



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

AI Operational Analytics: A Conceptual Measurement Framework for Enterprise AI Adoption and Value Realization

Anwesha Sharma

SAP, USA. ORCID ID: <https://orcid.org/0009-0009-0825-9213>

Abstract

Artificial intelligence has become a strategic resource across product development, customer service, workforce management, financial planning, and operational governance in most modern organizations. Proof that this investment pays off, though, is still hard to come by, even as companies push past exploratory pilots into system-wide deployment. McKinsey's latest worldwide survey puts a number on that gap: 88% of companies report using AI in at least one functional area, but only 39% can point to any measurable impact on operational profit, and a mere 6% qualify as high performers. This paper takes up the problem of devising measures that close the distance between widespread adoption and demonstrated returns.

Key words

Artificial Intelligence, Enterprise Operational Intelligence, AI Operational Analytics, AI Adoption, Value Realization, Decision Analytics, AI Governance

I. Introduction

Modern enterprises now treat artificial intelligence as a strategic capability touching product development, customer service, workforce management, financial planning, and operational governance. Executives shifting from experimental AI projects to enterprise-wide deployment want proof the investment pays off in measurable operational and business terms, and that proof has been unusually hard to produce. McKinsey's most recent global survey found 88 percent of organizations reporting regular AI use in at least one business function, yet only 39 percent could attribute any enterprise-level EBIT impact to it, and a mere 6 percent qualified as "high performers" with significant, broad-based value (Singla et al., 2025). This paper takes up the problem of closing that gap between widespread adoption and demonstrated value.

Three largely disconnected layers currently make up how AI's contribution gets measured. Technical metrics like model accuracy, inference latency, and response quality describe system performance but little else. Adoption metrics such as active-user counts, prompt volume, or licensed-seat utilization tell you whether AI is being used, but not whether that use is improving decisions or outcomes. A third layer, a growing body of work on responsible AI governance, focuses on accountability, explainability, and compliance (Xia et al., 2023; National Institute of Standards and Technology [NIST], 2023). That layer matters. It just cannot demonstrate business value on its own any more than the other two can. No existing framework known to the authors ties all three together into one coherent measurement approach running from initial adoption through to realized business value.

This fragmentation is not just an academic problem. Karountzos's (2025) systematic review of AI-adoption literature across FinTech, retail, and advanced manufacturing found a broadly positive relationship between AI adoption and reported productivity, but sector-specific barriers got in the way of any single, universal measurement approach. Retail AI systems struggled with trust deficits, manufacturing with skill gaps, and FinTech with regulatory uncertainty. A similar conclusion comes out of Valiulla's (2025) case-study analysis of DORA and SPACE metric-framework adoption at organizations including Meta, Etsy, Spotify, and Mercado Libre: measurement-framework success hinges more on organizational context, culture, size, and engineering maturity than on the technical design of the framework itself. That finding bears directly on any attempt to standardize AI-specific measurement across a heterogeneous set of enterprises.

This paper proposes AI Operational Analytics (AIOA), a structured measurement framework for evaluating enterprise AI across its full value lifecycle: how it is adopted, how it is used, how it is integrated into operational workflows, how it shapes decision-making, and how it ultimately affects business outcomes. The contribution

here is explicitly conceptual, a structured set of dimensions and representative indicators meant to guide future empirical measurement, not the results of an empirical study already run. Section II lays out AIOA's five dimensions and a proposed measurement model. Section III walks through an illustrative cross-domain application and discusses the framework's governance implications. Section IV covers limitations and a future research agenda, and Section V concludes.

II. Materials and Methods The AI Operational Analytics Framework

AIOA was built through conceptual framework synthesis. It draws together three literatures that Section I established have not previously been connected with one another: enterprise AI adoption and productivity research (Karountzos, 2025), software/AI measurement-framework adoption research (Valiulla, 2025), and AI governance and accountability research (NIST, 2023; Xia et al., 2023). The method itself is conceptual design rather than empirical study. No primary data collection, controlled comparison, or statistical validation went into building the framework, and the illustrative application in Section III and the indicators below function as tools for future empirical validation rather than a stand-in for it.

Five interconnected dimensions make up AIOA, described below. Together they trace how an AI initiative moves from first being adopted through to delivering measurable business value.

AI adoption, as used here, means the extent to which enterprises have worked AI applications into their regular day-to-day operations. Companies often lean on simple indicators like the number of models deployed or the number of licensed users. That leaves out the question that actually matters, though: are people using AI in their daily decision-making, or does it sit mostly dormant after rollout? More telling questions are how many staff actually open the tools, how many parts of the company have taken them up, how deep that reach goes, how often people return to them, and whether that use persists months later. McKinsey's 2025 study captures the scale of this gap well. 88% percent of organizations report using AI somewhere, but far fewer are using it deeply enough within their workflows for that use to register as anything more than surface-level.

Once AI is adopted, a separate question follows: how well is it actually being used? Two organizations with identical adoption rates can land in very different places depending on how deeply AI gets woven into operational workflows. Useful measures here include how much of the workflow runs automatically, how often staff draw on AI when making a decision, how often they accept its recommendation, how often they override it, and how many functions use it at all. Valiulla's (2025) finding that measurement framework success ties more closely to organizational context than to framework design is directly relevant. Any reading of utilization indicators needs to account for an organization's own engineering and operational maturity, since comparing raw numbers across dissimilar organizations will mislead more than it informs.

Adoption and Utilization describe user behavior; Operational Impact breaks from both to ask whether AI is producing tangible improvements in enterprise operations. Relevant signs include process cycles finishing faster, plans tracking closer to actual outcomes, problems being caught and resolved sooner, delivery dates holding more reliably, and a greater share of the process running without manual intervention. This dimension also overlaps most with the existing operational analytics literature. Karountzos's (2025) review found AI adoption correlating with productivity gains, particularly through automation, predictive analytics, and customer personalization, though the review's own method, a qualitative synthesis of over 150 academic papers and 30 corporate filings, stopped short of a single standardized effect-size estimate that would hold across sectors. That gap is precisely why this dimension needs a structured set of indicators rather than one universal metric.

The topic of how artificial intelligence influences business decision-making revolves around Decision Quality. The crucial aspect is not if AI substitutes for the decision-making process conducted by humans, but it positively speeds up the decision process. The proportion of accepted interventions from AI, accuracy in forecasting processes, confidence in decision-making, and clarity of recommendations can serve as good indicators of that. According to the paper by Dogru and Krämer through structural equation modeling (2025, n=529) it has been proved that it is necessary to have the user competence, trust calibration, and self-confidence in the AI system for it to be used properly, while every behavioral characteristic on its own is not sufficient. Therefore, decision making needs to be evaluated not only as a stand-alone phenomenon. This dimension is also part of the AI governance issue at the same time. According to NIST (2023), the AI Risk Management Framework states that it is essential to monitor the effectiveness of AI systems during the operation, while the accountability measures by Xia et al. point out that auditability and explainability can serve as two requirements concerning the operation of governance measures.

Business Value is at the highest level of the AIOA since the objective is to determine whether the AI initiative is able to fulfill the overall strategy of the organization and not just improve a specific process. The metrics helping to measure it are AI-driven revenue growth, cost reduction, increased productivity, customer

satisfaction, speed of innovation, and returns from AI investments. According to the 2025 survey by McKinsey, the thing that distinguishes AI "high achievers" from the rest is the way they use organizational practice rather than which technology they use. Their workflows were reorganized around AI, there was support of top management for using AI, and they had efficient processes for monitoring AI outputs' performance (Singla et al., 2025). Consequently, business value metrics should be considered along with these organizational practice signals and not just considered as technology-related indicators.

Beyond the five dimensions, AIOA recommends assessing each one through a balanced mix of quantitative and qualitative indicators tied to organizational objectives; no single metric is sufficient. This reflects Valiulla's (2025) finding that measurement-framework success is context-dependent, and it reflects the AI-governance literature's insistence on measuring the process and oversight structures behind outcomes, not just the outcomes themselves (NIST, 2023). In practice, an organization applying AIOA should expect to track a small handful of indicators per dimension rather than a single composite score, with the specific indicators shifting somewhat by business function and industry, the same sector-specific pattern Karountzos (2025) documented across FinTech, retail, and manufacturing.

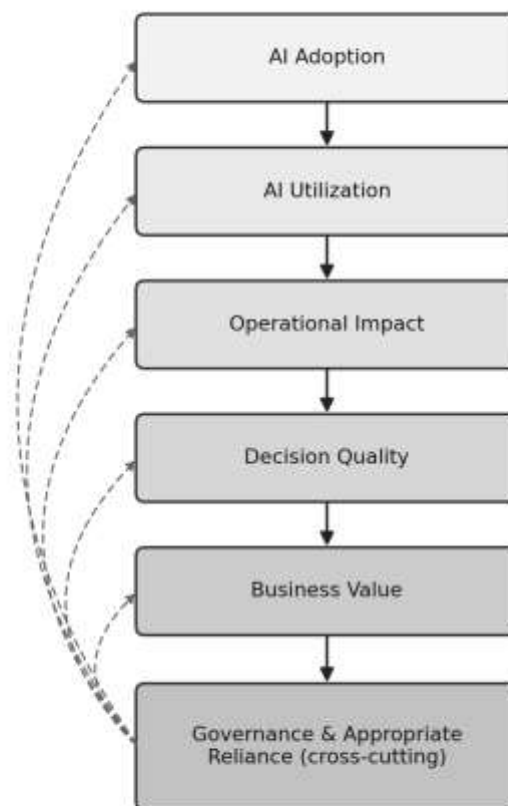


Figure 1. AI Operational Analytics (AIOA) Value Lifecycle

Governance and appropriate-reliance considerations apply across all five dimensions

Figure 1. AI Operational Analytics (AIOA) Value Lifecycle

III. Results and Discussion

This section walks through how AIOA's five dimensions would apply in practice, across a representative set of enterprise domains. As Section II made clear, this is illustrative, not a completed empirical case study; no organization-specific data was collected. The walkthrough below describes how the dimensions would apply to representative scenarios; none of it reports measured outcomes from an actual deployment.

Take Enterprise Software Delivery: AIOA would track AI-assisted planning accuracy (Adoption/Utilization), the acceptance rate of AI-assisted release-governance decisions (Decision Quality), and the resulting improvement in delivery predictability (Operational Impact) in an application consistent with the AI-augmented delivery-pipeline evidence already established in the software-engineering measurement literature (Valiulla, 2025). In customer experience, the relevant measures are AI-driven issue-resolution

adoption and utilization; the quality and acceptance of AI-generated resolution recommendations (decision quality); and the resulting shift in customer satisfaction and service efficiency (business value). Workforce management looks different again: AI support for capacity planning and workforce allocation (operational impact) and whether managers actually trust and act on those recommendations (decision quality), a framing drawn directly from Dogru and Krämer's (2025) finding that appropriate reliance turns on user expertise and trust calibration, not automatic acceptance. Financial services shifts the emphasis toward risk management, fraud detection, and compliance, with particular weight on auditability and explainability given the sector's regulatory scrutiny (NIST, 2023; Xia et al., 2023). And in healthcare, AIOA would track AI-enabled operational efficiency and resource allocation alongside clinical decision-support acceptance and explainability—the same governance considerations, but at higher stakes.

Table 1 summarizes this illustrative mapping between AIOA's five dimensions, representative indicators, and the enterprise domains discussed above. The indicators listed are illustrative examples of the kind of measurement each dimension would capture, not measured values from an actual deployment.

Table 1. Illustrative Mapping of Enterprise Domains to AIOA Dimensions and Indicators

Enterprise Domain	Primary Dimensions Emphasized	Representative Indicator (Illustrative)
Enterprise Software Delivery	Adoption/Utilization; Decision Quality; Operational Impact	AI-assisted release-decision acceptance rate; delivery-predictability change
Customer Experience	Adoption/Utilization; Decision Quality; Business Value	AI-generated resolution-recommendation acceptance; satisfaction change
Workforce Management	Operational Impact; Decision Quality	Manager acceptance of AI-assisted allocation recommendations
Financial Services	Decision Quality (governance-weighted); Business Value	Explainability/auditability of fraud-detection recommendations
Healthcare	Decision Quality (governance-weighted); Operational Impact	Clinical decision-support acceptance and explainability

AIOA's central implication is simple: adoption alone is the wrong thing to track. McKinsey's (2025) finding that 88 percent of organizations report AI use while only 39 percent report any enterprise-level EBIT impact is exactly the gap AIOA is built to close. Usage on its own is not the same as value, and an organization that only tracks adoption risks confusing the two. AIOA's five-dimension structure builds that distinction directly into the framework, separating Adoption and Utilization (behavioral indicators) from Operational Impact, Decision Quality, and Business Value (outcome indicators), so high usage never gets treated as self-evidently valuable. Governance is where the second implication lands. Rather than treating governance as a compliance exercise bolted on afterward, AIOA folds it directly into decision quality measurement. Explainability, auditability, and human-oversight indicators (NIST, 2023; Xia et al., 2023) become measurement dimensions in their own right. This echoes a broader argument in the responsible AI literature, namely that accountability functions as a governance mechanism regardless of an AI application's assessed risk level, and belongs in continuous measurement rather than a one-time check at deployment.

Valiulla (2025) also emphasizes the importance of organizational context in terms of AIOA adoption. It posits that context determines the value of which indicators to consider in each dimension, rather than simply using one single set of indicators. Author validates the previous statements in his research where he studied the effects of various obstacles and drivers to allowing companies to realize value from AI across regions and spheres of activity.

A key aspect of achieving decision ideal is correct reliance on AI, not maximum implementation of it. The empirical evidence from Dogru and Krämer (2025) shows that both reliance on AI leads to poor decision-making. This is not a coincidence, meaning that in order to achieve the right level of reliance on AI, it is crucial to take all three elements into account. Thus, a high acceptance rate of AI recommendations should not be seen

as something positive because it can be a signal of the fact that an organization relies on AI too much, not appropriately utilizes it.

It is worth being clear about where AIOA stands right now. It's an early conceptual proposal, roughly where DORA metrics stood in the DevOps literature before large-scale empirical validation took place. Valiulla (2025) offers a useful reminder here, that even a well-validated framework like DORA still produces mixed outcomes depending on organizational context. The five dimensions and their representative indicators mark a starting point for that same validation process, not a measurement standard already proven.

IV. Limitations and Future Research

Because this paper's contribution is conceptual, its limitations follow directly from that scope. AIOA has not been empirically validated. The illustrative application in Section III shows conceptual plausibility and internal coherence, not measured organizational outcomes, and no primary data collection or statistical testing was performed. The representative indicators proposed for each dimension are illustrative rather than exhaustive, since different organizations and sectors will likely need different specific indicators within the same five-dimension structure, in line with the sector-specific variation Karountzos (2025) documented. The framework also does not resolve the measurement framework adoption challenges that Valiulla (2025) documented, which organizations adopting AIOA should expect to see in the same context-dependent success and failure patterns seen with DORA and SPACE. And AIOA's Decision Quality and Business Value dimensions lean on reasonably mature underlying operational data. Organizations with substantial data fragmentation or immature AI governance practices should expect these two dimensions to be harder to measure reliably than Adoption or Utilization.

These limitations point toward several directions for future research. Empirical validation of AIOA itself is the most immediate priority. Applying the proposed indicators longitudinally, before and after structured AIOA adoption, would show whether the five-dimension structure actually produces more actionable measurement than existing fragmented approaches, and which specific indicators turn out to predict realized business value within each dimension.

Standardization presents a second useful direction. Karountzos (2025) and the broader AI-adoption literature make clear that organizations currently measure AI performance using highly inconsistent metrics, which limits cross-organizational benchmarking. Whether a subset of AIOA's proposed indicators can be standardized across sectors, in the way DORA metrics achieved broad standardization in software delivery performance measurement, while still preserving the sector-specific flexibility Karountzos's findings suggest is necessary, is worth investigating further.

How AIOA's decision-quality dimension interacts with the trust-calibration literature is a third open question. Because Dogru and Krämer (2025) found appropriate reliance depends on the interaction of expertise, trust, and self-confidence, future AIOA validation studies would do well to incorporate individual-level trust-calibration measures alongside organization-level decision quality indicators, rather than treating decision quality as a purely organizational-level construct.

One last question follows from where AI systems are heading. As autonomous, agentic components become more common, does AIOA's five-dimension structure still adequately capture the measurement needs of agentic AI specifically, or does autonomous decision-making and multi-agent coordination call for additional measurement dimensions beyond the five proposed here?

V. Conclusion

AI adoption on its own does not guarantee organizational value. What enterprises need is a structured way to measure how AI shapes operations, decision-making, and business performance across its full lifecycle, not just whether it has been deployed. This paper introduced AI Operational Analytics (AIOA), a conceptual measurement framework spanning five interconnected dimensions: Adoption, Utilization, Operational Impact, Decision Quality, and Business Value. By moving the measurement focus from technology usage to enterprise outcomes, and by building governance and appropriate reliance into measurement rather than treating them as a separate compliance step, AIOA gives organizations a structured starting point for closing the gap between widespread AI adoption and demonstrated business value that recent industry survey data has made clear. Its contribution remains explicitly conceptual: the indicators and illustrative application here guide future empirical validation rather than stand in for it.

Conflicts of Interest

The author declares no conflicts of interest relevant to this manuscript.

Acknowledgment

The author used Claude (Anthropic, claude-sonnet model) to assist with manuscript drafting, literature synthesis support, and language refinement. All conceptual contributions, framework design decisions, and final content judgments are the author's own.

This research received no external funding.

Data Availability: Not applicable this is a conceptual framework paper; no new data were generated. The illustrative application in Section III is a hypothetical walkthrough, not a report of measured organizational data.

References

- [1] Dogru, E. O., and Kramer, N. C. (2025). Investigating appropriate reliance on AI-based decision support systems: The role of expertise, trust, and self-confidence. *Journal of Decision Systems*, 34(1). <https://doi.org/10.1080/12460125.2025.2593251>
- [2] Eken, B., Pallewatta, S., Tran, N. K., Tosun, A., and Babar, M. A. (2024). A multivocal review of MLOps practices, challenges and open issues (arXiv:2406.09737) [Preprint]. arXiv. <https://arxiv.org/abs/2406.09737>
- [3] Karountzos, V. (2025). The impact of AI adoption on productivity across FinTech, Retail, and Advanced Manufacturing: A systematic quantitative literature review approach (Productivity Insights Paper No. 061). The Productivity Institute, Alliance Manchester Business School.
- [4] National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. <https://doi.org/10.6028/NIST.AI.100-1>
- [5] Singla, A., Sukharevsky, A., Hall, B., Yee, L., Chui, M., and Balakrishnan, T. (2025). The state of AI in 2025: Agents, innovation, and transformation. McKinsey & Company, QuantumBlack. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>
- [6] Valiulla, A. (2025). An empirical study of measurement framework adoption DORA and SPACE: How organizational context shapes success and failure (SSRN Working Paper No. 5344643). SSRN. <https://doi.org/10.2139/ssrn.5344643>
- [7] Xia, B., Lu, Q., Zhu, L., Lee, S. U., Liu, Y., and Xing, Z. (2023). Towards a responsible AI metrics catalogue: A collection of metrics for AI accountability (arXiv:2311.13158) [Preprint]. arXiv. <https://arxiv.org/abs/2311.13158>