How Copilot Changed the Pace of Work in Word — And How We Measured It

By Gaurav Verma, Siddharth Suri, Scott Counts

June 18, 2026

Read the article online · Read the methodology paper →

Key takeaways

- Word tasks that took similar coworkers about eleven minutes took Copilot adopters about six — roughly 40 to 45 percent faster after twelve weeks.
- The change carried over: even on Word tasks where adopters did not use Copilot, they worked faster than their matched peers.
- This measures how long activities took, not what they were worth. Faster is not automatically better.

How we know this — How to read this card

Evidence type	**Linked paper is methodology only** **Activity time is not output** Quasi-experimental
Design	72,186 people in 1,021 US organizations; matching and staggered difference-in-differences.
Limit	Activity time is not output quality or total productivity; role was uncontrolled.

Source / posture	Reader analysis · Jun 18, 2026 WorkflowView documents the method, not these estimates. Read methodology · Microsoft analysis; not peer-reviewed. **Use as a lens, not a mandate.**

Using data from more than 72,000 Word users, researchers compared sustained Copilot adopters with similar coworkers who adopted later, and found the pace of work in Word changed markedly after adoption. The study introduces a new way to measure AI impact as products and people evolve together.

Knowing whether an AI tool is actually helping people work is harder than it sounds. Adoption dashboards and license counts show who has access, not what changed for the people using it. The tool also keeps changing, the models keep changing, and people keep changing how they use them, so a single rollout study goes stale fast. What organizations need is a way to keep measuring impact after the experiment ends.

Gaurav Verma, Siddharth Suri, and Scott Counts take that problem into Word. Using everyday usage data from 72,186 people across 1,021 US organizations, they compare sustained Copilot adopters (People who used Copilot in Word at least weekly through the full 12-week window, rather than trying it once.) with near-identical coworkers at the same organization who had Copilot available but hadn't started using it. They ask a narrower question, and answer it without running a new experiment. Every click, edit, and formatting change leaves a signal. Grouped together, those signals reveal the rhythms of the document lifecycle and provide the foundation for understanding how Copilot changes the pace of work within Word. What happened to measured activity time (The time telemetry recorded for a stretch of Word work like editing or formatting. It measures observed activity, not the quality or amount of what gets produced.) in Word once people started using Copilot consistently?

01

PART ONE

Before you can estimate impact, you need a unit of work.

Raw telemetry knows that someone opened a file, clicked a ribbon item, changed a style, accepted a suggestion, or typed text. What it doesn't know is what kind of work that was, whether the user was editing, creating, consolidating, etc. To bridge that gap, the team built WorkflowView (A method the researchers built that reads raw Word telemetry (clicks, keystrokes,

saves) and groups it into recognizable kinds of work, like editing or formatting.), which sorts those raw signals into recognizable kinds of work, such as active editing, refining layout, formatting text, and managing versions. Each bounded stretch of one such kind of work is an activity instance (A single, self-contained chunk of one kind of Word work, such as one run of formatting before the person moves on.). Here the team uses LLMs to abstract messy click-streams into plain-language activities across wildly different domains. It is a general-purpose translation layer from raw behavior to meaning.

Much of the discussion around AI productivity focuses on closed-loop tasks, the kind with a clear beginning, end, and measurable output, such as resolving a customer service ticket or completing a coding assignment. But for many knowledge workers, those tasks are the exception rather than the rule. Creative, marketing, strategy, and communication work is often open-loop. Ideas evolve, goals shift, and there is no obvious finish line. In those settings, measuring value becomes less about counting tasks completed and more about understanding how AI changes the pace, flow, and effort of work.

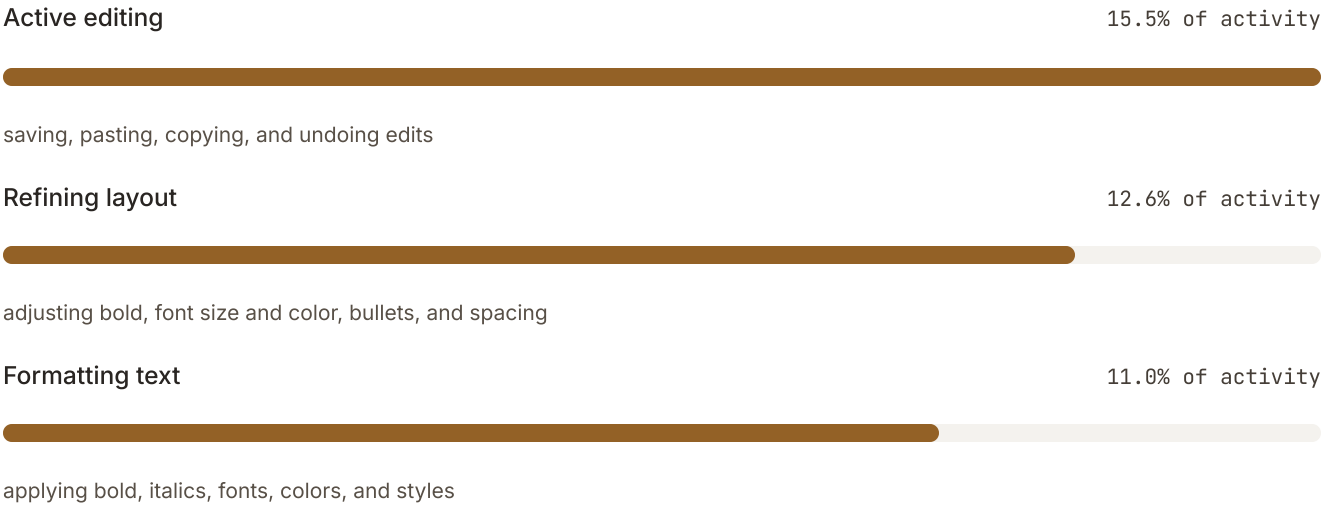
WorkflowView lets the researchers compare like with like. Rather than guess at productivity from total time in the app, they measure how long each kind of Word work activity actually takes, then watch how those times change once Copilot enters the picture.

FIGURE 1

The anatomy of work in Word

Ranked list of kinds of Word work by how often they occur. Active editing is most common at 15.5 percent of all detected activity, then refining layout at 12.6 percent and formatting text at 11 percent.

Each row is a kind of work people do in Word; the bar shows its share of all detected activity, and the grey text describes the actions involved. The top three make up nearly two-fifths of everything. The most frequent work in Word is editing, refining, formatting, and revising existing documents, not creating them from scratch. Understanding that activity mix is the first step toward measuring how AI changes the pace of work within Word.



Add-ins and automation

8.8% of activity



launching add-ins and automation

Version management

8.7% of activity



saving new versions of a document

New documents

6.2% of activity



creating and starting new documents

Content transfer

5.2% of activity



moving content with copy, cut, and paste

Themes and styles

5.1% of activity



applying themes, styles, and highlighting

Drafts and templates

5.1% of activity



starting drafts and using templates

Document management

3.8% of activity



opening, saving, and closing files

Reviewing comments

3.4% of activity



viewing and selecting comments

Real-time coauthoring

2.6% of activity



co-editing with others in real time

Final edits

2.0% of activity



accepting changes, reviewing, and printing

Collaboration

1.7% of activity



sharing and presence with collaborators

AI features 1.6% of activity



AI rewrite and text-prediction suggestions

Reconsideration 1.5% of activity



undoing, redoing, and reworking edits

Source: Verma, Suri, and Counts (2026), chart-data workbook. Shares are of all detected Word activity instances.

To keep it honest, each Copilot user is measured against a near-identical coworker who simply hadn't started using the tool yet.

WHY THE COMPARISON GROUP MATTERS

02

PART TWO

To measure change, you need the right comparison group.

Early adopters are rarely a random bunch. They may be more motivated, more technical, or just more willing to change their routine. Comparing them to everyone else would tangle Copilot's effect together with the kind of person who reaches for it first. To keep the comparison fair, the study pairs each Copilot user with a coworker at the same organization who had been working in Word much the same way beforehand, but hadn't started using Copilot during those 12 weeks. Those coworkers did go on to adopt soon after, so the two groups are similar people, separated mainly by *when* they started.

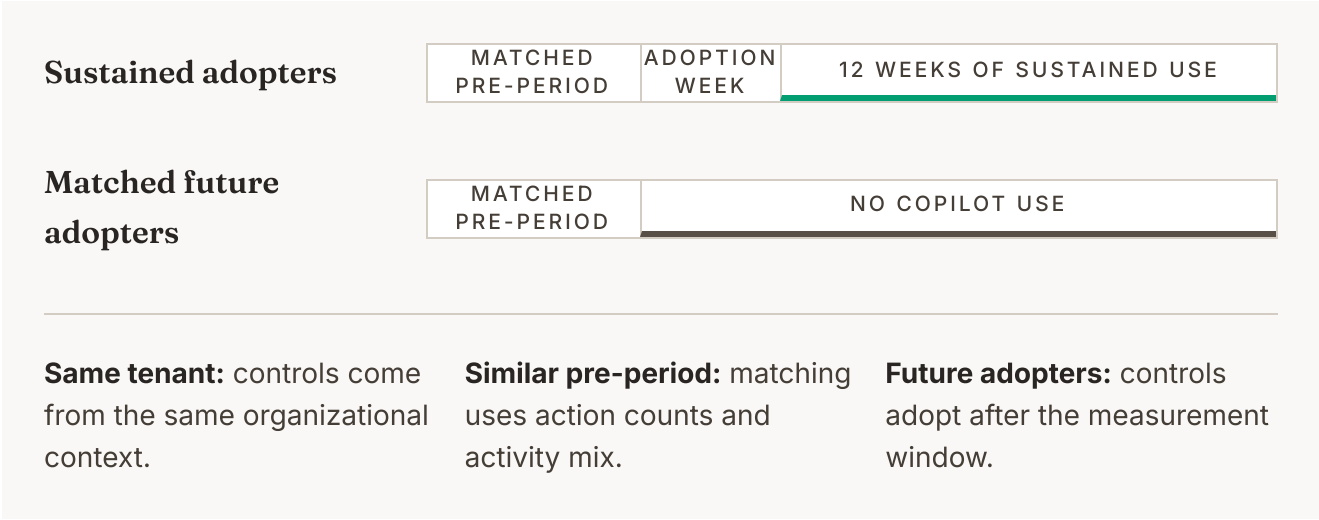
No one was randomly assigned to use Copilot in this study. Instead, the study relies on a simpler idea: if two very similar groups were moving in the same direction before adoption, then a change that appears only after one group starts using Copilot is more likely to reflect the tool itself. The groups track remarkably closely before adoption, making the divergence that follows more meaningful, even if it can never be proven with absolute certainty.

The design accounts for timing, workload, and rollout where the data allow, but it can't rule out every difference between groups. Role, notably, isn't controlled for.

> **What the statistical design can claim**
>
> The analysis uses coarsened exact matching on pre-treatment action counts and activity mix, then estimates effects with a staggered difference-in-differences framework across non-overlapping adoption cohorts. Adoption timing enters through the cohort structure, measured from each group's first week of adoption. Workload is matched on the number of instances of each activity, and rollout on enabled state.

DESIGN

Earlier adopters are compared with similar users who adopt later



Same tenant: controls come from the same organizational context.

Similar pre-period: matching uses action counts and activity mix.

Future adopters: controls adopt after the measurement window.

03

PART THREE

What happened after Word users adopted Copilot?

The result is clear. Before adoption, future Copilot users and their matched peers worked at nearly the same pace. After adoption, their trajectories diverged and stayed that way. Twelve weeks later, people using Copilot were completing the same kinds of Word activities in about six minutes, compared with roughly eleven minutes for similar coworkers who had not yet adopted it, a reduction of roughly 40 to 45 percent.

This wasn't a one-off effect confined to a single group of users. The pattern appears across adoption cohorts and persists over time. And when researchers look specifically at the activities where Copilot was actually used, the difference is even larger. Those Copilot-assisted activities averaged about 4.6 minutes by the end of the measurement window, well under half the pace of their matched peers and faster still than the adopters' own work without Copilot. The broader figure reflects all Word activities performed by adopters, whether they used Copilot in a given task or not.

> **Why measured time is not productivity**
>
> The outcome is telemetry-observed duration per Word activity instance. It does not measure document quality, total output, off-app thinking, review time, or business value. The measure also excludes the interval between submitting a Copilot prompt and accepting its output, so the result is not explained by model speed alone. Those other outcomes matter; they are outside this estimate.

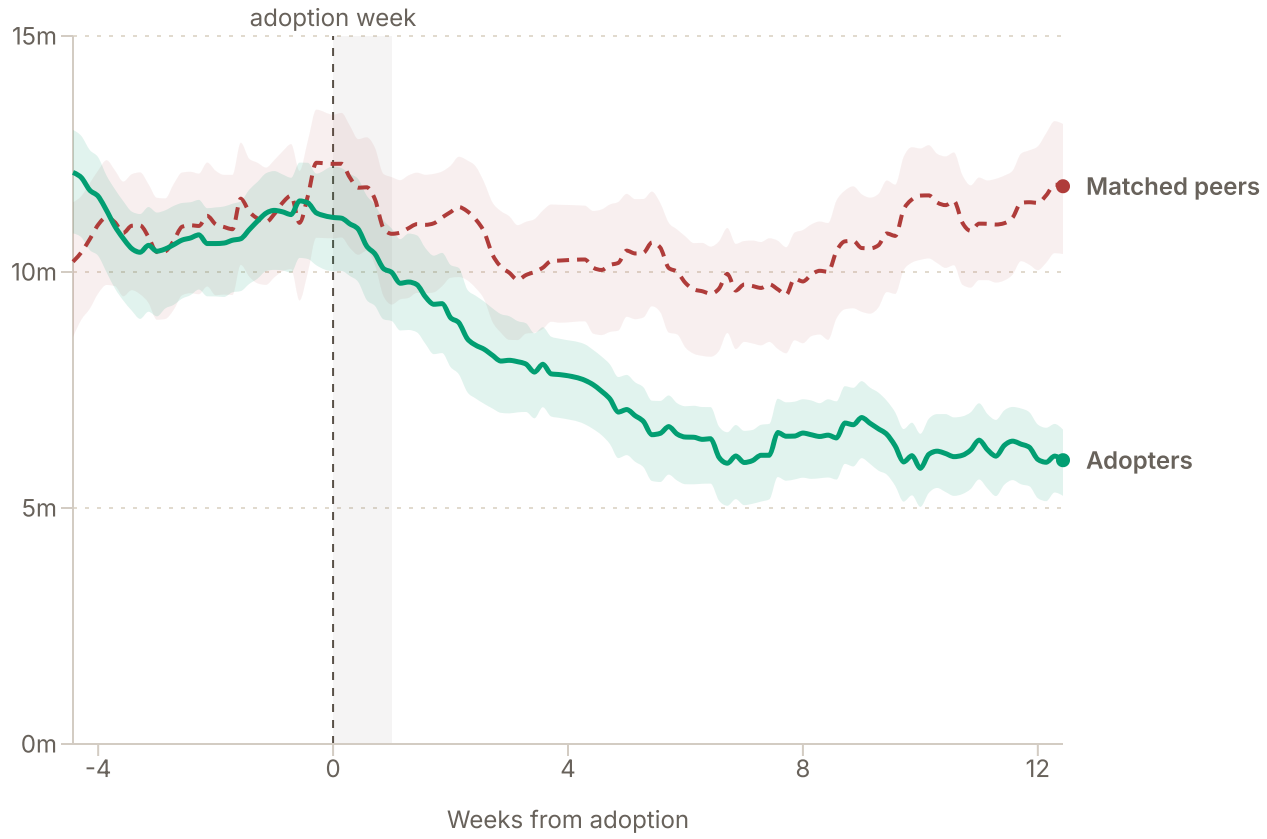
FIGURE 2

How Copilot changed the pace of work in Word

Difference-in-differences line chart. Matched future adopters and sustained adopters track together before adoption, then diverge. Sustained adopters fall to about six minutes per activity instance by week twelve while matched peers stay near eleven minutes.

Weighted time per Word activity instance, for sustained adopters and their matched future-adopter peers. The two groups move together until adoption, then separate.

ADOPTION COHORT



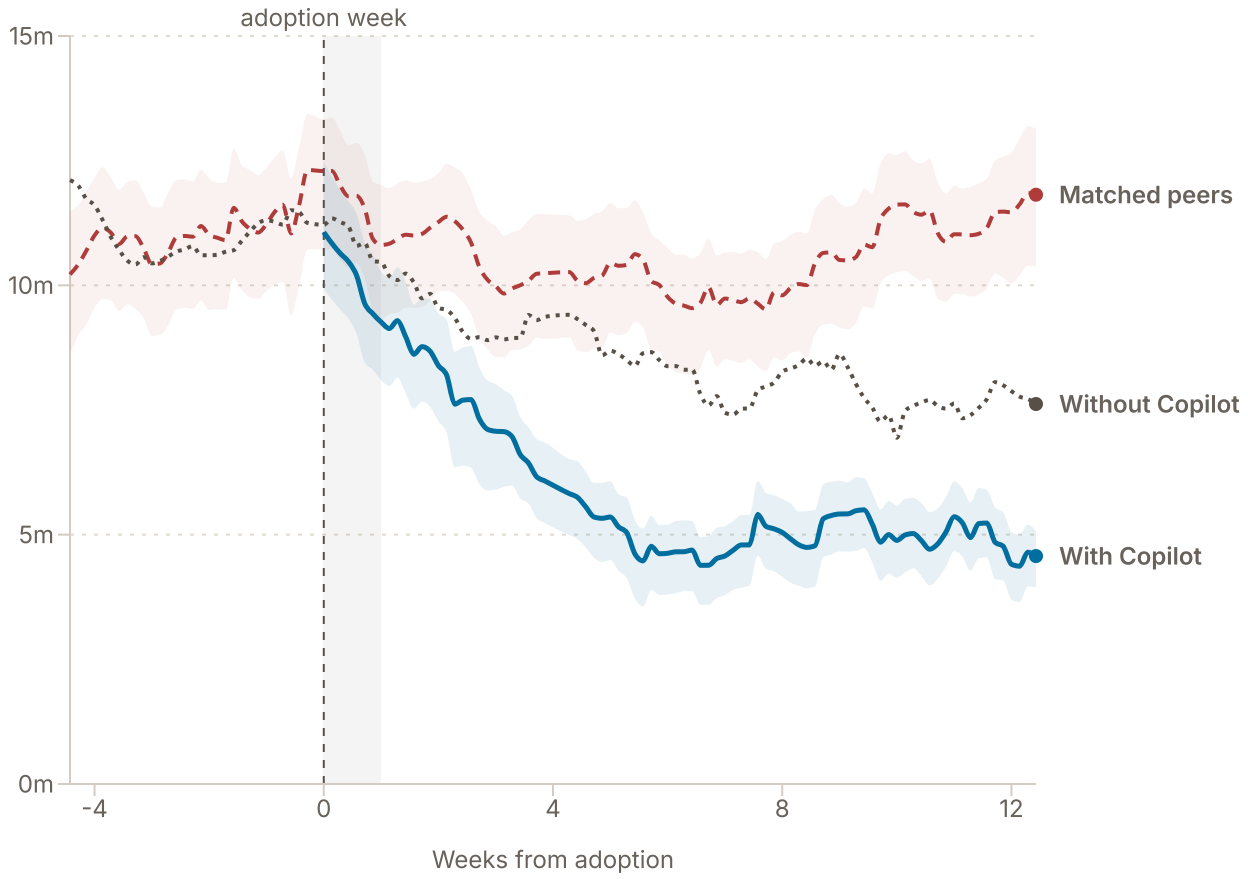
Source: Verma, Suri, and Counts (2026), weighted Feb/Nov cohort data. Shaded bands are 95% confidence intervals; measured time excludes the prompt-to-accept interval.

FIGURE 3

Even on tasks where Copilot wasn't used, adopters' pace of work accelerated

Three-line chart. Adopter instances using Copilot fall to about 4.6 minutes, and even adopter instances without Copilot settle near 7.6 minutes, both faster than matched peers, who stay near eleven minutes.

Where Copilot was used, activity times fell sharply. But even for Word activities where Copilot was not used, adopters worked faster than their matched peers. The gains spilled beyond the AI-assisted moments. Time per instance, with 95% confidence bands.



Source: Verma, Suri, and Counts (2026), split chart data. The with-Copilot series begins at the adoption week. Shaded bands are 95% confidence intervals.

04

PART FOUR

The gains weren’t limited to the easy work.

One obvious explanation for the result is that Copilot users may have been reserving the tool for simpler tasks. If that were true, the time savings would be less meaningful. It would suggest people were selecting easier work rather than completing the same work more efficiently.

The data points in a different direction. Activity instances with accepted Copilot output contain many more non-Copilot actions than the control average, and that gap holds across the whole window. In other words, the fastest activities were not necessarily the simplest ones. People were often moving through busy, multi-step workflows while using Copilot.

This measure is not a direct test of difficulty. More actions can reflect different working styles, and some benefits may carry over into activities where Copilot is not used directly. But taken together, the evidence argues against a simple “easy tasks” explanation. Copilot users were completing work faster while engaging in activities that appeared at least as involved as those performed by their peers.

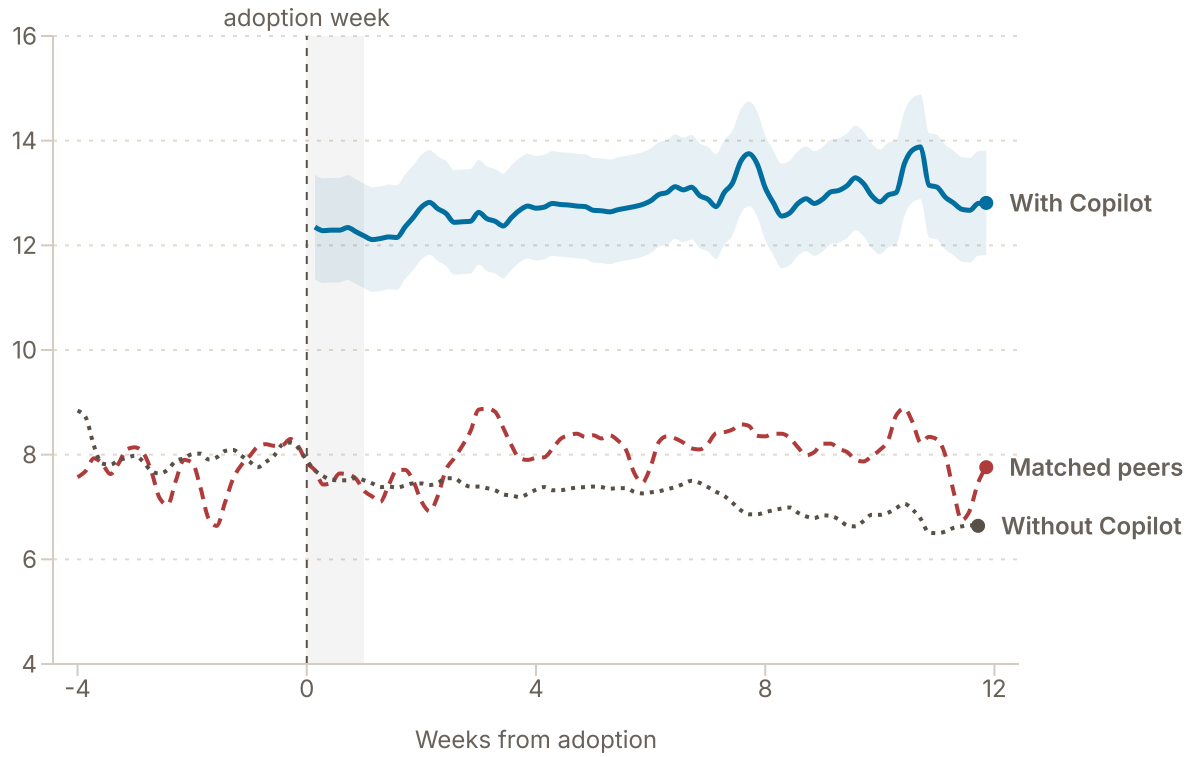
There is a second possible channel that current telemetry cannot yet isolate, where adopters carry AI-learned habits into work they do without Copilot. Adopters also seem to apply Copilot where it helps most, formatting a 20-row table with it, say, while handling a two-row one by hand. But the headline pattern argues against the easy-task explanation. Adopters were faster and working on busier instances at the same time.

FIGURE 4

Copilot was used on substantial work, not just quick wins

Three-line chart of unique non-Copilot actions per instance. Matched peers and adopter-without-Copilot instances sit near seven to eight actions. Adopter instances using Copilot sit near twelve to thirteen actions throughout the window.

Average number of distinct non-Copilot actions per activity instance. Tasks where Copilot appeared contained more non-Copilot actions on average, an imperfect but useful signal that people brought it into busier stretches of document work.



> Unique action count is a rough proxy for how busy a task is, not a validated measure of difficulty. It mainly tests the simplest easy-task explanation against the data.

Source: Verma, Suri, and Counts (2026), unique-action chart data. Copilot and AI-chat actions are excluded from the count.

> **How activities are compared**
>
> Comparisons are made at the level of matched activities. Each is the same kind of Word work, compared like-for-like between adopters and their peers, not across broad categories. Each instance is demarcated by its activity boundaries, so a given kind of work is always measured against the same kind of work. (We describe the concept here; the underlying signal is internal telemetry.)

05

PART FIVE

New adopters reached value sooner.

One of the study's most striking findings is how quickly the time savings appear. Employees who adopted Copilot later in the year reached their gains much faster than earlier cohorts. February 2025 adopters reached the gain in about 2.4 weeks. But by November, it had fallen to about 1.4 weeks, a roughly 40% compression of the ramp-up.

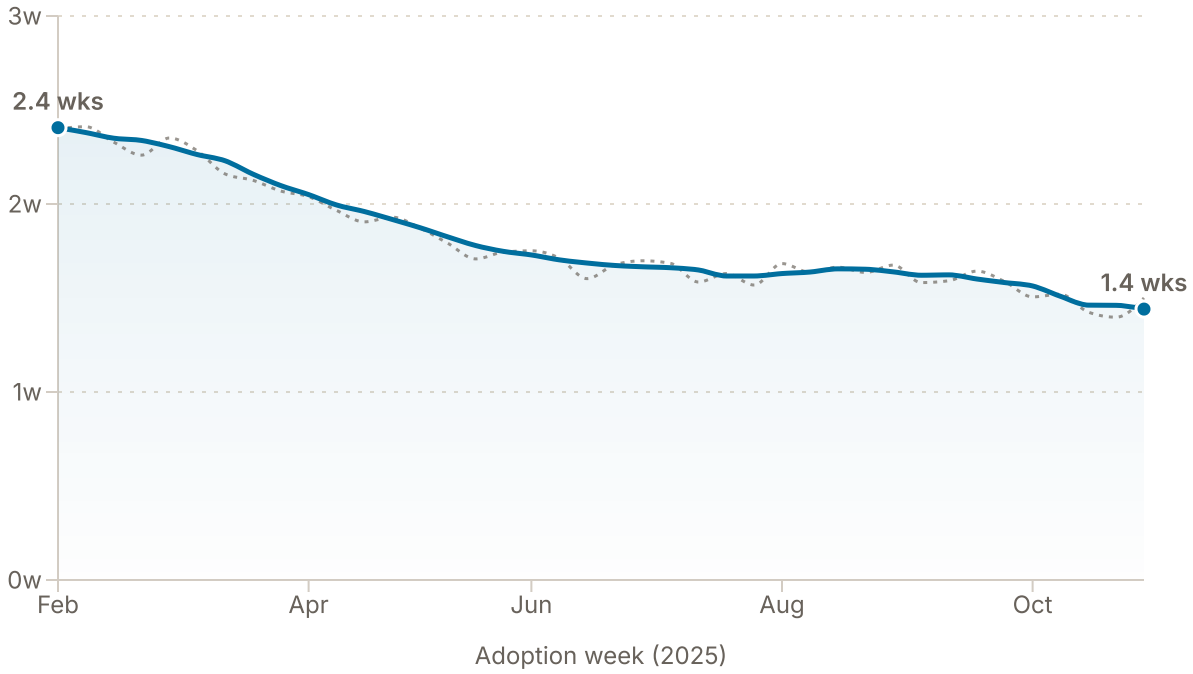
Multiple forces could be driving the shift: a better product, better onboarding, growing organizational know-how, or simply greater AI familiarity. The data cannot separate those effects. But it can show that the journey from adoption to value got shorter over the course of 2025.

FIGURE 5

Newer adopters reached their time savings faster than earlier ones

Line chart showing weeks to reach half of the eventual measured time reduction, by adoption week across 2025. It falls from about 2.4 weeks in February to about 1.4 weeks in November.

Weeks for each 2025 adoption cohort to reach half its eventual time savings. The bold line is the smoothed trend; the faint line is the weekly estimate.



Source: Verma, Suri, and Counts (2026), half-life chart data. The trend across cohorts reflects some mix of product improvement and growing familiarity that this data cannot fully separate.

How this line is built

Each point is a 2025 adoption cohort. For each, we fit a curve to its measured time-reduction trajectory and read off the half-life (the week by which the cohort had reached half of its eventual savings), across roughly forty weekly cohorts from February to November. We don't show confidence intervals here. The half-life is a derived estimate, and a dependable interval for it isn't straightforward.

THE BOTTOM LINE

What does this all mean?

Three things this should change about how you think about Copilot, and what to do about it.

I. **II.** **III.**

AI's impact can be measured in the flow of everyday work

For years, organizations have faced a measurement problem. Traditional experiments are expensive, slow, and often outdated by the time results arrive. This study shows that it may be possible to estimate AI's impact continuously, using the digital traces people leave behind as they work. As AI products evolve, measurement will need to evolve with them.

Copilot changed the pace of work, even where it wasn't used

The pace of Word work changed after adoption. Sustained adopters completed comparable activities in less time. But the interesting part isn't that Copilot made Copilot-assisted tasks faster; it's that work accelerated even on activities where Copilot wasn't used, suggesting the effects of adoption reach beyond the individual AI-assisted interaction.

People, products, and organizations learn together

Later cohorts reached the same gains faster than earlier adopters. The study cannot tell us exactly why, but the pattern suggests AI adoption is not static. Products improve, organizations refine rollout practices, and workers become more familiar with new ways of working. The result is a moving target. The experience of adopting Copilot today may be different from the experience of adopting it six months ago.

AUTHORS

The team behind the analysis.

Gaurav Verma

Researcher in IDEAS Research at Microsoft, studying how telemetry can be repurposed to make work with AI more intuitive and measurable.

Profile LinkedIn

Siddharth Suri

Senior Principal Researcher at Microsoft Research, studying computational social science, work, and human behavior at scale.

Profile LinkedIn

Scott Counts

Senior Principal Researcher and Research Manager, leading the IDEAS Research team at Microsoft.

LinkedIn

With thanks to Sonia Jaffe for feedback on the causal design and framing, and to the Computational Social Science working group for discussions that sharpened the analysis and its presentation.

Findings by Verma, Suri & Counts (Microsoft), first reported here. The WorkflowView method they build on is described separately in Verma & Counts, *Abstracting Cross-Domain Action Sequences into Interpretable Workflows* (arXiv:2606.14654). Analysis covers Word activity among US tenants across 2025 and into early 2026.

AN INTERACTIVE READING