WORKWISE SOLUTIONS

WHITE PAPER

Measuring AI ROI in Private Equity

A Framework for Decision Velocity vs. Decision Quality

Dr. Leigh Coney

Founder & Principal Consultant, WorkWise Solutions

Q1 2026

workwisesolutions.org

Contact@workwisesolutions.org

Abstract

Private equity firms are investing aggressively in AI-powered deal sourcing, due diligence, and portfolio monitoring. Yet the industry lacks a coherent framework for measuring whether these investments generate genuine returns. The dominant metrics—deal throughput, time-to-completion, and analyst hours saved—capture decision velocity but ignore decision quality. A deal team that screens three times as many opportunities per quarter has not created investment value if the additional screening produces no incremental alpha, even if it provides secondary benefits in market intelligence and deal flow coverage. Conversely, an AI system that slows a single diligence process by introducing a previously undetected risk factor may generate outsized returns that never appear in a speed metric.

This paper introduces the Decision Velocity–Quality Framework (DVQF), a measurement model designed specifically for private equity's investment lifecycle. The DVQF provides a structured methodology for evaluating AI's impact across four dimensions: throughput efficiency, analytical depth, outcome attribution, and risk-adjusted return contribution. Drawing on established research in decision science, performance measurement, and organizational behavior, the framework addresses the critical gap between what PE firms currently measure and what actually determines whether AI creates or destroys investment value.

The paper proposes specific metrics, measurement protocols, and implementation guidance for general partners, operating partners, and chief technology officers responsible for justifying and optimizing AI investments across the fund lifecycle.

Keywords

AI ROI, private equity, decision quality, decision velocity, AI value creation, investment decision-making, deal screening AI, due diligence automation, portfolio monitoring, AI measurement framework, AI productivity metrics, PE operating model, AI adoption, performance measurement, investment committee

Table of Contents

1. Introduction: The Measurement Crisis in PE AI

The private equity industry is spending heavily on artificial intelligence. Firms are deploying AI systems across deal sourcing, due diligence acceleration, portfolio monitoring, and investor reporting (McKinsey, 2026; PwC, 2026). The operational case for these investments appears strong: deal teams can process more information, identify opportunities faster, and compress timelines that previously stretched over weeks into days or hours.

Yet a striking disconnect has emerged between AI spending and the ability to measure its returns. A 2025 IBM Institute for Business Value study of 2,000 global CEOs found that only 25% of AI initiatives had delivered expected ROI over the preceding three years, and just 16% had achieved enterprise-wide scale (IBM IBV, 2025a). While this study covered all industries, the challenge is amplified in private equity, where the relationship between operational efficiency and investment performance is indirect and often ambiguous. A deal team that processes CIMs faster has not necessarily selected better investments. A portfolio monitoring system that detects performance deterioration earlier has not necessarily generated better exit outcomes. Speed and quality are related but distinct—and the industry's measurement infrastructure conflates them.

The problem is not that PE firms lack data. It is that they measure the wrong things. The dominant AI performance metrics in PE today—deals screened per quarter, hours saved per deal, time-to-IC-memo—are velocity metrics. They quantify how much faster the process runs with AI. They do not quantify whether faster processes produce better investment decisions. This distinction matters enormously in an asset class where a single misjudged investment can offset years of operational efficiency gains (Kaplan & Schoar, 2005), and where the best deals are often the ones that required the deepest, most time-intensive analysis.

This paper introduces the Decision Velocity–Quality Framework (DVQF), a measurement model designed to help PE firms evaluate AI's actual contribution to investment performance—not merely its contribution to operational speed. The framework draws on established research in decision science (Kahneman, Sibony & Sunstein, 2021; Simon, 1955) and performance measurement, applied to the specific characteristics of the PE investment lifecycle.

A methodological note: while the DVQF draws on well-established principles of decision analysis, performance measurement, and organizational behavior, there is limited published empirical research on AI ROI measurement specific to PE investment workflows. The framework is designed as a practical tool based on these principles, and its specific metrics should be calibrated to each firm's context through implementation experience. The Bain & Company and StepStone Group 2026 GP Outlook, which found that roughly 40% of GPs do not expect material AI financial impact in 2026, underscores the urgency of developing rigorous measurement approaches.

2. What PE Firms Currently Measure—and Why It's Insufficient

2.1 The Velocity Bias in AI Measurement

When PE firms evaluate AI investments today, they overwhelmingly rely on speed and volume metrics. This is understandable: speed is easy to measure, improvement is easy to demonstrate, and the results are immediately visible to partners and investors. Typical metrics include the number of deals screened per analyst per quarter, the time required to produce a due diligence summary, the number of data room documents processed per hour, and the turnaround time from CIM receipt to preliminary investment committee memorandum.

These metrics are not useless. They capture a genuine dimension of AI value: operational efficiency. A deal team that can screen 200 opportunities per quarter instead of 60 has more optionality. A due diligence process that compresses from three weeks to three days frees capacity for other activities. But velocity metrics alone create three specific problems for PE firms.

Problem 1: Speed Does Not Equal Selection Quality

The fundamental question in PE deal screening is not "how many deals can we evaluate?" but "are we selecting the right deals from the ones we evaluate?" (Hammond, Keeney & Raiffa, 1998). An AI system that enables a deal team to screen three times more opportunities has created genuine value only if the expanded screening surface yields investments that the team would not otherwise have found. If the additional deals screened are no better than those the team would have found through traditional channels, the AI investment has generated operational efficiency without investment value.

Problem 2: Compression Can Degrade Analytical Depth

Time compression in due diligence is presented as unambiguously positive, but it can come at a cost. A manual due diligence process that takes three weeks includes substantial "soak time"—periods during which the deal team absorbs information, develops intuitions, identifies patterns, and surfaces concerns that emerge only through extended engagement with the material. The value of cognitive incubation in complex problem-solving is well-established in psychological research

(Kahneman, 2011), and practitioners widely report that extended immersion in deal materials surfaces insights that compressed analysis does not. An AI-accelerated process that compresses the same work into three days may produce a technically complete analysis while eliminating the cognitive incubation that often generates the most valuable insights. Measuring the time saved without measuring the analytical quality of the output creates a misleading picture of AI's contribution.

Problem 3: Throughput Metrics Mask Error Introduction

AI systems introduce new categories of error that do not exist in manual processes (Parasuraman & Manzey, 2010). These include hallucinated data points in AI-generated summaries, systematic biases inherited from training data, and pattern-matching errors where the AI applies an analytical framework that is technically appropriate but contextually wrong. When the primary measurement is throughput, these errors may go undetected unless they produce visible failures downstream. For illustration, consider a deal screening system that processes 200 opportunities per quarter with a 2% systematic error rate: it has introduced four potentially flawed assessments that would not have existed in a manual process handling 60 deals with human-generated analysis. The velocity metric shows a 3.3x improvement; the error metric, which is unmeasured, tells a different story.

2.2 What Is Missing: Quality Metrics

The metrics that would actually capture AI's investment value—selection accuracy, analytical depth, error rate, and contribution to risk-adjusted returns—are largely absent from how PE firms evaluate their AI investments. This is not because firms do not care about quality. It is because quality metrics in investment decision-making are genuinely difficult to construct, require long feedback loops (the quality of a deal selection is not known for four to seven years), and demand attribution methodologies that isolate AI's contribution from the many other factors that determine investment outcomes (Einhorn & Hogarth, 1981). The DVQF is designed to address these challenges.

3. The Decision Velocity–Quality Framework (DVQF)

3.1 Framework Overview

The DVQF evaluates AI's contribution to PE investment performance across four dimensions. Each dimension captures a distinct aspect of how AI affects the investment process, and each requires its own measurement methodology:

Dimension	What It Measures	Core Question

I. Throughput Efficiency	Speed and volume gains from AI deployment across the deal lifecycle.	How much faster and more broadly does the team operate with AI?
II. Analytical Depth	Quality and comprehensiveness of AI-assisted analysis compared to manual baselines.	Is the analysis better, worse, or equivalent to what humans produced alone?
III. Outcome Attribution	Degree to which AI-influenced decisions produce superior investment outcomes.	Did AI improve actual investment results, not just the process?
IV. Risk-Adjusted Return Contribution	AI's net impact on fund-level returns after accounting for implementation costs and error costs.	After all costs and risks, is the fund better off with AI than without it?

3.2 Design Principles

The DVQF is built on four design principles that reflect the specific characteristics of PE decision-making:

- **Long feedback loops:** PE investment outcomes take four to seven years to materialize. The framework incorporates leading indicators (measurable within weeks or months) and lagging indicators (measurable at exit) to provide actionable measurement at each stage of the fund lifecycle.
- **Multi-causal outcomes:** Investment performance is determined by market conditions, operational execution, competitive dynamics, and macroeconomic factors in addition to deal selection and diligence quality. The framework uses counterfactual and comparative methodologies to isolate AI's specific contribution.
- **Asymmetric consequences:** In PE, the cost of a bad decision (writing off an investment) vastly exceeds the benefit of a marginal efficiency gain. The framework weights error costs and risk avoidance more heavily than throughput improvements.
- **Behavioral integration:** AI does not make investment decisions in isolation; it operates within human decision-making processes subject to well-documented cognitive biases (Tversky & Kahneman, 1974). The framework measures AI's influence on human decisions, not just the quality of AI outputs in isolation.

4. Dimension I: Throughput Efficiency

Throughput efficiency is the dimension most PE firms already measure, but the DVQF refines how it should be measured and—critically—establishes the conditions under which throughput gains actually translate to value.

4.1 Core Metrics

- **Deal Screening Velocity (DSV):** Number of deals screened per analyst per quarter, before and after AI deployment. For DVQF purposes, a 'screened deal' is defined as an opportunity that has received sufficient analysis to produce a documented pass/advance recommendation; firms should establish a consistent definition that reflects meaningful analytical engagement, not merely automated ingestion. DSV should be reported alongside selection rate (the percentage of screened deals that advance to preliminary evaluation) to distinguish between productive expansion and undifferentiated volume.
- **Diligence Cycle Time (DCT):** Elapsed time from data room access to completed diligence summary, measured in business days. Track separately for AI-assisted and manual processes during the transition period to establish comparative baselines.
- **Capacity Release Rate (CRR):** Percentage of analyst hours freed by AI automation that are reallocated to higher-value activities (deeper analysis, relationship development, original research) versus absorbed by increased throughput targets. As a proposed starting point, a CRR below 20% suggests that AI speed gains are being consumed by volume expansion rather than quality improvement; this threshold should be calibrated to each firm's context through baseline measurement, as empirical benchmarks for AI capacity reallocation in PE do not yet exist.
- **Cost Per Screened Deal (CPSD):** Total cost of AI infrastructure, licensing, and human oversight per deal screened. For firms using integrated AI platforms that serve multiple functions, CPSD should be calculated using activity-based cost allocation—apportioning platform costs based on the proportion of compute, API calls, or user time attributable to screening activities specifically. This metric prevents the illusion of free throughput; AI-generated analysis has real costs that must be compared against the cost of manual analysis on a per-unit basis.

4.2 Interpretation Guidance

Throughput metrics should never be reported in isolation. A deal team that reports a threefold increase in deals screened per quarter without corresponding data on selection quality and outcome attribution has demonstrated that AI accelerates process, not that it creates value. The DVQF requires that throughput metrics always be reported alongside at least one quality metric from Dimensions II or III.

> *The velocity trap: when faster screening becomes the goal rather than the means, firms optimize for throughput at the expense of the analytical depth that drives returns. Velocity is valuable only when it expands the opportunity set without degrading selection quality.*

5. Dimension II: Analytical Depth

Analytical depth measures whether AI-assisted analysis is as comprehensive, accurate, and insightful as the manual analysis it replaces or augments. This dimension captures the quality cost of speed—the dimension that velocity metrics alone cannot reveal.

5.1 Core Metrics

- **Coverage Completeness Index (CCI):** For a sample of AI-assisted due diligence outputs, measure the percentage of material risk factors, value drivers, and analytical dimensions covered, compared to a benchmark set derived from the firm's best manual diligence processes. As a proposed starting point, a CCI below 85% suggests the AI is producing superficially complete but substantively shallow analysis; this threshold should be calibrated based on the firm's own baseline quality audit.
- **Error Detection Rate (EDR):** Percentage of AI-generated analytical outputs that contain material errors detected during human review. Track over time: a declining EDR indicates the AI is improving; a stable or rising EDR with declining review intensity (fewer human hours spent reviewing) indicates that errors are being missed, not eliminated.
- **Novel Insight Rate (NIR):** Percentage of AI-assisted analyses that surface a material insight, risk factor, or analytical connection that the deal team did not identify independently. This metric captures AI's contribution to analytical value-add, not just analytical replication. Track using post-analysis surveys or structured debrief protocols with deal team members.
- **Human Override Rate (HOR):** Percentage of AI-generated recommendations or assessments that human reviewers substantively modify before use. An HOR near zero may indicate rubber-stamping rather than accuracy—a phenomenon known as automation complacency (Parasuraman & Manzey, 2010; see also Coney, 2026, for a detailed treatment of behavioral governance design for AI oversight in PE). An HOR above 40% may indicate the AI outputs require so much correction that the time savings are illusory. These thresholds are proposed starting points; the

healthy range depends on the task and should be calibrated through baseline measurement.

5.2 Measuring What AI Removes

A critical and frequently overlooked dimension of analytical depth is what AI compression eliminates from the analytical process. Manual due diligence includes activities that do not appear in a formal deliverable but contribute to analytical quality: informal conversations between team members about patterns they are noticing, the cognitive “soak time” during which analysts develop intuitions about a business, and the serendipitous connections that emerge when an analyst spends days immersed in a data room rather than minutes reviewing an AI summary.

These activities are difficult to measure directly. The DVQF recommends using deal team retrospectives as a qualitative supplement to quantitative metrics. After each deal that progresses beyond preliminary screening, the deal team should conduct a structured debrief that includes the question: “What, if anything, did we learn from the AI-assisted analysis that we would not have learned from a manual process? What, if anything, did we miss because the AI-assisted process was faster?” These qualitative inputs should be logged and analyzed for patterns over time.

The CRR metric (Dimension I) measures whether freed capacity is reallocated to quality-enhancing activities. However, CRR alone does not capture whether AI compression eliminates cognitive incubation that cannot be replaced by additional hours after the fact—the insight emerges from the process of slow engagement, not from time tacked on after a fast summary. The qualitative retrospective described above is designed to surface this more subtle dimension of quality loss.

6. Dimension III: Outcome Attribution

Outcome attribution is the most methodologically challenging dimension of AI ROI measurement in PE. The question it addresses is fundamental: did AI-assisted decisions produce better investment outcomes than non-AI-assisted decisions would have? Answering this question requires navigating long feedback loops, multi-causal outcomes, and small sample sizes.

6.1 The Attribution Challenge

In most industries, AI ROI can be measured through A/B testing: run the AI-assisted process alongside the manual process, compare outcomes, and attribute the difference to the AI. In PE, this approach faces severe constraints. Fund-level sample sizes are small (a typical buyout fund makes 10–15 platform investments over a fiveyear investment period). Outcomes take years to materialize. And the counterfactual—what would have happened without AI—is inherently unobservable for any specific investment.

The DVQF addresses these constraints through a combination of process-level attribution (measurable in the near term) and outcome-level attribution (measurable at exit), using comparative rather than experimental methodologies.

6.2 Process-Level Attribution Metrics (Leading Indicators)

- **AI-Surfaced Deal Conversion Rate:** For deals sourced or identified primarily through AI systems, track the conversion rate through each stage of the deal funnel (preliminary screen to deep dive, deep dive to LOI, LOI to close) compared to deals sourced through traditional channels. A higher conversion rate for AI-surfaced deals suggests the AI is identifying higher-fit opportunities.
- **Risk Factor Discovery Index:** For completed due diligence processes, count the number of material risk factors identified by the AI that were subsequently validated by the deal team, compared to the number identified through manual analysis alone. A positive differential suggests AI is expanding the analytical surface in ways that reduce the probability of undetected risks.
- **IC Decision Confidence Score:** After each investment committee decision, survey committee members on their confidence in the quality and completeness of the information available to them. Track over time and compare periods before and after AI deployment. This is a subjective metric that should be interpreted with caution: research on expert calibration demonstrates that confidence and accuracy are often poorly correlated (Kahneman, 2011), and AI outputs may increase confidence through comprehensiveness of presentation rather than accuracy of content. The IC Decision Confidence Score is most useful as a trend indicator over time, not as a standalone measure of information quality.

6.3 Outcome-Level Attribution Metrics (Lagging Indicators)

- **Vintage Comparison Analysis:** Compare the performance of investments made with AI assistance (post-deployment vintage) to investments made without AI assistance (pre-deployment vintage), controlling for market conditions, sector mix, and fund maturity. Vintage comparison should also control for changes in the risk profile of selected investments, since AI-assisted selection may systematically shift the portfolio toward or away from certain risk-return profiles compared to historical vintages. This requires patience—meaningful data will not be available until the post-deployment vintage has had sufficient time to mature—but it is the most direct measure of AI's impact on investment outcomes.

- **Avoided Loss Attribution:** Track instances where AI analysis identified a risk factor that contributed to a decision to decline or restructure an investment, and where subsequent developments validated the concern. The value of a deal not done—or a deal restructured to mitigate a specific risk—is difficult to quantify precisely but represents one of AI's most significant potential contributions to fund performance. Importantly, this metric captures only one side of the ledger—validated risk avoidance. Firms should supplement it with periodic review of deals declined based on AI analysis that subsequently performed well for other acquirers, to calibrate the AI's false positive rate on risk flagging and prevent systematic overestimation of avoided loss value.
- **Portfolio Alert Accuracy:** For AI-powered portfolio monitoring systems, track the percentage of early warning alerts that correspond to genuine performance issues subsequently confirmed through standard reporting. Also track the lead time—how many weeks or months before standard reporting did the AI flag the issue? The combination of accuracy and lead time quantifies the monitoring system's value to the GP.

7. Dimension IV: Risk-Adjusted Return Contribution

The fourth dimension integrates the preceding three into a single question: after accounting for all costs, risks, and benefits, is the fund better off with its AI investments than it would be without them?

7.1 The Full Cost Picture

AI ROI calculations in PE typically undercount costs. The DVQF requires a comprehensive cost accounting that includes direct costs (technology licensing, infrastructure, data procurement), implementation costs (integration engineering, workflow redesign, training), ongoing operational costs (human oversight, quality assurance, system maintenance), and—critically—error costs (the estimated financial impact of AI-introduced errors, including both detected errors that required remediation and undetected errors whose costs manifest downstream).

Error costs are the most frequently omitted category and potentially the most significant. An AI-generated due diligence summary that contains a flawed EBITDA adjustment, if not caught, flows into the investment thesis, the valuation model, and ultimately the bid price. The cost of that single error may exceed the total savings from AI-assisted diligence across the entire fund. As IBM IBV research has documented, even among organizations reporting AI productivity gains, translating those gains into measurable financial impact remains the central challenge (IBM IBV, 2025b). Error costs for detected errors can be estimated at the time of detection; error costs for undetected errors are inherently uncertain and face the same long-feedback-loop challenges as outcome attribution. In early-stage measurement, the error cost component should be treated as a risk-weighted estimate based on the Error Detection Rate and the firm's assessment of the potential magnitude of undetected errors.

7.2 The DVQF Return Calculation

The DVQF proposes a modified return calculation that captures AI's net contribution:

> *AI Net Value Contribution = (Throughput Value + Analytical Value + Outcome Value) – (Direct Costs + Implementation Costs + Operational Costs + Error Costs)*

Each component is defined as follows:

- **Throughput Value (Dimension I):** The dollar value of capacity freed by AI, measured as the cost-equivalent of analyst hours reallocated to higher-value activities. Hours consumed by increased throughput targets rather than reallocated to quality improvement are excluded.
- **Analytical Value (Dimension II):** The estimated value of AI-generated insights that contributed to investment decisions, measured through the Novel Insight Rate and Coverage Completeness Index. This is inherently an estimate, but structured deal team retrospectives can provide reasonable approximations.
- **Outcome Value (Dimensions III and IV):** The attributed financial impact of AI on investment outcomes, integrating process-level leading indicators (AI-Surfaced Deal Conversion Rate, Risk Factor Discovery Index, IC Decision Confidence Score, Portfolio Alert Accuracy) with outcome-level lagging indicators (Vintage Comparison Analysis, Avoided Loss Attribution). This component will be unavailable or highly uncertain in the early years of deployment and becomes more reliable as the post-deployment track record develops.

The DVQF explicitly acknowledges that a precise calculation of AI Net Value Contribution is not achievable in the early years of deployment. The framework's value lies not in producing a single ROI number but in structuring the conversation about AI value around the right dimensions and metrics, rather than defaulting to velocity metrics that capture only a fraction of AI's impact.

8. The Velocity–Quality Tradeoff Matrix

The four dimensions of the DVQF interact in predictable ways. The Velocity–Quality Tradeoff Matrix maps the most common patterns observed when PE firms deploy AI, based on the relative performance across the throughput and quality dimensions:

		Decision Quality	
		High	Low
Decision Velocity	High	Quadrant A: AI Alpha. Speed + quality. The target state. AI expands the opportunity set and improves selection.	Quadrant B: The Velocity Trap. Fast but shallow. Throughput masks analytical degradation. Most common failure mode.
	Low	Quadrant C: Deep but Slow. Quality preserved but speed gains unrealized. Often caused by excessive oversight or poor integration.	Quadrant D: Worst Case. AI adds cost without improving speed or quality. Typically indicates fundamental workflow misfit.

The matrix provides a diagnostic tool for evaluating AI deployments. Most firms begin in what we term a ‘default Quadrant B’ state—achieving immediate speed gains while quality metrics remain unmeasured. Without quality measurement, these firms cannot distinguish between Quadrant A (the target state) and Quadrant B (the velocity trap). The DVQF is designed to resolve this ambiguity, enabling firms to course-correct toward Quadrant A before velocity-driven degradation becomes embedded in the operating model.

Quadrant C, while less common, is particularly relevant for firms with strong governance cultures that may over-apply human review to AI outputs. In these firms, the AI is producing high-quality analysis that is then extensively re-reviewed by humans, negating the speed benefit without adding proportional quality improvement. The DVQF’s Human Override Rate metric helps identify this pattern.

9. Implementation: Building a Measurement Infrastructure

Implementing the DVQF requires organizational commitment beyond selecting metrics and building dashboards. It requires changes to how deal teams document their work, how technology teams instrument AI systems, and how firm leadership evaluates AI investments. The DVQF is designed to be implemented incrementally—firms should begin with the metrics most relevant to their specific AI deployment and expand measurement scope as the infrastructure matures. Not all metrics need to be tracked from day one. The measurement system itself has costs that should be considered when evaluating the net value of AI governance and performance infrastructure.

Phase 1: Baseline Establishment (Weeks 1–6)

Before deploying any AI measurement, establish baselines for each metric in the pre-AI or current-state environment. This is the most frequently skipped step and the most consequential omission. Without a reliable baseline, no amount of post-deployment measurement can determine whether AI improved or degraded performance.

- **Process baselines:** Document current deal screening volumes, diligence cycle times, analyst hours per deal, and cost per screened deal. These baselines must reflect actual practice, not aspirational targets.
- **Quality baselines:** Conduct a retrospective quality audit on a sample of recent, completed diligence processes. Assess coverage completeness, error rates, and the number of material risk factors identified. This provides the benchmark against which AI-assisted quality will be measured.
- **Outcome baselines:** Compile historical data on deal funnel conversion rates, IC decision patterns, and portfolio company performance by vintage. These data will be essential for outcome attribution analysis as the post-AI track record develops.

Phase 2: Instrumentation (Weeks 7–12)

Configure AI systems and deal management platforms to capture the data required for DVQF metrics. This includes logging all AI-generated outputs with timestamps and version control, tracking human modifications to AI outputs (what was changed, by whom, and why), recording deal team assessments of AI contribution through structured debrief protocols, and establishing the sampling and audit procedures that will support ongoing quality measurement.

Phase 3: Initial Measurement (Months 4–12)

During the first year of measurement, focus on Dimensions I and II (throughput efficiency and analytical depth). These dimensions provide actionable data on relatively short timescales and enable early course correction. Dimensions III and IV (outcome attribution and risk-adjusted return contribution) require longer time horizons and should be tracked but not expected to yield definitive results in the first year.

Phase 4: Mature Measurement (Year 2+)

As the post-deployment track record develops, begin incorporating outcome-level metrics. The vintage comparison analysis becomes increasingly meaningful as AI-assisted investments mature. The Avoided Loss Attribution metric can begin to be calculated as deals that were declined or restructured based on AI-identified risks reach the point where the counterfactual can be assessed. Report DVQF results quarterly to the management committee, with annual deep dives that integrate all four dimensions.

10. Common Pitfalls and How to Avoid Them

Pitfall 1: Measuring What Is Easy Instead of What Matters

Velocity metrics are easy to capture and produce impressive numbers quickly. Quality metrics require more effort to construct and may initially produce ambiguous or unflattering results. Firms that default to velocity metrics because they are easier will systematically overestimate AI's value. The DVQF addresses this by requiring that velocity metrics always be accompanied by at least one quality metric.

Pitfall 2: Claiming AI Credit for Market-Driven Outcomes

A fund that deploys AI during a favorable market environment may attribute strong investment performance to AI when the primary driver was market beta. Research on PE fund performance has demonstrated the importance of controlling for market conditions and vintage effects when evaluating manager skill (Kaplan & Schoar, 2005). The DVQF's vintage comparison methodology controls for market conditions by benchmarking AI-assisted vintages against appropriate market indices and peer fund performance, not just against the firm's own historical returns.

Pitfall 3: Ignoring Error Costs

AI systems introduce errors that are qualitatively different from human errors. They may be more systematic (affecting all outputs generated by the same model), more subtle (producing plausible-sounding but incorrect analysis), and more difficult to detect (because reviewers develop automation complacency). The DVQF requires explicit error cost estimation as a component of the return calculation. Firms that omit this component will overstate AI ROI.

Pitfall 4: Conflating Adoption with Value

High adoption rates—measured as the percentage of deal teams using AI tools or the percentage of deals that incorporate AI-generated analysis—are frequently reported as evidence of AI success. Adoption is a prerequisite for value creation, not evidence of it. A firm in which 100% of deal teams use AI but none of them produce better investment outcomes has achieved full adoption with zero value creation. The DVQF treats adoption as an input metric, not an output metric.

Pitfall 5: Optimizing AI for the Wrong Decision

The most valuable decisions in PE are not the most frequent ones. Deal screening happens constantly; portfolio construction decisions happen rarely. As decision science research has long emphasized, the quality of infrequent, high-stakes decisions has a disproportionate impact on outcomes compared to the quality of frequent, routine decisions (Hammond, Keeney & Raiffa, 1998). AI systems optimized for high-frequency, lower-stakes decisions (screening speed) may not be optimized for low-frequency, high-stakes decisions (portfolio construction, add-on acquisition timing, exit timing). The DVQF's four-dimensional structure ensures that firms evaluate AI's contribution at each decision point in the investment lifecycle, not just at the highest-volume decision point.

11. Conclusion and Recommendations

The private equity industry's AI measurement infrastructure is not keeping pace with its AI deployment ambitions. Firms are investing heavily in AI-powered deal sourcing, due diligence, and portfolio monitoring, but evaluating those investments almost exclusively through velocity metrics that capture operational efficiency without measuring investment value. This creates a dangerous feedback loop: speed gains are reported as success, throughput targets increase accordingly, and the question of whether faster processes produce better investment outcomes goes unasked.

The Decision Velocity–Quality Framework provides a structured approach to breaking this cycle. By measuring AI's impact across four dimensions—throughput efficiency, analytical depth, outcome attribution, and risk-adjusted return contribution—the DVQF enables PE firms to evaluate AI investments on the terms that actually matter: their contribution to fund performance.

The recommendations for PE firms are as follows:

1. Establish baselines before deployment. The most common and most costly measurement failure is deploying AI without documenting pre-deployment performance. Without baselines, ROI calculation is impossible.

2. Measure quality alongside velocity. Every velocity metric should be accompanied by at least one quality metric. A deal team that reports screening throughput without selection quality data has not measured AI's value.
3. Budget for error costs. AI systems introduce errors that must be measured and costed. Omitting error costs from ROI calculations systematically overstates AI's contribution.
4. Use the Velocity–Quality Matrix as a diagnostic tool. Determine which quadrant your AI deployment occupies, and design interventions to move toward Quadrant A (high velocity, high quality).
5. Be patient with outcome attribution. The most meaningful measures of AI's investment value require years to mature. Build the measurement infrastructure now so that the data is available when the post-deployment track record is long enough to analyze.
6. Report to the management committee. AI measurement that lives exclusively in the technology function will be treated as a technology metric. AI measurement that is reported at the management committee level, alongside investment performance data, will be treated as a strategic priority.
7. Resist the velocity trap. The competitive pressure to screen more deals, faster, is real. But PE returns are generated by finding and executing the best deals, not the most deals. AI measurement should reinforce this distinction, not obscure it.

> *The firms that will generate the highest returns from AI are not those that move fastest. They are those that move fastest without sacrificing the analytical depth and decision quality that drive investment performance. Measuring the right things is the first step toward achieving this balance.*

References

Bain & Company and StepStone Group. (2026). 2026 Private Equity GP Outlook.

Coney, L. (2026). Agentic AI governance in private equity: A behavioral framework for autonomous decision systems. WorkWise Solutions White Paper Series.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32(1), 53–88.

Hammond, J. S., Keeney, R. L., & Raiffa, H. (1998). The hidden traps in decision making. Harvard Business Review, 76(5), 47–58.

IBM Institute for Business Value. (2025a). IBM CEO Study: CEOs Double Down on AI While Navigating Enterprise Hurdles. In cooperation with Oxford Economics.

IBM Institute for Business Value. (2025b). From AI Projects to Profits. IBM Corporation.

Kahneman, D. (2011). Thinking, Fast and Slow. Farrar, Straus and Giroux.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). Noise: A Flaw in Human Judgment. Little, Brown Spark.

Kaplan, S. N., & Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. The Journal of Finance, 60(4), 1791–1823.

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

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.

Simon, H. A. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 69(1), 99–118.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131.

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 "The Skill Erosion Paradox: Preserving Analytical Capability in AI-Augmented Teams" and "Agentic AI Governance in Private Equity: A Behavioral Framework for Autonomous Decision Systems" (both 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