2025; Coney, 2026a), provides the governance layer that sits atop the MAOF’s technical architecture.

4. Architecture Patterns for PE Deal Workflows

The MAOF can be instantiated in several architectural patterns, each suited to different workflow characteristics. This section describes four patterns, their appropriate use cases, and their respective tradeoffs.

4.1 Pattern 1: Sequential Pipeline (Deal Screening)

In the sequential pipeline pattern, agents execute in a defined order, with each agent passing its structured output to the next. The pipeline follows a strict sequence: Ingest, Extract, Score, Summarize. Failure at any stage halts the pipeline and triggers escalation rather than allowing corrupted data to propagate downstream.

This pattern is best suited for high-volume, standardized workflows where the input format is relatively consistent. A PE firm processing fifty to one hundred CIMs per quarter through an initial screening pass is a natural candidate. The sequential structure is simple to implement, easy to monitor, and straightforward to debug - each stage has a defined input and output, and failures are localized to specific agents.

The tradeoff is speed. Because each agent must complete before the next begins, total processing time is the sum of individual agent processing times. For initial screening, where the goal is to identify the twenty percent of opportunities worth deeper examination, this tradeoff is acceptable - reliability matters more than marginal speed gains at this stage.

4.2 Pattern 2: Parallel Fan-Out (Due Diligence)

In the parallel fan-out pattern, a lead agent analyzes the scope of the task and spawns multiple specialist agents that operate simultaneously. A due diligence workflow, for example, might spawn four parallel agents: a financial DD agent analyzing the quality of earnings, a legal DD agent reviewing material contracts and regulatory compliance, an operational DD agent assessing management capability and organizational structure, and a market DD agent evaluating competitive positioning and industry dynamics.

These agents operate independently, each with its own clean context window, and their results are synthesized by a coordinator agent that identifies cross-domain interactions - a legal risk that affects the financial model, an operational constraint that limits the market opportunity, or a regulatory change that alters the competitive landscape.

This pattern directly mirrors Anthropic’s production research system architecture, which uses an orchestrator-worker pattern where a lead agent coordinates specialized subagents operating in parallel (Anthropic, 2025). The key advantage is that each agent maintains a clean context window, avoiding the degradation that occurs when a single model attempts to hold financial, legal, operational, and market analysis simultaneously. As Anthropic’s research demonstrated, multi-agent systems excel specifically at tasks involving heavy parallelization and information volumes that exceed single context windows.

The tradeoff is coordination complexity. The coordinator agent must reconcile outputs from agents that may have operated on different assumptions, identified contradictory information, or reached conclusions that interact in non-obvious ways. This coordination is an engineering

challenge, not a prompt engineering challenge - it requires structured output schemas, defined reconciliation protocols, and explicit handling of inter-agent disagreements.

4.3 Pattern 3: Iterative Refinement (IC Memo Preparation)

In the iterative refinement pattern, a draft agent produces an initial output, a critique agent evaluates it against defined quality standards, and the draft agent revises based on the critique. This loop continues until a quality threshold is met or a maximum iteration count triggers human intervention.

IC memo preparation is the natural use case. The draft agent produces an initial memo from the deal data assembled during screening and diligence. The critique agent reviews the memo against the firm's IC standards: Are all required sections present? Does the investment thesis address the key risk factors? Are the financial projections internally consistent? Does the memo present a balanced assessment or read as advocacy? The draft agent then revises, and the cycle repeats.

This pattern is best suited for high-stakes deliverables that require judgment, not merely extraction and assembly. The critique agent serves a function analogous to a senior partner reviewing a junior analyst's work - not generating the analysis, but evaluating its completeness, coherence, and rigor. The iterative structure builds quality through successive refinement, producing outputs that are closer to what a human reviewer would accept.

The tradeoff is compute cost and latency. Each iteration of the draft-critique loop consumes additional tokens and processing time. Anthropic's research documented that multi-agent systems consume approximately fifteen times more tokens than standard chat interactions (Anthropic, 2025). For an IC memo - a document that may determine a multimillion-dollar investment decision - this cost is trivially justified. The architecture should include circuit-breaking logic that limits iterations and escalates to a human if the quality threshold is not reached within a defined number of cycles.

4.4 Pattern 4: Persistent Monitoring (Portfolio Surveillance)

The first three patterns operate on demand - triggered by the arrival of a new deal or the initiation of a workflow. The persistent monitoring pattern is different: agents run continuously, monitoring data feeds, financial reports, news, regulatory filings, and market signals for anomalies and early warning indicators across the portfolio.

In a portfolio surveillance context, monitoring agents track key performance indicators against forecast, detect deterioration patterns in financial metrics before they appear in quarterly reporting, flag regulatory or market developments that affect specific portfolio companies, and alert the portfolio management team when predefined thresholds are breached.

The value of persistent monitoring agents lies in early detection. A portfolio company's EBITDA deterioration that would not appear in standard quarterly reporting for another three months may be detectable through real-time analysis of customer churn data, supplier payment patterns, or industry sentiment shifts. Production implementations of this pattern have demonstrated early

detection windows of six or more weeks ahead of standard reporting cycles - time that can be used for proactive intervention rather than reactive crisis management.

The tradeoff is operational overhead. Persistent agents require infrastructure that runs continuously, processes streaming data, and manages alert fatigue - a well-known challenge in any monitoring system where excessive false positives cause users to ignore genuine alerts. The alert threshold calibration problem is as much a behavioral design challenge as an engineering one (Parasuraman & Manzey, 2010).

5. Security and Data Architecture for Agent Systems

5.1 Zero-Retention in Multi-Agent Contexts

The security architecture of a multi-agent system must account for the fact that each agent is a separate processing endpoint, and that data flows between agents create additional exposure surfaces that do not exist in single-agent systems. The MAOF addresses this through three principles.

First, each agent's working memory is ephemeral. When an agent completes its task, its context is discarded - no source documents, extracted data, or intermediate analyses persist in the agent's memory beyond the current session. Second, inter-agent communication uses structured schemas rather than raw documents. Agent B does not receive the full CIM from Agent A - it receives a structured reference to the financial sections, with the raw document data fetched directly from the secure document store on demand. This minimizes the volume of sensitive data in transit between agents. Third, audit logs record decisions and provenance metadata, not source content. The audit trail documents that Agent B extracted an EBITDA figure of a specific value from page 47 of the CIM; it does not store a copy of page 47.

5.2 Private Deployment Requirements

Multi-agent systems multiply the attack surface of AI deployments. A single-agent system has one processing endpoint to secure; a five-agent screening pipeline has five endpoints, plus the communication channels between them, plus the orchestration layer that coordinates their execution. Each of these components must be independently secured, and the entire system must achieve SOC 2 compliance across the full agent graph - not merely at the entry point.

For PE firms, this has practical implications for deployment architecture. Cloud-based multi-tenant agent systems, where multiple firms' data flows through shared processing infrastructure, are fundamentally incompatible with the confidentiality requirements of PE deal workflows. The architecture must support private deployment within the firm's security perimeter, or use dedicated single-tenant cloud instances with encryption at rest and in transit, rigorous access controls, and contractual zero-retention guarantees.

5.3 The Model Training Risk

Agent systems that improve over time must do so without incorporating proprietary deal data into model weights. This distinction is critical and frequently misunderstood. It is acceptable for an agent system to learn from interaction patterns - for example, learning that CIMs from a particular industry sector tend to structure financial information in a specific way, so that the ingestion agent can improve its document mapping. It is not acceptable for the system to fine-tune its models on the content of specific deals - the financial details, risk assessments, or investment theses contained in proprietary documents.

The boundary between these two categories is not always clean, and PE firms should require explicit technical documentation from any vendor about exactly how their system learns, what data contributes to model improvement, and what contractual and technical safeguards prevent proprietary deal content from entering training data. The governance requirements for AI model training in PE contexts are addressed in detail in Coney (2026a).

6. Failure Modes and Honest Limitations

Any responsible treatment of AI in PE must address not only what agent systems can do, but where they fail and what they cannot yet do at all. The following failure modes are not hypothetical - they are recurring patterns observed in production AI deployments.

6.1 Cascading Errors

In a multi-agent system, an error in an upstream agent propagates to every downstream agent that consumes its output. If Agent A (document ingestion) misclassifies a section of the CIM - identifying a pro forma adjustment table as historical financial data, for example - Agent B will extract the wrong figures, Agent D will score the opportunity against incorrect financial metrics, and Agent E will produce a memo that presents fabricated numbers with full confidence.

The MAOF mitigates this through cross-verification checkpoints - downstream agents are designed to validate upstream outputs rather than accept them on faith. But this mitigation is imperfect. Cross-verification increases computational cost, adds latency, and cannot catch all errors, particularly those involving plausible but incorrect categorizations. The honest assessment is that cascading errors remain the most significant risk in multi-agent architectures, and no current design eliminates them entirely.

6.2 Coordination Overhead

More agents do not always mean better performance. Each agent added to a system increases the coordination overhead: more communication between agents, more opportunities for miscommunication, more latency from waiting for outputs, and more complexity in debugging failures. Anthropic’s research documented that some domains with heavy inter-agent dependencies are not a good fit for multi-agent systems, and that early versions of their system made errors such as spawning fifty subagents for simple queries (Anthropic, 2025).

The optimal number of agents for a given workflow is the minimum number that maintains clean context windows and appropriate specialization - not the maximum possible. A screening

workflow might require five agents; a simple financial extraction might require two. The architecture should be designed to scale agent count based on task complexity, not to apply a fixed multi-agent architecture to every problem regardless of whether it benefits from decomposition.

6.3 The Demo-to-Production Gap

Agent demonstrations typically operate on clean, well-structured sample data - a professionally typeset CIM with clear section headings, consistent formatting, and complete information. Production PE data is the opposite: scanned PDFs with OCR artifacts, Excel models with circular references and hidden sheets, data rooms where the same financial metric is stated differently in three documents, and CIMs with material information buried in footnotes.

The engineering challenges of handling production data - state management across agent interactions, error handling for malformed inputs, graceful degradation when a document cannot be parsed, retry logic for intermittent failures - are genuine distributed systems engineering problems. They are not solved by better prompts. Firms evaluating agent systems should insist on testing with their own historical deal data, including the messiest examples they can find, rather than accepting demonstrations on curated sample sets.

6.4 What Agents Cannot Do (Yet)

There is a category of judgment that current AI systems cannot perform and that should not be delegated to agents, regardless of architectural sophistication. These include:

Relationship judgment. Assessing whether a management team has the capability, cohesion, and drive to execute a value creation plan requires interpersonal perception that AI systems do not possess. Agents can assemble a management team’s track record from public and proprietary sources, but they cannot assess the dynamics of a management presentation or the credibility of a CEO’s strategic vision.

Negotiation dynamics. Evaluating whether a seller’s EBITDA adjustment is defensible, or whether a specific deal term is achievable in the current competitive context, requires judgment informed by years of transaction experience, market awareness, and interpersonal reading. Agents can provide the analytical foundation - comparable transaction data, precedent analysis, contractual benchmarks - but the judgment itself remains human.

Strategic conviction. The most consequential decision in PE is not analytical but strategic: Does this deal, at this price, in this market, fit our portfolio construction thesis and warrant the commitment of our capital and management attention? This is a judgment that integrates quantitative analysis with qualitative assessment, pattern recognition from decades of experience, and institutional self-knowledge. No agent architecture can replicate it.

> *The purpose of agent systems in PE is not to replace these forms of judgment. It is to handle the data work - extraction, structuring, analysis, synthesis - so that the humans*

responsible for these judgments can focus their time and cognitive capacity on the decisions that actually determine investment returns.

7. Competitive Implications and Conclusion

The private equity firms deploying agent systems at Level 2–3 today are building institutional capability that compounds over time. This compounding is not primarily technological - it is organizational. Each deal cycle in which a team works with an agent system produces learning: the team develops intuition for where agents add value and where they fail; the agents are calibrated based on observed errors and edge cases; the confidence thresholds are refined based on the firm's specific error tolerance; and the workflow decomposition is adjusted based on the types of deals the firm typically evaluates.

This organizational learning is the genuine first-mover advantage in agentic AI for PE. It is not about which firm has the most advanced model - models improve rapidly and are available to all firms on similar terms. It is about which firm has developed the most effective integration between its AI systems and its human decision-making processes. That integration cannot be purchased from a vendor. It must be built through the iterative process of deploying, measuring, adjusting, and redeploying across multiple deal cycles. The Stanford HAI AI Index Report 2025 documented that while business adoption of AI accelerated to seventy-eight percent of organizations in 2024, translating adoption into measurable value remains the central challenge (Stanford HAI, 2025). In PE, where value is measured in fund returns, the gap between adoption and value creation is particularly consequential.

The DVQF framework introduced in our prior paper (Coney, 2026b) provides the measurement infrastructure to track this organizational learning. Throughput metrics tell you how fast your agents operate; analytical depth metrics tell you whether their outputs are trustworthy; outcome attribution metrics, over time, tell you whether the combination of agent and human intelligence produces better investment decisions than either alone.

The architectural requirements outlined in this paper - task decomposition, confidence-based escalation, self-correction loops, zero-retention security, and graceful degradation - are not optional features that can be added later. They are foundational design decisions that determine whether an agent deployment scales beyond its initial use case or collapses under the weight of production complexity. Firms that treat agent deployment as a prompt engineering exercise will build demos. Firms that treat it as a distributed systems engineering challenge will build capabilities.

The firms that understand this distinction will deploy agents that scale. Those that don't will deploy demos.

References

Anthropic. (2025). How we built our multi-agent research system. Anthropic Engineering Blog. https://www.anthropic.com/engineering/multi-agent-research-system

Coney, L. (2025). Combating automation complacency in financial due diligence: A deep dive into verification atrophy, cognitive interventions, and interface design for epistemic humility. WorkWise Solutions White Paper Series. DOI: 10.2139/ssrn.6111107

Coney, L. (2026a). AI governance across the deal lifecycle: From sourcing through portfolio monitoring. WorkWise Solutions White Paper Series. DOI: 10.2139/ssrn.6274559

Coney, L. (2026b). Measuring AI ROI in private equity: A framework for decision velocity vs. decision quality. WorkWise Solutions White Paper Series.

Coney, L. (2026c). The skill erosion paradox: Preserving analytical capability in AI-augmented teams. WorkWise Solutions White Paper Series.

Hong, K., Troynikov, A., & Huber, J. (2025). Context rot: How increasing input tokens impacts LLM performance. Chroma Research. https://research.trychroma.com/context-rot

Karpathy, A. (2025). 2025 LLM year in review. https://karpathy.bearblog.dev/year-in-review-2025/

Modarressi, A., Deilamsalehy, H., Dernoncourt, F., Bui, T., Rossi, R. A., Yoon, S., & Schütze, H. (2025). NoLiMa: Long-context evaluation beyond literal matching. Proceedings of the International Conference on Machine Learning (ICML 2025). arXiv:2502.05167

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

Stanford Institute for Human-Centered Artificial Intelligence. (2025). The 2025 AI Index Report. Stanford University. https://hai.stanford.edu/ai-index/2025-ai-index-report

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 their impact on organizational performance can be rigorously measured.

This paper is part of an ongoing research series on responsible AI adoption in financial services. Previous publications include “Closing the Accountability Gap: A Governance Framework for AI in Private Equity, Venture Capital, and Strategic Consulting” (December 2025), “Combating Automation Complacency in Financial Due Diligence” (Q1 2026), “The Skill Erosion Paradox: Preserving Analytical Capability in AI-Augmented Teams” (Q1 2026), “AI Governance Across the Deal Lifecycle” (Q1 2026), and “Measuring AI ROI in Private Equity” (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