McKinsey & Company

Industrials Practice

Cutting the ‘coordination tax’: How agentic AI can reshape workflows

Organizations must redesign workflows to realize value from AI, but the challenge is where to begin. An opportunity lies between workflow steps, where work passes between teams and systems.

by Ali Sankur, Bernhard Mühlreiter, and Steffen Fuchs



September 2026

Most large organizations have spent the past several years investing heavily in AI. They have launched copilots, built proofs of concept, and are increasingly scaling AI across individual functions. Eighty percent of respondents to McKinsey's latest Global Survey on AI report that AI has improved their individual productivity. Yet only 37 percent attribute any EBIT impact to their organization's use of it.[1]

Why? The problem is less about the sophistication of the technology than where it has been deployed. Most organizations use AI to make individual workflow steps more efficient. A larger opportunity lies elsewhere: in the coordination that takes place between those steps.

Over the past several years, organizations have learned another lesson, as well: AI rarely creates enterprise value by improving isolated tasks. The biggest gains come from reimagining end-to-end workflows around what the technology makes possible. That insight is now well established. The more difficult question is how. Where should leaders begin retooling complex workflows? Which parts of a workflow deserve the most attention? Why do some redesign efforts fundamentally change performance while others deliver only incremental improvements?

One answer is that rather than focusing on the work performed inside each process step, leaders should begin at the interfaces where work passes from one team, function, or system to the next. Those interfaces are where organizations confront what we call "the coordination interface tax"—the total cost an organization pays for human-mediated handoffs between workflow steps.

Such interactions have been largely ignored by previous generations of automation. While those efforts excelled at making individual steps faster and more efficient, they could not reliably perform the judgment required at the interfaces between them. Agentic AI changes that. In our experience, it is the first generation of technology capable of automating much of the verification, reconciliation, and routing that previously required human coordination. The opportunity is not only to make existing workflows more efficient but to redesign workflows in ways that reduce the hidden costs of coordination—costs that, in our experience, are the largest unaddressed expense in enterprise operations.

Most companies can draw their workflows on a whiteboard: order to cash, procure to pay, demand to production. After decades of process improvement, the steps within these workflows are designed, measured, and optimized. Yet, across knowledge-intensive organizations, coordination—aligning, verifying, reconciling, waiting, and handing work from one step to the next—consumes 35 to 60 percent of total work time.[2]

But the costs that these delays exact are largely invisible. The reason is simple: No function owns the interfaces. The demand planning team is responsible for its forecast. The supply planning team owns its allocation. But nobody is accountable for the gap between them, where the forecast waits for reconciliation against capacity constraints. That gap is organizational white space, unowned and unmeasured.

[1] "The state of AI in 2026: On the road to ROI," McKinsey, August 25, 2026; n = 1,719 participants at all levels, fielded May 4 to June 8, 2026. Eighty percent of respondents report AI has improved their individual productivity; 37 percent attribute at least some EBIT impact to their organization's AI use; 6 percent qualify as high performers, attributing at least 5 percent of EBIT to AI and describing the impact as significant.

[2] The 35 to 60 percent range is a convergent estimate drawn from studies using different methods and definitions, not a single measured figure. See Thomas Malone and Kevin Crowston, "The interdisciplinary study of coordination," *ACM Computing Surveys*, March 1994, Volume 26, Number 1, which provides the conceptual framework, defining coordination as the management of dependencies between activities. Mintzberg's observational studies report 50 to 80 percent for managers specifically, measuring oral communication. Microsoft's Work Trend Index 2023 finds 57 percent of time spent communicating in meetings, email, and chat, measured as a share of activity within Microsoft 365 applications rather than of total work time. Asana Anatomy of Work Index 2023 reports 58 percent in self-reported coordination activities.

The coordination tax hides in three places, only one of which appears as an item in selling, general, and administrative expenses (SG&A), an organization's day-to-day overhead costs that are not related to the production of goods or services. The visible portion is people and structures dedicated to coordination: planning teams, project management offices, governance forums, and review meetings. In manufacturing and industrial companies, this typically runs 3 to 5 percent of revenue, McKinsey research has found. Traditional cost programs target it regularly.[3]

Across the companies we have studied, that visible cost typically represents only about one-third of the total coordination tax. The remainder hides in places that no cost program has historically addressed. The second portion can be found in cycle time and working capital. Every day of latency at a coordination interface carries a financial cost, such as inventory held longer, receivables collected later, and capacity committed but not yet producing revenue.

The third portion appears nowhere in financial statements. It consists of the revenue not captured because the organization could not coordinate fast enough, such as products launched months later because engineering-to-production handoffs added delay, or market share not won because pricing decisions traversed too many approval layers. This is the largest portion and the least visible, because it represents what did not happen. No executive receives a report showing revenue lost to coordination latency. The costs may be real, but the measurement is absent.

What the coordination tax looks like in practice

McKinsey recently measured actual latency at each coordination interface in a standard demand-to-production workflow at a Fortune 500 industrial manufacturer. The workflow has seven steps. Between them sit six coordination interfaces (table).[4]

Interface latency exceeds processing time by nine to 36 times. The workflow takes roughly one and a half to two and a half weeks to traverse. The actual work inside the steps takes half a day to a day. The rest is coordination, such as verifying that one step's output meets the next step's requirements, reconciling discrepancies, obtaining approvals, and waiting for human availability.

We estimate that the total coordination interface tax for this single workflow came to $140 million to $240 million annually. The figure comprises latency-driven working capital cost, underutilized capacity, and missed revenue from delayed response to demand signals. At this manufacturer, the visible portion—the people dedicated to coordination—accounted for less than 7 percent of that total, well below the typical one-third. The business is capital-intensive and carries a high working-capital base, so latency cost dominates. The remaining 93 percent appeared as latency and opportunity costs, which showed up nowhere on the income statement.

This is why traditional cost programs miss it. Such programs target the 7 percent they can see: head count, meetings, and reporting layers. The other 93 percent is in cycle time, working capital, asset utilization, and revenue that never materializes because work moves too slowly between functions.

[3] Directionally consistent with S&P Capital IQ sector median SG&A ratios for industrials. Coordination-intensive functions typically constitute 3 to 5 percent of revenue. This captures only dedicated coordination head count; it excludes coordination time embedded in operational roles.

[4] Disguised client example: Fortune 500 industrial manufacturer, demand-to-production workflow redesign engagement (2024–25). Specific figures drawn from observed data, with magnitude preserved and identifying details altered. The $140 million to $240 million range represents the sum of working capital cost from interface latency, capacity underutilization, and estimated revenue foregone.

Table

Process step	Interface latency	Processing time within step
Sensing to planning	1–2 days	2–4 hours
Planning to supply	3–5 days	4–8 hours
Supply to scheduling	1–2 days	2–4 hours
Scheduling to procurement	1–3 days	1–2 hours
Procurement to execution	2–5 days	1–2 hours
Execution to fulfillment	1 day	2–4 hours
Total	**9–18 days**	**12–24 hours**

Research increasingly suggests that these interfaces are where AI has the potential to create the greatest value. A recent field experiment by researchers at INSEAD and Harvard Business School randomized 515 start-ups into two groups, each of which received identical AI tools. One group redesigned workflows across handoff points. The other used AI within existing workflow steps. Firms that redesigned across interfaces generated 1.9 times more revenue and were 11 percentage points more likely to acquire paying customers. The study involved early-stage start-ups rather than large enterprises, so the magnitude of the effect may differ at scale. But the finding is striking in its consistency with the industrial evidence: What matters is not simply the sophistication of the AI but where it is deployed.[5]

Why this has never been addressed: Five distinct eras; one common blind spot

Before exploring how to seize the newfound opportunities posed by agentic AI, it helps to take a moment to reflect on how we arrived here. The history of workflow improvement is a story of cumulative progress. Each era solved a real problem and created enormous value. Yet each also left the coordination interfaces largely untouched.

Seeing the flow (1980s–90s)

Quality movements, lean thinking, and Six Sigma made work visible as a connected system. Organizations learned to map the current state, identify waste, and measure cycle time across functions. The conceptual advance was enormous: For the first time, workers understood their activities as a system rather than a collection of departments.

[5] Hyunjin Kim, Dahyeon Kim, and Rembrand Koning, *Mapping AI into production: A field experiment on firm performance*, INSEAD working paper 2026/20/STR, March 30, 2026.

Redesigning the work (1990s)
Proponents of business process reengineering (BPR) often made a bold claim: "Don't automate, obliterate." When executed well, BPR delivered cycle time reductions of 50 to 90 percent. Ford reduced its North American accounts payable head count by 80 percent. Hallmark cut time to market by 70 percent. The ambition was transformative, but the unit of redesign was still a step.[6]

Connecting the data (1990s–2000s)
Enterprise resource planning (ERP) systems created integrated data across functions. For the first time, a transaction could flow across functional boundaries without manual reentry. Implementations delivered inventory reductions of 20 to 50 percent in the first two years. The promise was a single source of truth, but in reality, shared data did not eliminate the need for shared judgment at each handoff.[7]

Owning the outcome (2000s–10s)
Process management formalized ownership, measurement, and continuous improvement. Organizations created process offices and governance around their end-to-end flows. The discipline was meaningful, but in our experience, formalizing the interfaces documented rather than reduced their costs.[8]

Automating the task (2010s–20s)
Robotic process automation (RPA) targeted the manual effort persisting within individual steps. Process mining revealed where actual execution diverged from design. Targeted tasks were compressed by 30 to 70 percent when applied. But automating a task within a step does not address the verification, reconciliation, and approval work required before that step is handed off to the next.[9]

The pattern across all five eras is clear: Each was driven by a technological advance; each required organizations to redesign around what the technology made possible; and each rewarded those that moved first. But in every case, the unit of optimization was the process step. Those individual steps may have improved, but the arrows between them remained largely unchanged.

In our experience, no previous generation of technology could operate at the interface itself with the judgment required to verify, reconcile, and route work across competing constraints. That capability did not exist. Now it does.

[6] Michael Hammer and James Champy, *Reengineering the Corporation: A Manifesto for Business Revolution*, HarperBusiness, 1993, which reports Ford's North American accounts payable reduction at 80 percent; Thomas H. Davenport, *Process Innovation*, Harvard Business School Press, 1993, provides similar benchmarks.
[7] Sheri Shang and Peter B. Seddon, "Assessing and managing the benefits of enterprise systems," *Information Systems Journal*, October 2002, Volume 12, Number 4; this benefits-framework study was developed from 233 vendor-reported implementation cases and interviews at 34 organizations. Panorama Consulting Group ERP reports provide practitioner benchmarks on inventory reduction in initial deployment years.
[8] *Business process model and notation (BPMN) 2.0*, Object Management Group, January 2011. The standard formalized process representation. The observation that this documented rather than reduced interface costs is our own.
[9] *State of the Automation Professional Report*, UiPath, 2020–24; *Gartner Magic Quadrant for robotic process automation*, Gartner, 2023. Task-level RPA implementations report time reductions of 30 to 70 percent on targeted activities. Vendor-reported figures may overstate due to selection bias in case publication.

How agentic AI works within interfaces

AI creates value not so much by performing individual tasks as by connecting them into continuous flows that previously required human mediation at each junction.[10] A copilot that helps a planner write a demand forecast faster does not address the three days that the forecast waits for approval. An RPA bot that automates data entry within a step does not reduce the verification overhead at the boundary where that step is handed off to the next. Despite the individual productivity gains AI tools deliver, the interface remains a constraint.

Agentic AI systems can verify, reconcile, route, and act at workflow handoffs with judgment, at machine speed, and at a fraction of the cost of the human coordination they replace.[11] They operate with the reasoning capability that previous generations of automation lacked: evaluating exceptions, applying contextual rules, and escalating only what genuinely requires human judgment. Unlike RPA, which executes predetermined rules within a single step, agentic systems can reason across the boundaries between steps. The mechanism operates at three levels:

— *Verification:* The system confirms that an upstream step's output meets the downstream step's input requirements: data completeness, format consistency, and threshold compliance.

— *Reconciliation:* The system resolves discrepancies between competing constraints: demand against capacity, budget against timeline, specification against availability.

— *Routing:* The system directs work to the appropriate next step based on contextual evaluation, escalating only those exceptions that exceed its confidence threshold.

Each function previously required a human coordinator who understood both sides of the interface. That individual's calendar, availability, and cognitive load determined the interface's latency. When a machine performs these functions, the latency collapses from days to hours or minutes.

Consider how such a system might transform the Monday morning experience of a demand planner at a manufacturer. In the past, this executive might arrive at work to find a screen showing last week's forecast, awaiting manual review. That's no longer the case. Instead, the system has already ingested new demand signals, reconciled them against current capacity and inventory, flagged three exceptions that exceed its confidence thresholds, and staged a revised production schedule for the remaining 94 percent of items.

The planner's role is no longer to review every forecast and approve routine decisions. It is to investigate the three anomalies that require judgment. Rather than spending four hours reconciling routine demand against capacity constraints, the planner spends that time on the decisions that genuinely require human expertise, such as responding to a customer requesting an unusual delivery pattern, managing a raw-material shortage across product lines, or evaluating the implications of a new product launch with little historical data.

The routine coordination happened overnight, autonomously, at the interfaces. The human worker arrived to find a curated set of decisions that only a human should make.

[10] Erik Brynjolfsson, Danielle Li, and Lindsey Raymond, "Generative AI at work," *Quarterly Journal of Economics*, February 2024, Volume 140, Number 2. The task-chaining hypothesis holds that AI's primary value-creation mechanism is not task-level productivity improvement but the connection of previously discrete human-mediated tasks into continuous automated flows.

[11] Longju Bai et al., "How do AI agents spend your money?," arXiv, April 29, 2026, arXiv:2604.22750. Note that aggregate AI operating costs, including token costs, are an ongoing constraint: McKinsey's 2026 State of AI survey reports that approximately one in five organizations are constrained by them.

What an interface redesign produces

The results are operating leverage, not a reduction in head count. Freeport-McMoRan raised mill throughput by 5 to 10 percent from the same asset, with no new capital deployed. Toyota did not reduce its workforce when AI compressed resource allocation from weeks to minutes; its planners were redeployed rather than released.[12] Google Cloud, which reports more than 140,000 companies building enterprise AI deployments on its platform, describes the same pattern: "Almost none of our clients have let anyone go," Thomas Kurian, CEO of Google Cloud, said in an October 2025 interview with *India Today* magazine.[13]

Rather than eliminating people, reducing the coordination tax gives employees back the time they currently lose to coordination overhead. The planner who no longer spends four hours reconciling a demand signal against capacity constraints can spend those hours evaluating scenarios, managing exceptions, and advising the business on trade-offs. An engineer who no longer waits three days for cross-functional approval of a design change can iterate in hours. The procurement manager who no longer manually reconciles supplier confirmations with production schedules can focus on supplier relationships and risk management.

Over time, SG&A as a percentage of revenue declines because the same people produce more revenue, not because organizations employ fewer people overall. Roles change composition even where total employment holds: fewer people coordinating and more people deciding. The denominator grows faster than the numerator, which is the source of the operating leverage.

At the manufacturer we studied, for example, demand-to-production planning had often operated on a 30-day cycle spanning six functional domains and more than 75 planning tools, most of them spreadsheets and legacy systems. Regional sales teams submitted forecasts, central planning reconciled them against manufacturing capacity and supplier constraints, and allocation guides eventually made their way to production. More than 50 planners, each working within a largely siloed scope, spent ten nights or more during each cycle manually adjusting allocation and operating plans as conditions changed. The work performed within each step was relatively straightforward. Much of the delay was due to information crossing functional boundaries and planners reconciling competing constraints and conflicting KPIs.

The redesign team began by mapping coordination intensity across the workflow's six interfaces. It focused on the demand–supply reconciliation junction, where thousands of part-level forecasts were reconciled against hundreds of supplier capacity positions each week and where delays cascaded most severely downstream. After deconstructing the work performed at that interface, the team discovered that roughly 65 percent of the work consisted of routine verifications, such as checking demand against capacity bands, validating forecast changes against stability thresholds, and confirming part-level requirements against orderable configurations. Another 25 percent involved constrained reconciliation across competing functional KPIs: resolving capacity conflicts, responding to supplier gaps, and rebalancing allocations as conditions changed. Only about 10 percent represented genuine exceptions requiring human expertise and cross-functional negotiation.

[12] Eric Lamarre, Kate Smaje, Robert Levin, Alex Singla, and Alexander Sukharevsky, *Rewired: How Leading Companies Win with Technology and AI*, Wiley, April 2026.

[13] Unnati Gusain, "Google Cloud CEO Thomas Kurian says AI will not take your job, it is here to do this instead," *India Today*, October 12, 2025. The 140,000 figure is Google Cloud's own reported count of companies building enterprise AI deployments on its platform. Kurian's remark was made in the context of clients of Google Cloud's Customer Engagement Suite retaining their customer-service teams.

The team spent four months mapping and classifying these tasks and establishing the boundaries for AI autonomy before configuring any technology. Routine verification within agreed thresholds could proceed without human review; constrained reconciliations could be handled autonomously within established business rules and cost limits; genuine exceptions, along with cases where AI confidence fell below threshold, would be routed to a planner with the relevant context and alternatives already assembled.

The results were dramatic. Planning clock speed improved tenfold, from 30 days to three, and 80 percent of planning became touchless. Reaction time to supply shocks and demand opportunities fell from weeks to less than one day. The more than 50 planners who previously worked within siloed scopes gave way to roughly ten end-to-end planners who governed a system that processed routine decisions at machine speed, while more than 150 team members were upskilled into product management, data science, and AI engineering roles. The 75 spreadsheets and legacy tools gave way to a single platform, and the ten late nights of manually adjusting allocations became a five-minute optimizer run. The structural profit impact, validated by the finance function, reached billions within the first two years. That figure reflects the full planning transformation across the enterprise, including platform consolidation and a tenfold gain in clock speed, rather than the coordination tax on the single workflow measured above.

This manufacturer succeeded in large part because it understood that interface redesign cannot be delegated to existing functions. A finance team cannot redesign the planning-to-procurement interface alone because it owns only one side. A technology team cannot redesign it either, because the redesign is primarily organizational rather than technical.[14]

What is needed is a cross-functional team accountable for a specific workflow and empowered to redesign the interfaces within it. Its charter is to separate tasks from individuals and reassemble them into AI-mediated chains. In practice, this requires a COO-sponsored team with representation from each function the interface touches, accountable for a single workflow's end-to-end cycle time and reporting progress to the executive committee.

The process is sequential. Before AI can coordinate work at an interface, the team must identify the tasks currently performed there, determine which require human judgment and which can be automated, and redesign the flow before introducing technology. Organizations that deploy AI on top of existing coordination patterns add a new tool without removing the underlying source of delay.

Six priorities for addressing the coordination tax

Reimagining workflows to reduce inefficiencies at the handoffs requires leaders to take the following six initiatives:

— *Map interfaces, not just processes.* Make handoff points visible: what information crosses each interface, what verification occurs, how long it takes, and what it costs.

— *Start where interface density meets core competence.* The highest-value opportunities combine frequent coordination, moderate judgment requirements, and activities central to the business.

[14] "The AI transformation manifesto," McKinsey, April 7, 2026, which finds that leading companies can reach breakeven on tech and AI transformations in one to two years.

Find more content like this on the
McKinsey Insights App
Scan • Download • Personalize

— *Design for verification, not generation.* The value of AI at interfaces lies in verifying and reconciling competing constraints at machine speed.

— *Invest primarily in people and processes.* In our experience, successful redesign efforts devote roughly 70 percent of their effort to organizational redesign rather than technology alone.

— *Govern by interface quality.* Define escalation thresholds, accountability, and the boundaries between human and machine judgment.

— *Treat delay as a cost.* In our experience, every quarter spent postponing interface redesign increases the cost of catching up as competitors accumulate coordination data and operating experience.

Interface redesign typically takes six to 12 months, with roughly 70 percent of the investment going to people and process redesign, 20 percent to technology and data integration, and 10 percent to AI model development. Where the coordination tax is high, organizations can expect measurable payback within one to two years, consistent with the breakeven that leading companies report on tech and AI transformations more broadly.

The next frontier of workflow optimization

Every major technology wave lowers the cost of something important. Steam reduced the cost of power. Electricity reduced the cost of light and motion. The internet reduced the cost of information. AI is reducing the cost of coordination.

For more than a century, organizations designed themselves around the assumption that coordinating work across functions is expensive, slow, and constrained by human bandwidth. That assumption justified layers, meetings, review cycles, and approval chains. It was not incorrect. It was simply the constraint around which organizations were built.

Agentic AI changes the economics of that constraint. Organizations can now redesign workflows around a dramatically lower cost of coordination—not by improving the work performed within each step, but by reducing the friction between them.

Organizations that eliminate their coordination tax will carry a cost structure that their competitors cannot easily match. The advantage will compound because AI-mediated interfaces improve with every coordination event. Every transaction that passes through a redesigned interface generates information that makes the next one faster and more accurate.

It's clear that organizations need to redesign workflows. The question is where to begin. For four decades, organizations have looked first at the boxes on the process map. The next generation of redesign will begin with the arrows.

Ali Sankur is a partner in McKinsey's Chicago office, **Bernhard Mühlreiter** is a senior partner in the Vienna office, and **Steffen Fuchs** is a senior partner in the Dallas office.

This article was edited by Larry Kanter, a senior editor in the New York office.

Designed by McKinsey Global Publishing
Copyright © 2026 McKinsey & Company. All rights reserved.