The Real Reason AI Doesn’t Scale

FROM ISOLATED USE CASES TO SYSTEMS OF DECISIONS

PERSONAL POV | BY ESTEFANÍA MOLINA

/ 2026

Shared in a personal capacity; Views expressed are my own.

The emerging gap between investment and impact

Across industries, a consistent pattern is beginning to emerge. Organizations are making significant investments in AI, identifying high-value opportunities, deploying models, and in many cases achieving tangible improvements within individual domains. Yet, despite this progress, **the overall impact at the enterprise level often falls short of expectations**. Gains remain localized, and the anticipated step-change in performance does not materialize at scale.

In many cases, the gap is not marginal. Organizations report measurable improvements within individual use cases yet struggle to translate those gains into enterprise-level outcomes. E.g., a pricing model may improve margins by 2-3% in isolation, while supply chain inefficiencies or demand misalignment offset those gains elsewhere. The result is that substantial investment *yields* limited net impact.

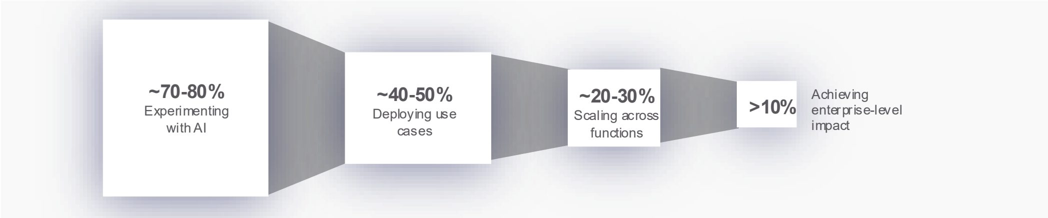
This pattern is increasingly reflected in industry research. While most organizations are experimenting with AI, only a *small* fraction achieve meaningful financial impact at scale, with many initiatives remaining confined to pilots or isolated deployments[1,2,3].

This dynamic is frequently interpreted as an execution challenge - a matter of refining models, improving data quality, or accelerating adoption. While these factors are not irrelevant, they do *not* fully explain the pattern. The more fundamental issue is **structural**.

The way organizations are currently designed constrains their ability to translate isolated improvements into system-wide performance.

EXHIBIT 1

From experimentation to enterprise value: where organizations stall *(Illustrative)*



Most organizations generate value in isolated use cases but struggle to scale impact across the enterprise

The limitations of optimizing in isolation

The difficulty in scaling impact is closely related to how AI initiatives are defined. In most organizations, they are framed as *discrete* use cases, each with its own scope, funding, and delivery model. Pricing optimization, demand forecasting, workforce planning, and customer targeting are treated as independent problems to be solved within functional boundaries.

This approach is effective when value can be captured locally. However, many of the decisions these use cases address are inherently **interconnected.** Pricing decisions influence demand patterns; demand shapes supply requirements; supply depends on workforce capacity and inventory positioning. These relationships are not sequential but interdependent.

When decisions of this nature are optimized independently, the result is not a more effective system but a more complex one. Improvements are achieved within individual steps, yet the structure of decision-making across the enterprise remains unchanged. As a consequence, gains do not accumulate. Instead, they are often redistributed, with benefits in one area introducing variability or cost in another.

What is often described as workflow redesign/transformation therefore tends to remain limited in scope. Rather than redesigning how decisions relate to one another, organizations improve components within an existing structure. **The underlying system continues to operate as a collection/sequence of loosely coordinated functions.**

Enterprise impact: isolated vs connected decisions

The implications of optimizing decisions in isolation become clearer when viewed through their financial impact.

Consider a consumer-facing organization deploying AI across **three high-impact areas**: (i) pricing, (ii) demand forecasting, and (iii) supply chain optimization. Each initiative delivers measurable improvements *within* its respective domain. E.g., pricing models can increase margins in the range of 2-5%. Demand forecasting can improve accuracy by 10-20%, reducing planning errors. Supply chain models can reduce inventory levels by 10-30% while maintaining service targets.

Evaluated independently, these results suggest meaningful performance improvement - at the enterprise level, however, the realized impact is often *significantly* lower.

Pricing decisions can introduce demand variability that the supply chain is not structured to absorb, resulting in stockouts, expedited shipping, and service degradation. Forecast improvements are not always reflected in pricing strategies, limiting their effect on revenue outcomes. Inventory optimization can reduce buffers that commercial activity depends on, increasing operational fragility.

As a result, a portion of the value created by individual models is offset elsewhere in the system.

Margin improvements generated through pricing are *diluted* by increased fulfillment and service costs. Inventory reductions are *eroded* by the reintroduction of safety stock to compensate for volatility. Forecast accuracy does *not* translate proportionally into either revenue growth or cost efficiency.

In directional terms, this often leads to outcomes such as e.g.,:
(i) Margin improvements of 2-5% at the use case level translating into approximately 1-2% at the enterprise level once downstream effects are considered
(ii) Inventory reduction targets of 10-30% resulting in realized improvements closer to 5-15% when service buffers are reintroduced
(iii) Incremental increases in logistics and service costs in the range of 1-3%, partially offsetting gains achieved elsewhere

Despite strong performance at the level of individual models, a meaningful share of value is either diluted or redistributed across functions.

This dynamic changes when decisions are structured as part of a connected system.

Pricing decisions are informed by supply constraints and service level requirements. Demand forecasts directly influence both pricing strategies and supply allocation. Supply chain decisions are optimized not only for efficiency, but for responsiveness to demand patterns created by pricing actions.

Trade-offs are defined in advance, rather than resolved during execution. Margin targets are balanced against service levels and fulfillment costs within a shared decision framework, and constraints are embedded in the system to ensure consistency.

Under these conditions, outcomes begin to reinforce one another.

In directional terms, organizations are more likely to observe:
(i) A greater share of margin improvements retained, often in the range of 2-4%, with limited downstream offset
(ii) Logistics and service cost leakage reduced to below 1% as variability is better managed
(iii) Inventory reductions in the range of 10-25% sustained without requiring the reintroduction of buffers

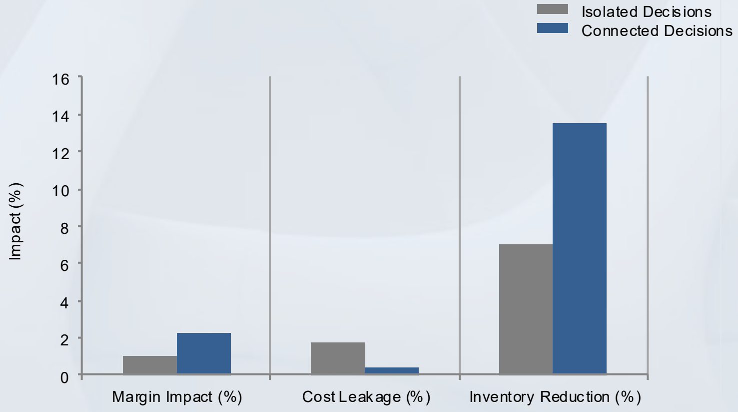
The exact magnitude will vary by organization. The pattern, however, is consistent.

The difference is *not* in the performance of individual models, but in how decisions are connected.

When decisions are managed independently, value is partially captured and often redistributed. When decisions are coordinated, value *compounds.*

EXHIBIT 2

Isolated vs Connected Decision System *(Illustrative)*



*Improvements from individual AI use cases often fail to translate into enterprise impact when decisions are managed independently. Coordinating decisions enables value to **compound** rather than offset.*

The AI spectrum and the point of structural pressure

The variation in enterprise impact is not random. It reflects differences in how AI influences decision-making and, more importantly, in how those decisions relate to one another. **Not all applications exert the same level of pressure on the enterprise, and not all require a change in the operating model.**

At one end of the spectrum, AI improves efficiency by **automating** tasks and accelerating workflows. These applications operate within existing processes and decision contexts, but do not change how decisions are defined or how they interact with one another. Human actors remain fully in control, and trade-offs are minimal or contained within a single function. As a result, they can typically be absorbed without materially altering the operating model.

As AI begins to **support** decisions, its influence extends into forecasting, prioritization, and planning. Decisions remain human-led but are increasingly shaped by system outputs. At this stage, the impact depends on the scope of the decisions involved. When decisions remain within a single function, coordination requirements are limited. However, when these decisions begin to influence outcomes across functions, trade-offs emerge that are no longer contained. These trade-offs are often defined locally and resolved during execution, typically through coordination between teams.

The constraint becomes structural when AI systems either **shape or directly make** decisions that are both continuous and interdependent across functions. Under these conditions, trade-offs can no longer be resolved during execution without introducing inconsistency or delay. Instead, they must be defined in advance and embedded into how decisions operate. Human participation shifts accordingly, from making individual decisions to defining objectives, constraints, and trade-offs within which decisions are executed.

This transition reflects a *broader* shift in how decision-making is structured. It is not driven solely by the level of autonomy, but by the combination of autonomy and interdependence. **As decisions become both system-driven and cross-functional, the effectiveness of AI depends *less* on the performance of individual models and *more* on how decisions are structured, connected, and governed across the enterprise.**

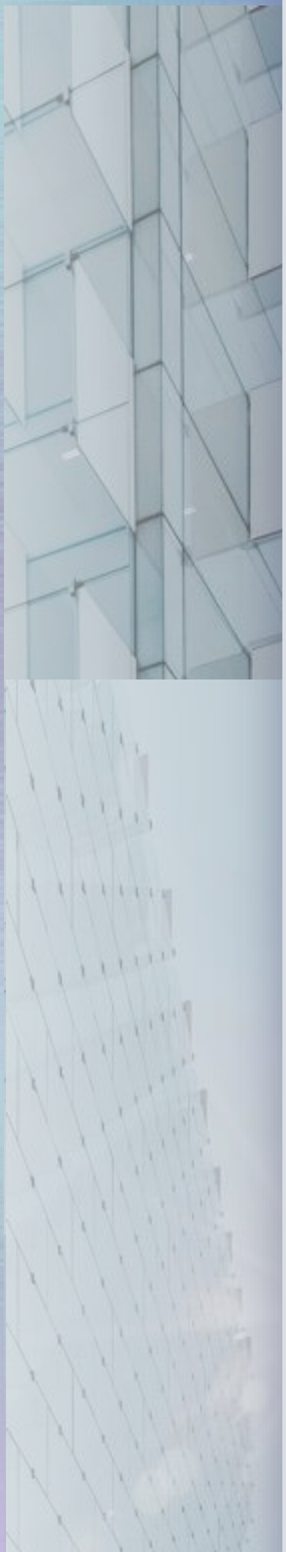
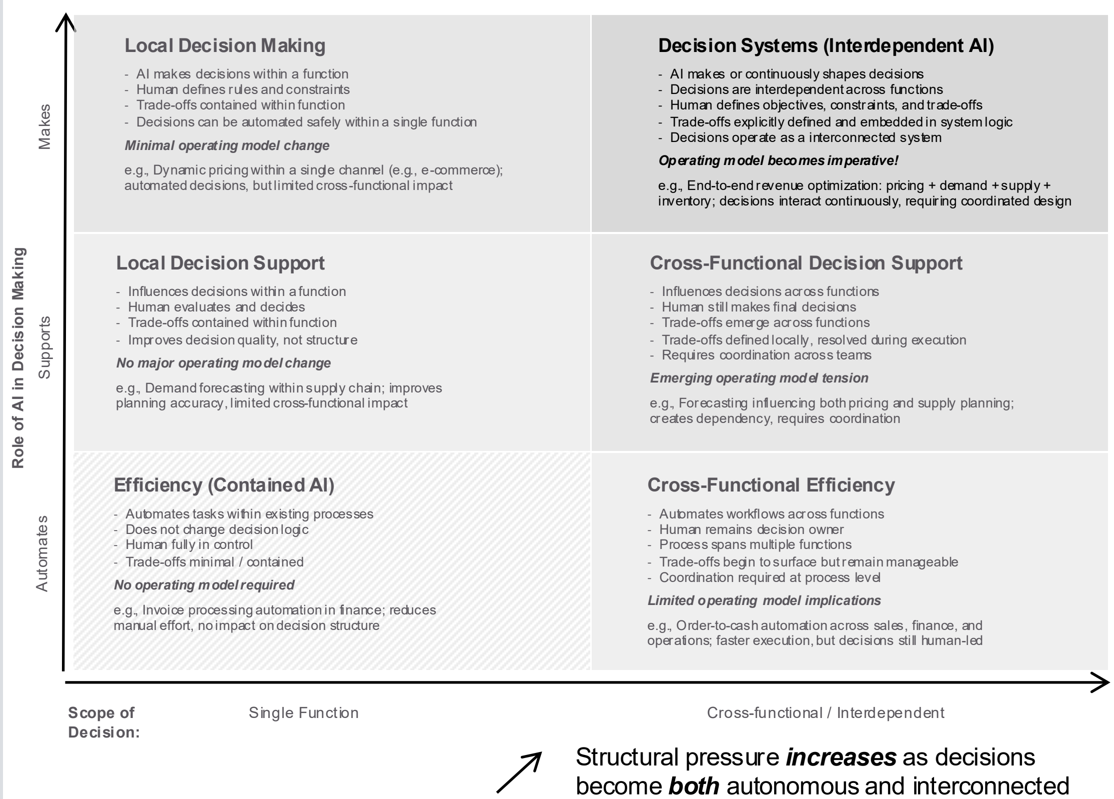


EXHIBIT 3

Not all AI requires an operating model shift *(Illustrative)*



Shared in a personal capacity; Views expressed are my own.

Coordination as the limiting factor

As AI moves beyond isolated applications and begins to influence interconnected decisions, the primary constraint shifts. It is no longer defined by the performance of individual models, but by the ability of the organization to **coordinate** how decisions interact.

In most enterprises, decision-making remains distributed across functions, each optimizing for its own objectives. This model is effective when decisions are largely independent. However, as AI increasingly shapes decisions that span functions, interdependencies become more pronounced, and the limitations of this structure begin to surface.

Under these conditions, even well-performing models can produce inconsistent outcomes. Decisions that are locally optimal may conflict when viewed at the system level. **Improvements in one area can introduce variability, cost, or risk in another,** not because the underlying models are flawed, but because the interactions between decisions are not *explicitly* managed.

The challenge is not that decisions are distributed, but that their interactions are not designed. In practice, coordination is often handled through informal mechanisms such as escalation, cross-functional alignment, or manual overrides. These approaches assume that decisions are discrete and can be reconciled through human intervention. They become less effective when decisions are continuous, embedded in systems, and operating at scale.

As a result, organizations experience diminishing returns on incremental AI investment. Each additional use case delivers value within its domain, but the overall enterprise impact does not scale proportionally. **The constraint lies not in the capability of the technology, but in the absence of a structure that defines how decisions should work together.**

This raises a more fundamental question: **can an organization operate as a coordinated system of decisions, rather than as a collection of functions optimizing independently, without relying on a centralized function to resolve trade-offs?**

Aviation as a reference for coordinated systems

This type of coordination problem is not unique to AI. There are environments in which outcomes depend on multiple actors making interdependent decisions under conditions of incomplete alignment, without relying on centralized intervention to resolve trade-offs. **Aviation provides a relevant reference point.**

Airlines optimize for asset utilization and profitability. Airports optimize for throughput and capacity. Air traffic control prioritizes safety and separation. Regulators enforce compliance and risk thresholds. These objectives are not fully aligned and, in many cases, are structurally in tension.

Yet, the system functions as one with a high degree of reliability.

This performance is not achieved through centralized decision-making, nor by requiring alignment across all actors. It is achieved by ensuring that independent optimization does not destabilize the system.

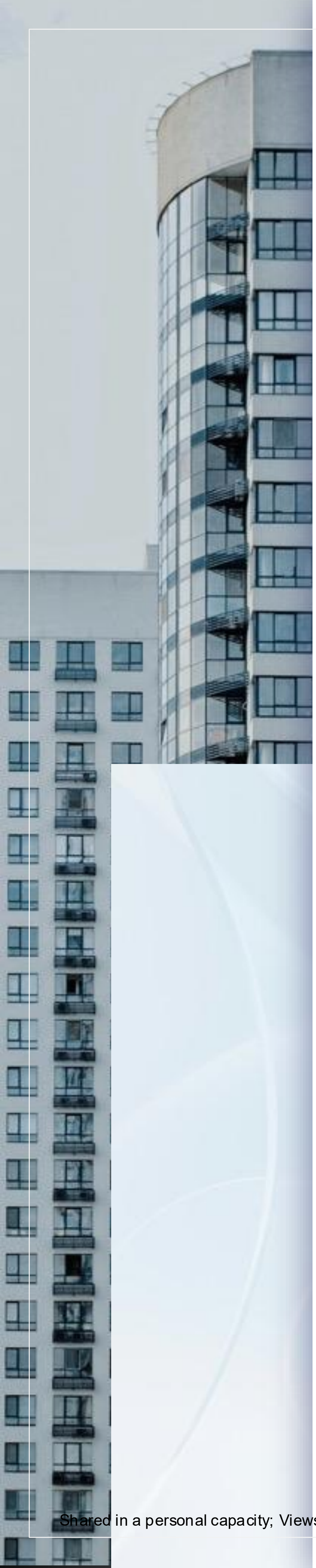
Three characteristics enable this:
(i) **Decision rights are explicitly defined**, providing clarity on who acts and under what conditions. At any given moment, it is unambiguous which actor controls a decision and which constraints apply.
(ii) **Constraints are embedded into the system.** Decisions are not made freely, but within clearly defined boundaries that preserve system stability. Safety, spacing, sequencing, and capacity limits are not negotiated in real time; they are designed into how the system operates.
(iii) **Interactions between actors are structured in advance.** The way decisions affect one another is not left to interpretation or coordination through discussion. It is governed by predefined protocols that ensure consistency even when incentives differ.

The system does not depend on alignment to function. **It depends on design that accommodates misalignment without allowing it to disrupt coordination.**
This distinction is critical… →

As AI distributes decision-making across functions, enterprises begin to operate under similar conditions. Decisions are interdependent, incentives are not fully aligned, and outcomes are shared.

However, most organizations continue to rely on structures designed for independent decision-making, where coordination is treated as a secondary layer rather than a core design requirement.

The relevance of aviation is *not* that it offers a model to replicate directly, but that it demonstrates what becomes necessary when decisions are both distributed and interdependent. **Coordination must be designed into the system itself, rather than managed through centralized intervention or alignment.**



From functional models to decision-based design

The implications of this shift extend beyond individual use cases or technologies. They point to a more fundamental question of **how the enterprise itself is structured.**

Most operating models are organized around functions, with processes and systems layered on top. This reflects an assumption that decisions can be decomposed, owned, and optimized within functional boundaries, with coordination handled as a secondary concern. As long as decisions remain largely independent, this model is effective.

As interdependence increases, this assumption becomes less viable**. Value is no longer generated primarily through the performance of individual activities, but through the interaction of decisions across the enterprise.** Under these conditions, improving decisions in isolation is insufficient. What matters is how those decisions interact.

This requires a shift in the unit of design.

Rather than organizing solely around functions, **the enterprise must increasingly be designed around decisions.**

This does not imply removing functional structures. Functions remain necessary for capability, expertise, and execution. However, they are no longer sufficient as the primary organizing logic for how value is created. The critical layer becomes the system through which decisions are defined, connected, and governed.

As outlined earlier, not all AI requires this level of redesign. The need emerges when AI moves beyond efficiency into influencing or making decisions that are interdependent across functions. At that point, coordination cannot be managed through alignment alone; it must be designed into the system.

Without this shift, organizations are likely to continue capturing only a portion of the value created by individual AI initiatives. Improvements will remain localized, while variability, cost, and risk are introduced elsewhere in the system. Over time, this results in diminishing returns on additional use cases, as the underlying constraint is not the capability of the models, but the absence of coordination across the decisions they influence.

A practical starting point for leaders is to focus on a small set of these interconnected decisions, where this dynamic is already visible:

First, identify where AI is influencing interdependent decisions.
Prioritize use cases that sit beyond isolated efficiency gains and begin to shape decisions across functions, such as pricing, demand planning, supply allocation, or service levels. These are typically the areas where the gap between use case performance and enterprise impact is most pronounced.

Second, define how these decisions interact.
Make the decisions explicit, clarify ownership, and map how one decision constrains or informs another. Rather than treating use cases as independent initiatives, define the relationships between them. This is where the shift from isolated use cases to a connected decision system begins.

Third, embed trade-offs into decision logic.
Instead of resolving competing objectives during execution, define in advance how trade-offs are balanced across decisions. This includes aligning objectives, constraints, and model logic so that decisions reinforce rather than offset one another.

Technology then becomes the mechanism through which this structure is enforced. It connects decisions through shared data and models, applies constraints consistently, and ensures that trade-offs are respected across use cases.

This does not require centralizing decision-making. Decisions can remain distributed but must operate within a clearly defined structure. **The shift is not from decentralization to centralization, but from implicit coordination to explicit design.**

KEY TAKEAWAYS

1. AI does not fail to scale because of models - it fails because **decisions are not designed to work together**

2. The primary constraint is not capability, but **the absence of a system that coordinates decisions**

3. Value is not created by individual use cases, but by **how decisions interact across them**

4. Organizations do not underperform because decisions are distributed, but **because their interactions are undefined**

5. Optimizing decisions in isolation leads to value leakage; **coordinating them enables value to compound**

6. **Not all AI requires an operating model shift** - only when decisions become both interdependent and system-driven.

7. The challenge is not that trade-offs exist, but that **they are resolved too late**

8. Trade-offs should not be resolved during execution; they **should be defined as part of the system**

9. Governance is no longer about overseeing decisions, but about **designing the conditions under which they operate**

10. The enterprise does not need more alignment; **it needs a system that works despite misalignment**

Shared in a personal capacity; Views expressed are my own.

References

[1]McKinsey & Company. (2023). *The state of AI in 2023: Generative AI's breakout year*. Retrieved from https://www.mckinsey.com

[2]Boston Consulting Group. (2023). *AI at scale: From experimentation to impact*. Retrieved from https://www.bcg.com

[3]Davenport, T. H., & Ronanki, R. (2018). *Artificial intelligence for the real world*. MIT Sloan Management Review, 59(4), 108–116.

About the Author

Estefanía Molina
https://www.linkedin.com/in/estefaniamolinae/

Strategy consultant with over eight years of experience advising Fortune 500 executives across the Americas on enterprise transformation, AI strategy, and organizational redesign. Her work focuses on translating complex, interconnected decisions into measurable business outcomes, including large-scale initiatives delivering hundreds of millions of dollars in value.

She has led engagements spanning enterprise AI agendas, operating model redesign, and global transformations, partnering directly with CxOs on high-stakes strategic decisions. Originally from Mexico and now based in New York City, she has a long-standing personal interest in op models and aviation, shaped since childhood, which continues to influence how she thinks about complex, coordinated systems.

Outside of her client work, she is a Master's in Finance candidate at Harvard University.

Shared in a personal capacity; Views expressed are my own.