

# Measure What Matters: A Three-Layer Measurement Framework for AI Projects

**SCALE|AI**

**Studies have shown Canadian companies are more hesitant to adopt AI than companies in peer nations. While some of this can be attributed to financial caution, we've detected a deeper cause: a fear of failure that often becomes paralyzing.**

The fear isn't unfounded. AI projects can fail, but the cause isn't typically due to the technology itself. Rather, it's more often a rush toward implementation without first answering two fundamental questions: *What are we trying to achieve? And how will we know if we've succeeded?*

These two elements - establishing clear objectives, and measurable success criteria - can determine the difference between whether a project delivers transformative value, or ends up on the cutting room floor. And it makes sense: if you don't know where you're going, how will you know when you've arrived?

In this paper we focus on the second question - have we succeeded? - and show why this is not as difficult to answer as it may seem. But to be effective, the answers must be defined before the AI project even begins.

Well-chosen metrics are the key, because they transform AI projects from black boxes into manageable initiatives, and enable organizations to assess progress during

development (not just at the end). They enable continuous progress assessment, informed investment decisions, and, crucially, the ability to confidently decide whether continued investment is justified.

But choosing the right metrics can be challenging, particularly because traditional performance indicators - like revenue growth or cost savings - often fail to capture whether an AI project has truly been effective or not. In particular, most traditional metrics can't translate technical progress into demonstrable business value.

However, there's a general lack of guidance on solid alternatives for meaningfully measuring AI project success. This paper aims to bridge that gap, by presenting a three-layer approach to translate technical progress into demonstrable business value.

# Metrics Matter

Metrics are essential to help an organization's stakeholders establish the shared understanding and common decision-making framework needed for successful AI adoption. Adopting the right metrics *before* the project begins offers many benefits:

01

## More timely decisions

Establishing a measurement approach from the outset allows for earlier feedback on whether the project is going as planned, and can help more accurately pinpoint when and where corrective action is required.

02

## Commit with confidence

Establishing metrics in advance provides a clearer sense of where the finish line is, what is needed to get there, and when goals have been reached, which empowers more confident decisions.

03

## Stay aligned

Innovation projects usually involve multiple technical and business teams, overseen by executives. Establishing metrics in collaboration with all stakeholders provides a more objective assessment of progress, a common language for communication, and a collective understanding of what success looks like.

04

## Eliminate false positives

Adopting robust approaches reduces the risk of mismatch between meeting technical milestones and meeting (or missing) overall business objectives.

05

## Build stakeholder support

Visible, measurable progress against meaningful metrics helps to build and sustain executive confidence. When leaders can tangibly see clear business value demonstrated beyond just technical milestones, they're able to champion the project more effectively within the organization.

## The 'Problem' with Traditional Indicators

Traditional key performance indicators (KPIs), like revenue generation or cost reduction, are often a 'go-to' metric for assessing AI projects. But since these business indicators cannot directly be applied to the technical outputs of an AI project, they often give a misleading or delayed picture of the project's impact - especially in the near-to mid-term.

These shortcomings leave organizations struggling to measure the success of AI projects:

- **Delayed Feedback:** Traditional KPIs typically provide insights only after a significant time lag.
- **Excessively Conservative:** Financial KPIs often miss capturing the full picture of AI's benefits.
- **Overlapping Influences:** External or unrelated factors may influence the KPI, making it difficult to isolate AI project impact.
- **Operational Disconnect:** Traditional reporting may split KPIs among silos, masking a project's true impacts.
- **Subjective Interpretations:** Traditional KPIs can be subjectively interpreted across organizations.



# The Three-Layer Approach: A New Perspective on AI Metrics

Through our work supporting hundreds of AI projects across diverse industries, we've observed a consistent challenge: how to translate AI's technical outputs and milestones into meaningful business-relevant metrics that organizations can actually use?

To address this gap, we've developed a three-layer measurement framework that bridges the divide between what AI systems produce and what business leaders need to evaluate success and plan next steps.

Each layer is detailed later in this paper, but at a high level, three layers are used to connect technical performance to business outcomes:

## 01 Business Outcome Layer

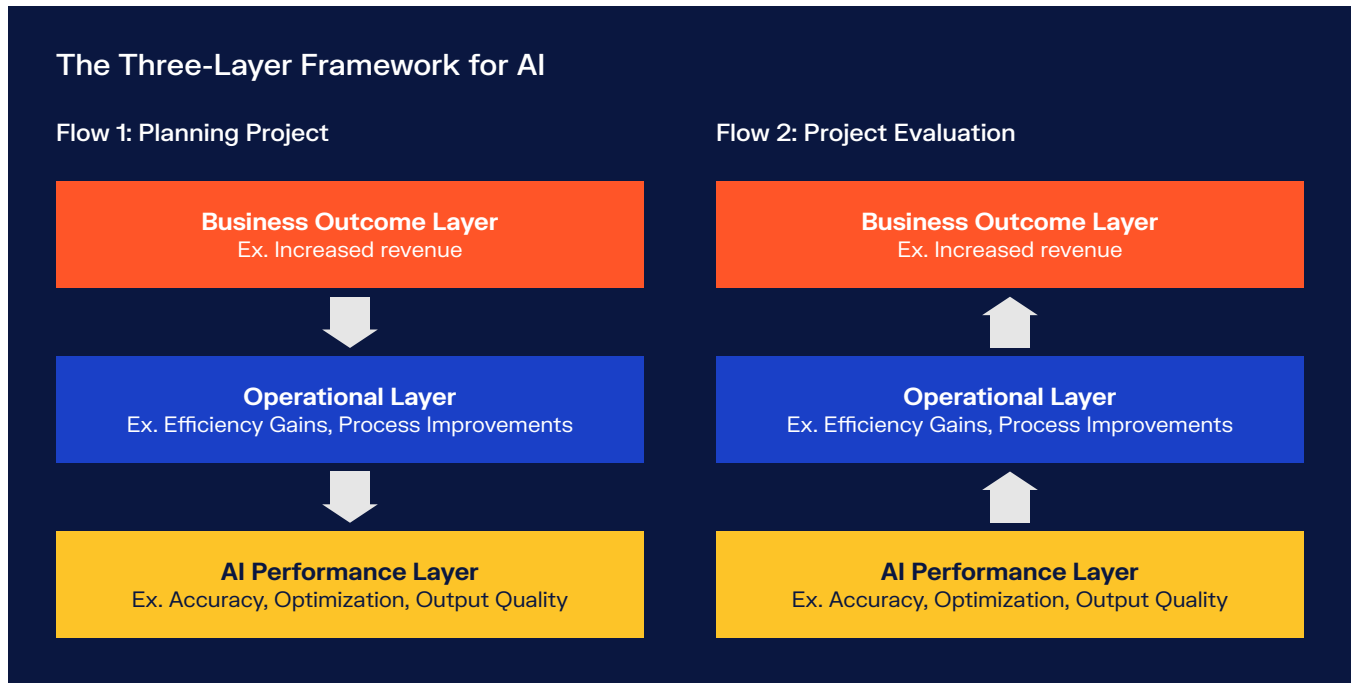
The desired business change, and the reason for doing the project. This is the end goal and ultimate success measure for the project.

## 02 Operational Layer

Achieving the desired business outcome requires productivity gains versus the existing business process. These gains are measured in the operational layer using productivity metrics.

## 03 AI Performance Layer

The AI solution produces predictions, optimizations, recommendations, generated content, etc. that are used to achieve productivity gains. The efficacy of these outputs in achieving those gains is measured as a function of AI model performance.



Think of these three layers as rungs on a ladder, allowing organizations to methodically connect AI solution performance to business impact.

Importantly, you'll travel this ladder in different directions depending on the project stage. During planning, start at the top: define your desired business outcomes first, then work down the ladder to identify the operational metrics and technical measures that will drive them. During execution and evaluation, use the reverse direction: measure technical performance first, then climb upward to ultimately assess business impact.

#### HOW THE LAYERS WORK IN PRACTICE

Typically, an AI solution is developed, deployed to production, and integrated into a business process. The AI Performance Layer has metrics that measure whether the solution's outputs meet their technical objectives, such as accuracy or reliability.

The metrics of the Operational Layer then capture the productivity gains from these technical outputs. These metrics translate AI capabilities into tangible process improvements, such as faster cycle times, reduced error rates, or increased throughput.

Finally, these Operational metrics link to Business Outcome metrics in the layer above. We say 'link' because many business metrics only become directly observable months or even years after deployment. However, by carefully connecting Operational metrics to Business Outcome metrics, organizations can estimate likely results far sooner, providing crucial near-real-time feedback on the AI solution's effectiveness.

This three-layer approach enables organizations to measure results more comprehensively and quickly, and communicate those results effectively across technical and business stakeholders.

# Choosing the Right Metrics

The framework's effectiveness hinges on selecting the right parameters: those that genuinely capture your business goals and how the AI project will achieve them. Several key considerations can guide this selection process.

## LAYER 1: BUSINESS OUTCOME METRICS

Choosing AI project metrics starts here: What are the ultimate goals the business seeks to achieve? These objectives are familiar to every executive, such as:

- **Revenue Growth:** ex. new revenue stream, increased conversion rate, market expansion
- **Cost Reduction:** ex. reduced waste, lower maintenance frequency, decreased downtime

However, as we've explained, using these metrics alone to measure the efficacy of AI projects isn't straightforward. The connection between AI outputs and business results is rarely direct or isolated from other factors.



To demonstrate why, let's consider two scenarios:

### **Scenario 1: The Confounding Event.**

A manufacturer deploys an AI-powered predictive maintenance solution that performs exceptionally well technically. Yet during the same period, a factory flood increases both downtime and maintenance costs. Looking only at Business Outcome metrics would suggest the AI project failed, masking its actual success.

**Scenario 2: The False Positive.** A company implements an AI solution to boost sales team performance, however engineering simultaneously introduces a popular new product feature. Sales increase – but was it the AI, the new feature, or both? An ineffective AI solution could easily claim undeserved credit.

Both scenarios demonstrate why Business Outcome metrics are essential but insufficient on their own. They tell you what happened, but not why – making it nearly impossible to isolate the AI solution's true contribution.

This is where the next layer becomes critical: Operational metrics provide the missing link in the measurement chain.

## LAYER 2: OPERATIONAL METRICS

Most AI projects aim to increase productivity in some measurable way.

AI solutions generally produce an output – predictions, optimizations, recommendations, generated content, etc. Sometimes the generated output directly improves productivity. More often, however, productivity gains materialize when people use these outputs in downstream business processes. Operational metrics capture and quantify these productivity effects in concrete terms.

Common Operational metrics are:

- **Time saved:** ex. reduced time per task or per employee hour
- **Throughput optimization:** ex. increased deliveries per truck per unit of fuel
- **Capacity utilization:** ex. increased machine uptime per hour
- **Error reduction:** ex. fewer overstaffed shifts due to inaccurate demand forecasting
- **Pricing optimization:** ex. improved profit margin per transaction

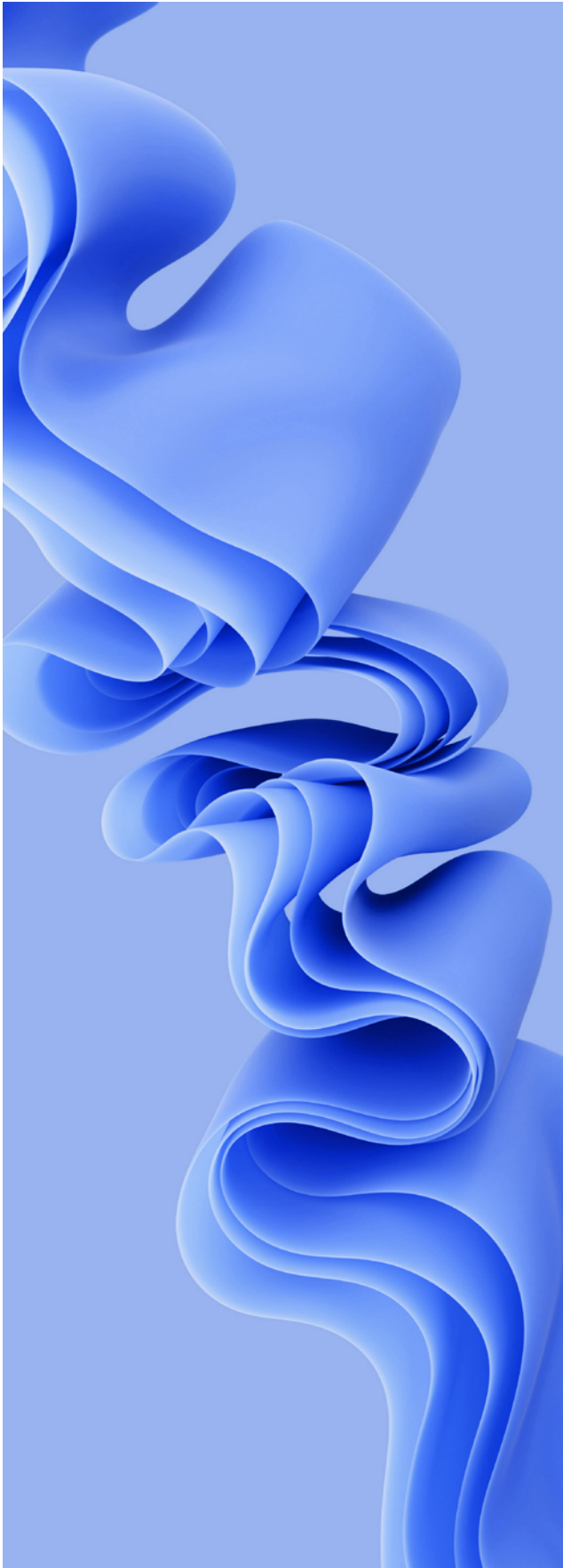
Selecting Operational metrics begins after your Business Outcome metrics have been chosen. Ask yourself: what specific process factors affect the desired Business Outcome metrics? What productivity inputs are required to produce the business outcome you're targeting? These factors become your Operational metrics - the levers where productivity gains must occur.

*Example:* A distribution company seeking cost reduction identifies that energy consumption is growing faster than revenue, eroding profitability. Within their transport division, truck fuel costs are particularly high. Throughput optimization - deliveries per truck per unit of fuel - becomes the operational metric. Improving this specific factor directly contributes to the Business Outcome of reduced costs.

Measuring Operational metrics without connecting them to Business Outcome metrics often obscures whether there's a net benefit to the organization. A process improvement that increases productivity but doesn't advance strategic goals may not justify the investment. This is why establishing clear, logical links between Operational metrics and Business Outcomes is essential.

With Business Outcomes and corresponding Operational metrics now defined, one critical piece remains: measuring the AI performance layer. Without this layer, the connection between the AI solution's efficacy and the resulting business process improvements remains invisible.





### LAYER 3: AI PERFORMANCE LAYER

Once deployed, an AI solution's output quality can be technically measured and assessed. In some respects, this is the most straightforward layer: Did the model produce accurate, reliable outputs when evaluated against expectations?

The specific metrics depend on the AI model type(s) involved. While selecting these metrics typically requires significant AI expertise, common examples include:

- **Predictive models:** ex. accuracy in forecasting known outcomes (error rate, precision, recall)
- **Optimization models:** ex. quality of decisions relative to operational objectives (objective value achieved, constraint feasibility)
- **Generative models:** ex. quality and reliability of outputs (e.g., human evaluation scores, relevance, consistency)

The AI team responsible for technical execution will select the appropriate model type to achieve desired outcomes. Once determined, the range of AI Performance metrics is relatively well-established through research and industry practice. However, selection still requires judgment to reflect operational context, risk tolerance, and downstream decision-making processes.

*Example:* Improving a predictive maintenance model's precision from 85% to 92% is meaningful to data scientists. But what does it mean for the business? Through Operational metrics, a 7-point technical improvement might translate to 15% fewer false alarms, which reduces unnecessary maintenance visits and ultimately significantly cuts annual costs. This translation from AI Performance metrics to Operational metrics is what enables executives to evaluate the AI investment's true business impact.

While technical teams will always fixate on AI Performance metrics, having the ability to translate which of these affect Business Outcome metrics, and how, is crucial to the efficacy of the three-layer approach.

# Putting It All Together

To illustrate how the three layers work in practice, let's examine a real-world scenario.

## EXAMPLE: INVENTORY OPTIMIZATION THROUGH AI DEMAND FORECASTING

### The Challenge



A manufacturer holds excess inventory due to unreliable sales forecasts. To avoid costly stockouts, they maintain high inventory levels, which erodes profit margins. The executive team believes that improving demand forecast accuracy and reducing days inventory outstanding will lead to lower inventory costs (the desired business outcome).

### Planning the Metrics (Top-Down)

The company defines metrics starting from the desired destination and working backward, down the ladder - like planning a road trip by first deciding where you want to go, then determining how to get there.

#### **Layer 1: Business Outcome Metric**

- Reduce inventory carrying costs

#### **Layer 2: Operational Performance Metric**

- Reduce days inventory outstanding by 10%, from 90 to 81 days
- This target was validated through a proof-of-concept early in the project, demonstrating achievability

#### **Layer 3: AI Performance Metrics**

The AI service provider recommends building two interconnected models – a demand forecast model, to predict demand for specific SKUs over defined time periods, and an inventory optimization model that uses the demand forecast to recommend order quantities for specific SKUs.

For each model, the following AI Performance metrics are selected to connect to the Operational metrics:

- **Demand Forecast Model Metric:** mean absolute error, comparing predictions to actual demand
- **Inventory Optimization Model Metrics:** Constraint feasibility (valid recommendations under business rules), optimality gap (proximity to minimum achievable cost), solution stability (consistency across planning cycles)

## Measuring Success (Bottom-Up)

Once the solution is deployed, measurement flows in reverse, from technical performance back up the ladder to business impact.

### *Layer 3: AI Performance Metrics*

→ AI models perform as designed. Both the demand forecast and inventory optimization models meet their technical performance targets for accuracy, feasibility, and stability.

### *Layer 2: Operational Performance Metrics*

→ Operational improvements materialize. Inventory managers now base ordering decisions on data-driven recommendations rather than individual judgment (which varied widely by employee tenure and experience). Days inventory outstanding decreases from 90 to 81 days.

### *Layer 1: Business Outcome Metrics*

→ Business outcome achieved. With less inventory outstanding, costs are reduced, and the business unit is able to reduce its budget allocation for inventory purchasing.

## The Value of the Three-Layer Approach

By measuring across all three layers, companies can:

- **Quickly assess** whether the business outcome is likely to be achieved, even before final results are observable
- **Pinpoint issues** when results fall short, identifying which layer is underperforming
- **Troubleshoot effectively** by addressing root causes rather than symptoms

For instance, if inventory costs aren't decreasing as hoped, the company can diagnose whether the problem lies with model accuracy (Layer 3), employee adoption of recommendations (Layer 2), or external factors affecting the business outcome (Layer 1). This systematic approach transforms what would otherwise be a black box into a transparent, manageable process.



# Conclusion

**This three-layer approach provides a methodical way to systematically connect the dots between technical progress and desired business outcomes. It enables organizations to measure AI impact accurately and, just as importantly, communicate results to decision-makers, to enable informed judgments about whether projects have met their objectives.**

Effective measurement demands a structured approach that connects AI technical performance to operational metrics, and operational metrics to business outcomes. Each layer serves a distinct purpose; together they provide a complete picture of whether an AI initiative is technically sound, operationally effective, and - ultimately - worth the investment.



## DEFINE METRICS EARLY

The framework's power lies not just in selecting the right metrics, but in establishing them before development begins. This enables organizations to:

- Assess progress early and govern projects effectively
- Distinguish genuine value creation from misleading signals
- Troubleshoot issues at the appropriate level
- Determine whether a solution needs minor adjustments or fundamental redesign

Organizations that treat metrics as a first-order design decision, rather than an afterthought, will be better positioned to move confidently from experimentation to sustained impact.

## THREE QUESTIONS TO ASK BEFORE YOU BEGIN

Before launching your next AI initiative, pause and reflect on three questions:

1. What business outcomes are we trying to achieve?
2. What operational processes must improve for these business outcomes to materialize, and how will we measure those improvements?
3. What AI technical performance is required to deliver the desired operational improvements?

When these questions can be answered clearly and consistently across technical teams, business leaders, and external partners, your project is well-positioned for successful execution and scale.

# A Final Word

AI adoption carries risk - but the risk of AI *inaction* is usually greater. The competitive landscape is shifting rapidly, and organizations that delay AI integration risk falling behind. This three-layer measurement framework allows you to proceed with confidence, transforming AI from a leap of faith into a managed, measurable business initiative.

The path forward is clear: define your outcomes, establish your metrics, and measure what matters.

Acknowledgments:

The production of this report required the combined efforts of many members of the Scale AI team. We would like to thank Sean Duckett and Yassine Lazraq for their significant contribution to this report, and Todd Bailey for the review and editing of the document.

We would also like to thank Guillaume Lavoie for his efforts in the publication and diffusion of this report.

## Reach Out ↓

For any question regarding project submissions or intellectual property:

[info@scaleai.ca](mailto:info@scaleai.ca)

**SCALE|AI**

