SAFWAN ALMODHAYAN



OPINION AND ANALYSIS

Editorial date: October 29, 2026

The Speed That Stopped at the Desk

> A company can buy a better capacity for writing and analysis, only to discover that what needs redesigning is how work moves between people. What do the experiments tell us when we follow the outcome beyond the employee's screen?

Between September 2023 and October 2024, 66 companies took part in an experiment involving Microsoft 365 Copilot. Access to the tool was randomly assigned to employees, and the sample included 7,137 knowledge workers. In the second half of the planned six-month experiment, being given access reduced time spent on email by about 1.4 hours a week; the local estimate of the effect of actual use among employees who responded to access was about two hours.1

But the researchers also say: "We do not observe the content of any work, nor do we have measures of output or performance evaluations." The experiment therefore does not establish that companies completed more work or that their customers waited less.1

That qualification is where the problem begins. We have evidence that some of what an individual does has changed, and a different question about the result produced by the organization. The distance between the two is not a detail that can be removed from a financial presentation.

Customers do not buy employee speed; they buy the completion of what they came for.

An email may be written in a minute, then sit for days awaiting information that no one was assigned to produce. An analysis may become easier to prepare while its acceptance criteria remain obscure. The speed is real, but the route that would turn it into a service is still under

construction.

What Is the Customer Actually Buying?

In a controlled experiment published by Shakked Noy and Whitney Zhang in *Science* in 2023, 453 professionals completed writing tasks related to their occupations. Access to ChatGPT reduced average task time by 40 percent and raised output quality by 18 percent. This is an important result about specified tasks, not an estimate of higher profits at the organizations where the participants worked.2

Confusion begins when the unit of inference expands faster than the unit of measurement. We test the writing of a memo, then speak about a department's productivity. We measure a department, then assign a percentage to the company. Eventually, helped by polished slides, a number born inside a short task becomes qualified to govern an annual budget.

Consider an engineering-services company preparing a technical and commercial proposal. One employee summarizes the client's needs, an engineer estimates the scope of work, a colleague checks team availability, and the proposal is then approved. An assistant can improve the draft, but the proposal is not ready to send until these parts agree. This is a hypothetical case illustrating task dependence, not an account of a particular company.

The evidence also offers a successful case that should not be concealed. In a study by Erik Brynjolfsson, Danielle Li, and Lindsey Raymond, published in 2025, the estimated increase in issues resolved per hour was about 15 percent in a sample of 5,172 customer-support employees. The analysis used the assistant's staggered introduction; deployment was not a comprehensive randomized experiment.3

One feature of that setting matters: employees largely handled cases individually. The personal task came close to the unit of completed service. The following is an interpretation. When an employee can complete what the customer needs within their own work, the benefit may travel quickly. When the employee supplies one part that must wait for others, the relationship among those parts becomes an economic factor no less important than the efficiency of any one part.

Knowledge Does Not Work Alone

Return to the hypothetical company. The engineer produces a more accurate scope description, but sales relies on a different version of the client's requirements. The financial estimate arrives early, yet the implementation schedule is updated only at a weekly meeting. Every document is better; the documents are not necessarily aligned.

The problem is that improving one element may require another before the improvement can bear fruit. Economists call this relationship organizational complementarities. In their 1990 study of modern manufacturing, Paul Milgrom and John Roberts examined how choices in technology, production, and organization reinforce one another, so that the value of one choice changes when the surrounding choices change.4

The idea is simpler than the term. If a proposal can be revised as soon as the client's requirements change, access to a shared version of those requirements becomes more valuable. If every department uses that shared version, rapid revision becomes more worthwhile. Buying one capability while leaving the other untouched may yield less than combining them.

In firm data studied by Timothy Bresnahan, Brynjolfsson, and Lorin Hitt, the evidence was consistent with complementarity among information technology, workplace organization, and employee skills. The study was published in 2002. It is observational evidence, not an experiment that permits every improvement to be attributed to reorganization alone.5

What this literature adds to today's discussion is a change in where the question is placed. It is not enough to ask whether the model can perform the task. We must also ask what has to be true elsewhere in the process for us to benefit from that performance.

The answer may be modest: a common definition of "complete scope," inputs accepted by both departments, or the ability to update information when it changes instead of waiting for another round of correspondence. Not every problem calls for wholesale restructuring. Sometimes the project a company lacks is a precise one-page agreement, which is why it is difficult to display at a launch event.

The idea also has limits. If the process is already integrated and the augmented task is the real constraint, the tool may produce a clear return without broad change. Complementarities are not an excuse to attach a consulting program to every software licence. They are an invitation to identify what depends on what.

What Lies Between Two Steps May Be Worth More Than the Steps

A task may be finished in an employee's account before its result becomes usable by a colleague. The completed file needs sufficient inputs and an agreed acceptance standard before the next step can begin. That is an operating cost that a writing-speed test alone cannot reveal.

In a 2026 working paper titled “Chaining Tasks, Redefining Work,” Mert Demirer and his co-authors model production as linked steps with a Leontief structure: the result requires complementary components, and an excess of one cannot compensate for the absence of another. The choice among human work, assistance, and automation is shaped by verification and coordination costs and by the order of steps, not only by how many steps can be automated.6

In the engineering example, imagine that the tool extracts the client’s requirements, connects them to items in the scope, and produces a list of assumptions and open questions. The engineer can review one connected package. Compare that with departments producing a summary, a proposal, and a financial estimate separately, in different formats and without a single reference version. The quality of each output might look similar when judged independently, while the cost of verifying the final result differs. The difference does not arise from the number of departments alone, but from whether their outputs agree and can be traced to the source.

Connecting steps does not mean leaving them unsupervised. A verification point can be placed at the end of a package, with internal checks retained where errors have serious consequences. The cost of checking one long result may also exceed the cost of reviewing shorter ones. “End to end” is therefore not, by itself, an economic argument.

The paper is a theoretical model with empirical patterns consistent with its predictions; it does not certify that linking steps raises any company’s profit. Our practical inference is narrower: an evaluation of the tool must include the cost of inserting its output into the work that follows.

This yields a testable engineering requirement. The output should carry the data its recipient needs, its sources, what remains unresolved, and a clear completion standard. If the next department must call the file’s owner and ask for a fresh explanation of everything that happened, the organization has not automated the transfer of knowledge. It has improved the appearance of the message announcing that the transfer is still needed.

Some of this work requires software changes; some requires a change in the agreement among the people who own the process. An application programming interface can move data quickly. Agreement that the data are sufficient to begin the next step requires the work itself to be designed.

Can the Tool Repair Coordination by Itself?

The strongest objection to this analysis is that AI may reduce some of the need for coordination in the first place. If an engineer can understand market requirements, or a commercial specialist can explore technical constraints, the relationship between disciplines

may change from within the task.

An experiment at Procter & Gamble supports this possibility. Its peer-reviewed study was published in *Organization Science* in 2026. The workshop was completed by 791 participants from technical and commercial functions, randomly assigned to work alone or in pairs, with or without AI.7

The quality of proposals by individuals using the tool was comparable to that of teams without it. AI-assisted proposals were also more balanced between technical and commercial considerations. But the experiment studied early product-development work in a short virtual workshop; it did not establish a shorter product-to-market cycle or higher sales.7

The finding deserves to be taken seriously. Not every functional boundary is a fixed wall. A tool may help an employee see how a proposal affects a colleague's work before sending it, reducing some misunderstandings that would otherwise require a later meeting. That is a possible mechanism inferred from the result, not a direct measurement of every meeting in a company.

Knowledge of constraints, however, does not automatically grant authority to commit to them. An engineer may understand pricing without making a proposed price approved. A commercial specialist may anticipate a manufacturing problem without making factory capacity available. Integrating knowledge and completing mutual commitments remain different things.

The more precise claim is therefore conditional: a tool can improve collaboration, and an organization can prevent that improvement from reaching execution if the tool's outputs remain unacceptable elsewhere in the process. The location of the barrier must be tested, rather than assumed always to be technical or always managerial.

When the Place to Search for Value Changes

Another 2026 working paper offers a complementary perspective. Hyunjin Kim, Dahyeon Kim, and Rembrand Koning studied 515 start-ups. According to descriptions by the researchers and the research institute, participants received technical resources and training. One randomly assigned group then received case studies about reorganizing production around AI, while the other received general entrepreneurial material.8

The reported results included greater discovery of uses and improvements in some performance indicators. What could be checked for this article, however, was the researchers' summary and the institute's account; the tables in the full manuscript were not reviewed. We therefore do not rely on widely repeated revenue multiples, or carry the finding from start-ups to large organizations as if it were a settled law.8

What merits attention is the subject of the intervention: teaching an organization where to look for value. It may possess the same tools while confining its imagination to the work visible on a screen. It improves proposal writing but does not ask why the proposal needs three rounds to define the same thing. It searches for a use inside every function and misses the possibility of recomposing the work between functions.

In the hypothetical engineering company, the larger improvement might come from receiving requirements in a form that allows scope, resources, and cost to be estimated together. What reaches the employee changes. We no longer ask someone to handle an incomplete file faster; we make it possible to complete the file before it travels to the next department.

This is a proposed strategic conclusion from the evidence taken together: choosing where AI enters a process may matter more than distributing the tool evenly across every seat. Equal access to licences does not create equal opportunities to produce value.

The Employee Who Reviews and the Customer Who Waits

The problem has a human dimension that text-quality tests do not show. In our hypothetical example, the proposal writer can generate many alternatives while a colleague remains responsible for reviewing them. If the handoff does not change, the constraint may simply move to the reviewer. We cannot then infer an organizational improvement from measuring the writer's work alone.

Nor should reviewers be accused of resisting change merely because they need sufficient evidence. The new instruction to them may be clear: review more. The more important operating question is whether what they review has become easier to inspect, or merely more abundant.

At the other end, a small client may have activities tied to the date of a proposal: choosing a provider, arranging finance, and fixing a start date. These are economic possibilities in our example, not losses measured in a sample. Their value is to remind us that an organization's elapsed time enters the decisions of people outside it.

In a possible Saudi application, a company or public service provider could choose one recurring transaction and follow it from the first complete request to the result received by the beneficiary. This is a proposal for a local test, not a description of an existing program. The endpoint must be explicit: an executable proposal, a completed service, or a reliable answer on which its recipient can act.

A more exact financial question then replaces the counting of licences. Did accepted output increase with the same resources? Did its quality improve? Did total time or rework decline? Higher quality may justify the investment even if volume is unchanged. Higher volume may

fail to do so if it creates expensive corrections later.

We do not need to worship throughput instead of worshipping minutes. We need a specified outcome, a quality standard, a total cost, and a beneficiary from whose position the improvement can be explained.

Four Versions of the Process Before One Is Scaled

This discussion can be turned into a falsifiable organizational experiment. Instead of comparing enthusiasts with their colleagues, similar units of work can be randomly assigned to four groups in a factorial design that separates the effect of the technology from that of a specified process change.

The following table is a proposed design, not the result of an experiment that has taken place:

Group	Access to an AI tool	Change to the way of working
00	No	No
10	Yes	No
01	No	Yes
11	Yes	Yes

The change is defined in advance. In the engineering-proposal example, it might consist of a shared requirements file, a common acceptance criterion at handoff, and an update of resource availability before the commercial version is approved. The amount of training and managerial attention should be controlled so that their entire effect is not attributed to the tool.

One possible outcome is the number of completed proposals meeting the quality standard per period, provided demand is sufficient. If demand is limited, total elapsed time or quality may be more suitable than volume. Review and rework costs must also be observed, so that the measurement window does not close before corrections arrive.

Define Y so that a higher value means a better outcome. If the measure is completion time, use its negative or reverse the interpretation of the effect's sign. The interaction can be estimated from group means:

$$\Delta = Y_{11} - Y_{10} - Y_{01} + Y_{00}$$

When Δ is positive with a reliable estimate, combining the tool and the specified process change produces an increase beyond the sum of their separate increases, on the chosen outcome measure. This is a test of complementarity in that setting, not a law for every organization.

The experiment may yield an uncomfortable result: the process change may suffice without the tool, the tool may work alone, or the package may be costly without adding enough to justify itself. Those results must be permitted before the project begins; otherwise the evaluation is only a delayed approval ceremony.

The unit of randomization requires care. If groups use the same reviewers or exchange files, one intervention may affect another group's results. The causal-inference literature calls this interference and distinguishes direct from indirect effects. Teams or workstreams should therefore be chosen to limit spillover, or the spillover should be modeled explicitly.10

A large number of transactions is not sufficient if there are few independent units. Statistical power must be estimated, differences in case difficulty and changes in demand considered, and a period chosen that reveals later outcomes. When randomization is impossible, a controlled time comparison may still be useful, but its causal force is weaker and it should not be presented as an equivalent experiment.

What Should Remain After the Pilot?

Building this capability may take time. In "The Productivity J-Curve," published in 2021, Erik Brynjolfsson, Daniel Rock, and Chad Syverson discuss how general-purpose technologies require complementary intangible investment and how their effect may take time to appear in measured productivity.9

But the curve does not promise every project eventual success. We should be able to name the asset being built: documented knowledge of task dependencies, a repeatable verification method, alignment between inputs and outputs, or the ability to redistribute work. If nothing inspectable is accumulating, the mere passage of time is not organizational investment.

This changes the decision budget. Resources go to preparing data, designing handoffs, training recipients, and measuring the outcome, as well as paying for use. A review point is established at which the organization can reduce or alter the project on evidence from its own service.

The hypothesis worth testing is clear: where task gains are large and dependence on other departments is dense, better handoffs may explain an important part of the difference between a successful demonstration and successful operation. If the specified changes do not

improve the outcome, with quality held constant and measurement adequate, we should revise the hypothesis or diagnose the constraint—not accuse reality of failing to understand the transformation.

We return to the email screen where we began. We should neither make the saved minutes prove more than they do nor dismiss their value to the employee. But an organizational decision must trace what happened after the message was closed. Did the next step become possible? Did the work reach an endpoint the customer could use?

The message may have arrived this morning. What has not yet arrived is agreement on what it means for the work to be complete.

References

1. Dillon et al., Shifting Work Patterns with Generative AI, version 4 record dated November 13, 2025; AER: Insights: forthcoming. Randomized experiment involving Microsoft researchers; event and measurement details appear in the text.
2. Shakked Noy and Whitney Zhang, Experimental evidence on the productivity effects of generative artificial intelligence, *Science*; published online July 13, 2023, issue dated July 14. Randomized controlled experiment supporting improvement in writing tasks. PubMed record.
3. Erik Brynjolfsson, Danielle Li, and Lindsey Raymond, Generative AI at Work, *QJE* 140(2): 889–942; published online February 4, 2025, May issue. Peer-reviewed study of staggered introduction; supports the customer-service issue-resolution result, not company-wide profit.
4. Paul Milgrom and John Roberts, The Economics of Modern Manufacturing: Technology, Strategy, and Organization, *American Economic Review* 80(3): 511–528; June 1990. Peer-reviewed theoretical study; foundation for complementarity among production and organizational choices.
5. Timothy F. Bresnahan et al., Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence, *QJE* 117(1): 339–376; February 2002. Peer-reviewed observational study; evidence of complementarity, with causal-identification qualifications.
6. Mert Demirer et al., Chaining Tasks, Redefining Work: A Theory of AI Automation. Working paper; first version February 22, 2026, updated text September 28, version 2 deposited October 2, 2026. Supports the role of step order and handoff costs; not an experiment on organizational returns.
7. Fabrizio Dell'Acqua et al., The Cybernetic Teammate: A Field Experiment on Generative AI and Teamwork, *Organization Science* 37(4): 1217–1242; published online June 12, 2026. Peer-reviewed experiment conducted May–July 2024; supports knowledge integration in early-stage innovation.
8. Hyunjin Kim, Dahyeon Kim, and Rembrand Koning, Mapping AI into Production: A Field Experiment on Firm Performance, working paper, April 2026; and the institutional account of the design, AI Institute at Harvard Business School, April 7, 2026. The original summary and account were used; the full manuscript tables could not be inspected, so revenue multiples were omitted.
9. Erik Brynjolfsson, Daniel Rock, and Chad Syverson, The Productivity J-Curve: How Intangibles Complement General Purpose Technologies, *AEJ: Macroeconomics* 13(1): 333–372; January 2021. Peer-reviewed modeling and measurement study; explains the role of complementary investment but does not guarantee that any particular project will recover.
10. Michael G. Hudgens and M. Elizabeth Halloran, Toward Causal Inference With Interference, *Journal of the American Statistical Association* 103(482): 832–842; June 2008. Peer-reviewed methodological study supporting attention to treatment spillovers among units.