Measuring AI ROI Beyond Cose Savings: A Practical Framework for Enterprise Decision Makers

Mostafa Hanafi
Independent AI & Digital Transformation Consultant
Dubai, UAE
mostafa@vimlyconsulting.com

https://doi.org/10.63962/SAEC6097

Abstract- Organizations are investing heavily in Artificial Intelligence (AI), yet many struggle to articulate and measure its true return on investment (ROI). Traditional ROI approaches focus narrowly on cost reduction or automation savings, often failing to capture the broader strategic and operational value AI can deliver. This paper proposes a practical, five-dimension framework for measuring AI ROI in enterprise settings, grounded in delivery experience and illustrated through a real-world application at a home furnishing company in the Middle East. The initiative evaluated was a custom-built AI-powered forecasting and planning software based on the Claude AI engine, covering demand sensing, production scheduling, and inventory allocation. The framework spans operational efficiency, decision quality, speed to action, risk reduction, and organizational capability uplift, and uses a weighted scoring model with causal decomposition logic to support aggregate investment decisions. Results show how this broader lens changes evaluation outcomes and can prevent premature cancellation of initiatives that deliver strategic but non-obvious value.

Keywords: AI ROI; value measurement; decision quality; risk reduction; enterprise AI management; generative AI

I. INTRODUCTION

AI has moved from experimentation to a board-level priority across industries, including retail and manufacturing. Yet many organizations still struggle with a basic question: what value did we actually get for the money we spent? In practice, ROI discussions often collapse into a single storyline—automation and cost savings—because it is familiar and easy to defend. That storyline is frequently incomplete for AI, especially when AI is deployed to augment decisions rather than fully automate work (Raisch & Krakowski, 2021). The problem is particularly acute in planning-heavy business environments—demand planning, inventory allocation, production scheduling— where decision latency and forecast uncertainty are major cost drivers. In these contexts, an AI system that halves planning cycle time or meaningfully improves forecast consistency can deliver more bottom-line impact than one that reduces headcount, yet the former rarely shows up in a traditional ROI model.

This paper makes the case that AI value is multi-dimensional and proposes a practical framework for capturing it. The framework covers five dimensions: operational efficiency, decision quality, speed to action, risk reduction, and organizational capability uplift. It is designed for enterprise leaders who need to prioritize AI investments,set realistic success metrics, and communicate value through governance and executive forums—without requiring a background in data science or econometrics.

The framework is illustrated through an application at a home furnishing company in the Middle East that deployed a custom-built AI-powered forecasting and planning software based on the Claude AI engine. The example shows how reframing the evaluation using five dimensions changed the investment decision and prevented an early cancellation.

II. BACKGROUND AND RELATED WORK

Three elements of work commonly appear in discussions on AI value. First, productivity research emphasizes that value from general-purpose technologies often arrives with a delay and requires complementary investments such as process redesign and change management (Brynjolfsson, Rock, & Syverson, 2021). Second, practitioner research estimates large productivity potential from generative AI while emphasizing that realizing value depends on integration and organizational change (McKinsey Global Institute, 2023). Third, risk and trustworthiness frameworks stress that AI outcomes depend on socio-technical context and that risk management is part of value realization (NIST, 2023; OECD, 2024). Despite this guidance, many organizations report that ROI remains elusive even as investment rises—suggesting a measurement gap and, sometimes, a prioritization problem (Deloitte, 2025). At the macro level, recent OECD work also highlights growing interest in measuring AI investments and

improving transparency, reinforcing the need for consistent measurement approaches (OECD, 2025).

This paper integrates these perspectives into a single, manager-friendly measurement approach that supports both investment decisions and post-implementation reviews.

III. METHODOLOGY

The framework was developed through a three-phase design-oriented research process. In the first phase, recurring measurement gaps were identified across enterprise AI initiatives, particularly cases where AI decision-support tools were being evaluated using automation ROI templates. Common failure patterns included: evaluating planning AI on headcount reduction alone, treating non-financial dimensions as anecdotal evidence rather than measurable value, and applying ROI frameworks retrospectively after implementation rather than designing measurement into the business case upfront.

In the second phase, five ROI dimensions were derived by mapping the types of value that planning-oriented AI systems typically create. Each dimension was anchored to observable outcome types, practical metrics already available in most enterprise environments, and theoretical grounding in the literature on technology value, decision science, and risk management.

In the third phase, the framework was applied and refined through a real-world initiative. The application was conducted at a home furnishing company operating in the Middle East, which had deployed a custom-built AI-powered forecasting and planning software based on the Claude AI engine. The system covers three core planning functions: demand sensing, production scheduling optimization, and inventory allocation across product categories and sales channels. Rather than replacing planners, the software provides AI-generated recommendations and scenario comparisons that planning teams use to inform and accelerate their decisions. All operational details have been generalized to protect confidentiality while preserving the measurement logic.

The framework is best suited to AI systems that augment human decision-making in complex operational environments. It is less directly applicable to fully automated back-office processes, where traditional cost-reduction ROI models remain appropriate. Organizations should determine the primary role of their AI system before selecting which dimensions to emphasize.

IV. A MULTI-DIMENSIONAL FRAMEWORK FOR AI ROI

The framework defines AI ROI across five complementary dimensions. The central idea is to measure AI based on the type of value it is designed to produce, rather than forcing every initiative into an automation-savings mold. Each dimension captures a distinct mechanism through which AI creates enterprise value, and together they provide a complete picture of investment impact. Fig. 1 shows an illustrative distribution of value across dimensions for the home furnishing company application discussed in Section V.

A. Operational Efficiency

Efficiency remains a valid and important dimension, but in AI-augmented planning environments it rarely manifests as role elimination. Instead, efficiency gains typically show up as reduced rework caused by late-stage forecast revisions, fewer manual handoffs between commercial, supply, and production teams, and less time spent reconciling conflicting data across systems. In the context of the Claude AI-powered planning software, efficiency is visible in how much faster planners can prepare a consensus plan and how often that plan needs to be revised after approval. These gains are real and measurable, but they reflect capacity freed up for higher-value work rather than workforce reduction.

B. Decision Quality

AI can improve decision quality by increasing consistency, revealing non-obvious patterns, and enabling scenario evaluation under uncertainty (Raisch & Krakowski, 2021). AI systems like the Claude-based planning software can improve decision quality by processing more variables than a planner can hold in mind simultaneously, surfacing non-obvious patterns in demand signals, and enabling rapid scenario comparison under uncertainty. Practical measurement proxies include: forecast error rates (e.g., Mean Absolute Percentage Error, or MAPE), exception rates, decision reversal rates (the proportion of approved plans subsequently revised before execution), and post-decision outcome tracking such as fill-rate impact in the weeks following a planning decision.

C. Speed to Action

In volatile environments, faster decision cycles can be more valuable than marginal cost savings. Generative AI and analytics can compress analysis and reporting time, allowing teams to respond quickly to disruptions (McKinsey Global Institute, 2023). For the home furnishing company, where

cross-border logistics and seasonal demand swings add significant complexity, reduced decision latency was consistently cited by management as one of the most tangible operational advantages of the Claude AI-based system. This dimension is measured through planning cycle time, time-to-decision, and time-to-mitigation following a disruption event.

D. Risk Reduction

Risk reduction should be treated as a ROI contributor, not a compliance afterthought. Trustworthy AI guidance emphasizes managing risks across the AI lifecycle and aligning controls with context (NIST, 2023). In operational settings, an AI-powered planning system reduces risk by identifying stock imbalances and supply bottlenecks earlier, before they become service failures. It also reduces the variance of planning outcomes, making performance more predictable across demand cycles. Risk reduction is measured through avoided incidents (such as stockout events for critical product categories), changes in demand or supply variance, and improvements in policy adherence. Where historical disruption cost data is available, avoided incidents can be partially translated into financial terms.

E. Organizational Capability Uplift

AI investments often create long-term value by building capabilities: better data literacy, improved cross-functional alignment, and a repeatable delivery and governance playbook. Because these benefits are intangible, organizations tend to ignore them—even though they strongly influence scaling success (Brynjolfsson et al., 2021; Deloitte, 2025). Yet because capability uplift is hard to assign a dollar value, organizations routinely omit it from ROI evaluations, even when it is the dimension most predictive of long-term AI scaling success. It is measured through AI adoption breadth and depth, time-to-launch for new use cases, and reduction in external vendor dependency for analytics and planning support.

Table I. AI ROI Dimensions and Practical Metrics (Examples)

ROI Dimension	What It Captures	Example Practical Metrics
Operational Efficiency	Reduction in planning friction and rework rather than role elimination	Planner hours per cycle; % rework; manual handoffs; data reconciliation time
Decision Quality	Improved consistency and effectiveness of human decision-making	Forecast accuracy bands; exception rate; decision reversal rate; post-decision service-level impact
Speed to Action	Faster response from signal detection to operational action	Planning cycle time; time-to-decision; time-to-mitigation; exception triage time
Risk Reduction	Lower exposure to operational and supply disruptions	Stockout incidents; disruption frequency; variance reduction; compliance exceptions
Organizational Capability Uplift	Long-term readiness to scale and sustain AI adoption	AI adoption rate; number of teams onboard; time to launch new use cases; reduced vendor dependency

V. ROI Aggregation Logic

Combining five dimensions into a single investment decision score introduces two risks that the framework is specifically designed to prevent. The first is double-counting, where the same underlying value improvement is credited to more than one dimension. The second is dimension dominance, where easily quantifiable dimensions—particularly operational efficiency—crowd out harder-to-measure dimensions regardless of their actual strategic contribution.

A. **Causal Decomposition**

The framework uses causal decomposition to assign each observable outcome to its primary dimension—the one where the value is most directly caused. For example, a reduction in stockout incidents may reflect both earlier risk detection and faster response time. The framework requires analysts to

document the causal path before assignment: if the improvement is primarily caused by earlier signal detection, it is assigned to Risk Reduction; if it is primarily caused by shorter decision cycles, it is assigned to Speed to Action. Where outcomes genuinely cannot be cleanly separated, a proportional contribution split is applied before summing, adapted from causal decomposition methods used in marketing attribution and operations value modeling.

B. **Weighted Scoring Model**

Once each dimension is measured and double counting is controlled, dimensions are aggregated using a weighted scoring model. Weights reflect organizational priorities and must be set before evaluation begins, not after results are known, to prevent post-hoc justification bias. The composite ROI score S is defined as:

$$S = w_1*E + w_2*D + w_3*R + w_4*K + w_5*C$$

Where E = Operational Efficiency, D = Decision Quality, R = Speed to Action (Responsiveness), K = Risk Reduction, and C = Capability Uplift. Weights w_1 through w_5 sum to 1.0. Each dimension score is a normalized 0„100 index derived from percentage improvement over baseline, averaged across the metrics selected for that dimension. A balanced default profile assigns equal weight of 20% to each dimension. Organizations should adjust these weights to reflect their specific strategic context—for instance, assigning higher weight to Risk Reduction in safety-sensitive environments, or to Capability Uplift when the primary objective is building the foundation for future AI scaling.

A. **Causal Attribution**

Isolating the causal contribution of an AI system from concurrent organizational changes and market trends is a real methodological challenge that the framework addresses directly. Where feasible, randomized or quasi-experimental designs are preferred: A/B testing or switchback experiments assign comparable planning units to AI-assisted and non-assisted conditions and compare outcomes. When full randomization is not possible, difference-in-differences analysis compares trends in AI-exposed units against comparable non-exposed units before and after implementation. In most enterprise deployments, a pre-post comparison with documented controls is the practical starting point. The framework requires the attribution method to be disclosed, and directional improvements from pre-post designs should be presented as “consistent with AI contribution” rather than as proven causal effects.

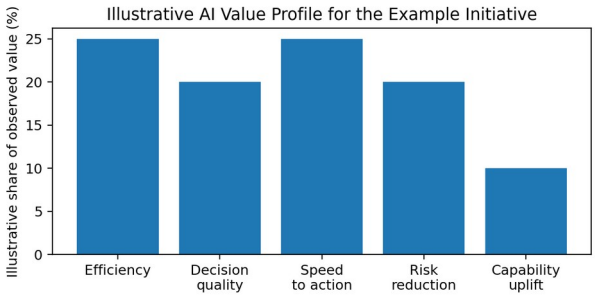
VI. EXAMPLE APPLICATION AND DISCUSSION

The framework was applied to an AI planning initiative at a home furnishing company operating in the Middle East. The company’s supply chain spans sourcing, warehousing, and retail distribution across product categories ranging from upholstery to decorative accessories. Before the AI initiative, planning processes were spreadsheet-based and heavily manual, with weekly consensus meetings across commercial, supply chain, and production teams involving roughly 12 planners and coordinators. Late-stage forecast revisions were common, planning cycles were slow, and service-level consistency suffered during peak seasonal demand periods.

The company deployed a custom-built AI-powered forecasting and planning software based on the Claude AI engine. The software integrates demand sensing, production scheduling optimization, and inventory allocation into a single planning environment. Planners interact with the system through natural language interfaces and structured scenario tools, receiving AI-generated recommendations that they can accept, modify, or override. The implementation covered a 9-month deployment period followed by a 3-month stabilization phase before formal measurement began.

The original business case leaned heavily on reducing manual planning effort, and early executive reviews questioned ROI when headcount did not change. Using the multi-dimensional framework, the evaluation was reframed. Instead of asking whether the AI system had replaced planners, stakeholders assessed how it had influenced the quality of planning decisions, the organization’s responsiveness to market volatility, and its exposure to operational risk. A 6-month pre-implementation baseline was established using historical operational data; post-implementation measurements were taken at the 6-month mark following stabilization.

Fig. 1. Illustrative distribution of value across ROI dimensions (example initiative).



Note: The chart is illustrative and intended to show how value can distribute across multiple dimensions even when direct cost savings are limited.

Dimension	**Metric**	**Baseline Value**	**Post-implementation value**	**Change**
Efficiency	Planner hours per planning cycle	~48 hrs	~31 hrs	−35%
Efficiency	Late-stage forecast rework rate	~22%	~11%	−50%
Decision Quality	Forecast MAPE (4-week horizon)	~18%	~14%	−4pp
Decision Quality	Decision reversal rate	~28%	~17%	−11pp
Speed to Action	Planning cycle duration (signal to approved plan)	~4.5 days	~1.8 days	−60%
Speed to Action	Time-to-mitigation after disruption	~3.2 days	~1.4 days	−56%
Risk Reduction	Stockout incidents per quarter (critical SKUs)	~34	~21	−38%
Risk Reduction	Demand variance (coefficient of variation)	0.41	0.33	−20%
Capability Uplift	Cross-functional teams using shared planning data	2	4	2×
Capability Uplift	Time to design and launch a new AI use case	~5 months	~2 months	−60%

Applying the balanced weight profile (20% per dimension), the composite ROI score was 61 out of 100, supporting a recommendation to continue and expand the initiative. Notably, Speed to Action and Capability Uplift were the two highest-contributing dimensions. Neither would have appeared in a traditional automation-focused evaluation.

The most significant finding for decision-making was contextual: the initiative had nearly been cancelled in its third month because headcount reduction had not materialized as originally projected in the business case. Applying the multi-dimensional framework reframed the conversation entirely. Leadership could see that the Claude AI-powered planning software was delivering meaningful value through planning cycle compression, improved forecast consistency, and organizational readiness to scale—none of which were visible in the original ROI model. The initiative was not only retained but subsequently expanded to additional product categories.

Two broader observations emerged. First, the type of AI system matters for measurement: because the Claude-based software is designed to augment planner judgment rather than replace it, the most meaningful value shows up in decision quality and responsiveness, not in labor cost reduction. Forcing an augmentation system into an automation ROI template guarantees an undercount of value. Second, capability uplift proved to be a leading indicator of long-term success. Teams that adopted the system more deeply also launched follow-on use cases faster and reduced dependence

on external analytics vendors—a compounding return that is entirely invisible in single-period ROI calculations.

Limitations of this evaluation should be acknowledged. This is a single-site application and generalizability across other industries or organizational contexts has not been tested. Several dimension measurements rely on practitioner-reported proxies rather than independently verified data. No control group was used; the improvements reported are directional indicators consistent with the AI initiative's contribution rather than causally isolated treatment effects.

VI. CONCLUSION AND FUTURE WORK

This paper proposed a practical framework for measuring AI ROI beyond cost savings. By combining operational efficiency, decision quality, speed to action, risk reduction, and organizational capability uplift, the framework provides a more realistic basis for evaluating AI investments in complex enterprise settings
The application to a custom-built AI-powered forecasting and planning software based on the Claude AI engine at a home furnishing company in the Middle East demonstrates the framework's practical utility. A composite score of 61 out of 100 supported continuation and expansion of an initiative that a traditional automation-only evaluation would have recommended cancelling. Speed to Action and Capability Uplift were the most impactful dimensions—a result that only becomes visible when the measurement framework is aligned with the actual role the AI system plays.
For leaders, the practical takeaway is to (1) match ROI metrics to the role AI is expected to play, (2) treat risk reduction and decision responsiveness as value drivers, and (3) explicitly account for complementary investments—data work, change management, and governance—that enable value to materialize over time (Brynjolfsson et al., 2021; NIST, 2023).

Future work should validate the framework across additional industries and deployment contexts, develop standardized baselines for each dimension, and explore more rigorous quasi-experimental attribution designs. There is also an opportunity to formalize the framework's integration with enterprise AI governance programs, particularly as organizations build repeatable processes for evaluating AI portfolio performance.

REFERENCES

[1] Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics, 13(1), 333–372.

[2] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. Harvard Business Review, 96(1), 108–116.

[3] Deloitte. (2025, October 22). AI ROI: The paradox of rising investment and elusive returns. Deloitte Insights.

[4] McKinsey Global Institute. (2023, June 14). The economic potential of generative AI: The next productivity frontier. McKinsey & Company.

[5] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1).

[6] OECD. (2024, May). OECD AI Principles (updated May 2024). OECD.AI Policy Observatory.

[7] OECD. (2025). Advancing the measurement of investments in artificial intelligence. OECD Publishing.

[8] S. Raisch and S. Krakowski, "Artificial intelligence and management: The automation-augmentation paradox," Academy of Management Review, vol. 46, no. 1, pp. 192–210, 2021.

[9] E. Brynjolfsson, D. Rock, and C. Syverson, "The productivity J-curve: How intangibles complement general purpose technologies," American Economic Journal: Macroeconomics, vol. 13, no. 1, pp. 333–372, 2021.