Shifting Work Patterns with Generative AI

Eleanor Wiske Dillon, Nicole Immorlica and Sonia Jaffe*

Microsoft Research

March 20, 2025

Abstract

We present early evidence from a large, multi-industry, randomized field experiment on how knowledge workers are using generative AI tools to change the way they work. We tracked the work activies of over 6,000 workers over 6 months while randomly giving half of them access to a new tool that integrates generative AI into the applications they already used to read and write emails, attend video meetings, and create documents. Throughout the 6 month study, over 40% of workers with access to this tool used each week. Workers who used the generative AI tool spent 3 fewer hours working on email each week (25% less), without reducing their overall responsiveness to emails. They also completed the documents they authored 7% faster. During this early adoption period, we find less evidence of behavioral changes that require coordination with colleagues, such as shifting time spent in meetings or taking responsibility for authoring more documents.

Recent advances in generative AI have opened new possibilities for using AI to assist with or automate a variety of tasks. Early studies have already shown that these technologies can substantially increase workers' productivity in targeted tasks (Brynjolfsson, Li and Raymond (2025), Peng et al. (2023), Noy and Zhang (2023), Dell'Acqua et al. (2023)). Over time, these advances have the potential to shift the nature of jobs (Eloundou et al., 2024) and perhaps the macroeconomy (Autor (2024), Acemoglu (2025)). However, the transition from the creation of new technologies to effective, widespread adoption in the workplace takes time (Brynjolfsson, Rock and Syverson, 2021). In lab experiments, participants are asked to focus on completing a single task and directed to use AI tools that are well-suited to that task. Outside the lab, workers must learn on their own which parts of their job benefit from the use of new tools and gradually change their habits of work. Larger productivity gains often require organizational changes to redesign processes and shift responsibilities across workers (Bresnahan, Brynjolfsson and Hitt, 2002).

In this paper, we present early evidence on how generative AI tools are changing the way knowledge workers do their jobs. Between October 2023 and October 2024, we partnered with 56 large firms to run a cross-industry field experiment to measure how access to an integrated generative AI tool, Microsoft 365 Copilot (hereafter, Copilot), changed work patterns. Each firm recruited workers for the study and we randomly assigned each worker to receive access to this new generative AI tool or continue working with their current technologies. Participants in the study span many occupations, but all participating workers do some kind of work that involves heavy use of the Microsoft Office tools into which Copilot was integrated (e.g., emailing, video meetings, chat, document creation). Firms maintained the random assignment for six months, giving treated workers sufficient time to explore

*Corresponding author: eldillon@microsoft.com. All authors were employed by Microsoft, the maker of 365 Copilot, at the time of the study. Microsoft reviewed this paper for privacy conterns, but authors had full discretion over the results and estimates presented. We thank the Microsoft Customer Research Program, especially Alexia Cambon, Sida Peng, Modern Work Marketing, and company partners for help carrying out this experiment, Abigail Atchison, Roman Basko, and Fabio Vera for superb data science support, Jack Cenatempo, Esther Plotnick, and Will Wang for excellent research assistance, and Rem Koning and Danielle Li for helpful comments.

the new tool and integrate it into their daily work.[1] We used highly detailed telemetry data collected by Microsoft products to track the behavior of the over 6,000 studied workers during the experiment and for several months before the introduction of the new tool. These telemetry measures capture the actions of workers, for example time spent reading emails or attending video meetings or time from creation to completion of documents. We do not observe the content of any work.

This study took place during the early rollout of Copilot, when firms were restricted to purchasing a small number of licenses[2]. This particular generative AI tool was new for all workers, and the modal worker in the study had no close coworkers with access to the tool. Firms varied in how much training in the new tool they provided to treated workers. These workers were then free to use or not use the tool during their regular work. Treated and control workers may have also used other generative AI tools during the study, for example ChatGPT had been publicly available for nearly a year by October 2023, but access to Copilot could make it easier for workers to use generative AI at work. Copilot integrates AI into applications these workers were already using regularly in their jobs, allowing some new uses, for example summarizing emails or searching for information within work content, and reducing the effort cost of using generative AI to create content within documents and emails.

We use the product telemetry data to both confirm that workers assigned to the treatment group received an Copilot license and to track what share of workers used Copilot tools in any given week. Patterns of Copilot usage are interesting in themselves. The usage data also allow us to estimate treatment effects that adjust for low usage or noncompliance with the randomized assignment using an instrumental variable approach. We present these estimates of the local average treatment effect (LATE) alongside intent-to-treat estimates based only on workers' assigned study arm.

On average, in each week of the study, 44% of the workers who were assigned to receive Copilot used it at least once. Usage rates varied widely across firms, from a low of 21% of treated study workers using Copilot each week to a high of over 75%. This variation is consistent with other surveys of AI adoption across firms (Bonney et al., 2024; McElheran et al., 2024), but notable within even the subset of large firms that were interested in being early adopters of this tool and participating in a study of its use. Industry and variations in workers' roles (proxied by pre-study worker behavior) explain some of these differences. Nonetheless, indicators for a worker's firm remain by far the most powerful predictor of usage in a horserace, suggesting an important role for firm-specific differences that we are unable to measure such as training or managerial practices.

We find clear evidence that workers who use Copilot are beginning to change their work patterns in meaningful ways. During this early adoption phase, we see larger changes in behaviors that workers can adjust independently and less movement in behaviors that require coordination with colleagues. Workers who use Copilot regularly spend nearly 3 hours less per week working on emails (1.4 hours less in the intent-to-treat estimates). This shift represents a 25% time savings from the 11.6 hours per week that the average studied worker spent on email in the pre-period. In contrast, while workers are most likely to use Copilot during video meetings, we find no changes for Copilot users in the time spent in meetings or the types of meetings attended. Similarly, we find only small and statistically insignificant increases in total time spent working in Word and the number of documents authored, indicating that workers have not shifted the nature of their work or taken on new responsibilities. However, workers who use Copilot complete the documents they author more quickly, particularly documents that include input from other colleagues which they complete 20% faster, and contribute edits to 10% more documents authored by their colleagues. We find few substantive differences in

[1] Additional details on the experiment are in Appendix A.

[2] Each company was able to purchase 300 licenses. In all cases, this represents only a small fraction of their total employees who used Microsoft's Office suite.

these patterns among workers with more close coworkers who also have access to Copilot or within firms with higher adoption rates, suggesting that larger shifts in responsibilities require time and broad institutional efforts, not just local team coordination.

This work joins a small set of early studies of generative AI in real workplaces, including Brynjolfsson, Li and Raymond (2025), Cui et al. (2024), Otis et al. (2023), and Toner-Rodgers (2024). These real-world studies complement a larger body of work understanding the impact of generative AI on controlled tasks in a lab setting.[3] These earlier studies all focus on the impact of AI tools on specific tasks or narrow categories of workers. Surveys indicate rapid adoption of generative AI at work across a much broader set of workers; Humlum and Vestergaard (2025) find 41% of Danish workers in exposed occupations used ChatGPT at work by the end of 2023 and Blandin, Bick and Deming (2024) find 24% of a representative sample of all U.S. workers used generative AI at work by August 2024. However, these surveys cannot measure the impact of generative AI use on work behavior. This study aims to bridge some of the gap between these two lines of research: measuring both take up and impacts of use for a generative AI tool on a broad set of workers, albeit at the cost of clear measures of worker productivity.

1 Patterns of Copilot Adoption

In this study, treated workers were given access to Copilot, but were not required or directed to use it in their work. The median treated worker who had access to Copilot used it at least once in 42% of the weeks we followed them (the mean was 45%).[4] Figure 1a shows the distribution of these individual usage rates. 2% of treated workers used Copilot every week, while 6% never used it at all. Figure 1b plots the share of treated workers using Copilot by weeks since their firm allocated the experimental licenses. Usage trailed off modestly after a peak of over 60% when workers were first gained access, presumably due to a combination of poeple wanting to try the new tool and the introductory training sessions offered by some firms. In the the final two months, overall usage held steady around 40%, for a mean usage rate of 44% of workers each week across the study.

Microsoft's 365 Copilot could be accessed through Outlook, Teams, Word, Excel, PowerPoint, and as a separate M365 Chat. Within each application, Copilot could be used for different tasks. We focus on three applications where we think generative AI has particularly strong complementarities with the typical tasks and where we have the potential to trace clear changes in worker behavior as a result of Copilot use. Figure 1b also shows usage rates over time separately for these three programs. Workers were most likely to use the Copilot in Teams, which can capture transcripts of video meetings and analyze them to generate meeting summaries and follow-up items or answer user queries like "what was discussed in the last 10 minutes?". The next most popular use was through Outlook, where Copilot can summarize email chains, review a worker's inbox and flag emails that require a response, answer questions about information contained in emails more flexibly than a keyword search, and generate draft emails in response to prompts. The Word Copilot can draft text from prompts and answer

[3] See, for example, Peng et al. (2023); Spatharioti et al. (2023); Vaithilingam, Zhang and Glassman (2022); Campero et al. (2022); Noy and Zhang (2023); Edelman, Ngwe and Peng (2024); Edelman et al. (2024), Dell'Acqua et al. (2023), Cambon et al. (2023)

[4] Note that the data reflect engagement with Copilot, not with generative AI tools overall. These adoption rates may not be indicative of broader or long-run trends. During this early adoption period workers had less familiarity with the capacities of generative AI and the tools themselves were less developed, which may have lowered usage. However, firms may have recruited workers for this study who were particularly well suited to being early and enthusiastic adopters. The mean weeks of Copilot use across workers differs slightly from the mean share of workers using Copilot in each event week because not all workers were followed for exactly six months. See Appendix A for more details on the experimental design.

questions about the contents of documents. This Copilot had the lowest usage of the three we study, partially because not all studied workers used Word regularly in their jobs (see Appendix B).

Some of the individual heterogeneity is also reflected in average firm usage rates, which ranged from as low as 21% to as high as 75%.[5] We do not observe many potentially importance sources of this firm variation, for example differences in the training provided to workers or managerial practices. Some of the variation across firms is due to different amounts of training and nudging that different firms did, which we cannot capture. However, we can look at how much of the vaiation is adoption is explained by (the proxies we have for) the type of work done at different firms (firm industry), the type of work done by study participants (pre-experiment individual work patterns[6]), and how firms distributed their Copilot licenses (the share of a worker's coworkers who have a license). We cannot estimate the causal effect of any of these factors, only the extent to which they can statistically explain the observed variation.

Table 1: Predicting Adoption

Individ. pre-experiment behavior	$\langle all \rangle$	$\langle all \rangle$	$\langle all \rangle$				$\langle all \rangle$
Share of coworkers with Copilot		0.044	0.053	0.099			0.038
		(0.024)	(0.024)	(0.023)			(0.024)
Firm FEs			Yes	Yes	Yes		
Industry FEs						Yes	Yes
R^2	0.093	0.094	0.235	0.189	0.184	0.050	0.139
Within R^2	–	–	.053	.006	–	–	.091

Note: This table reports OLS regressions where the dependent variable is the share of post-period weeks that an individual used Copilot. Coefficients for individual pre-experiment behavior variables are in Appendix Table 4.

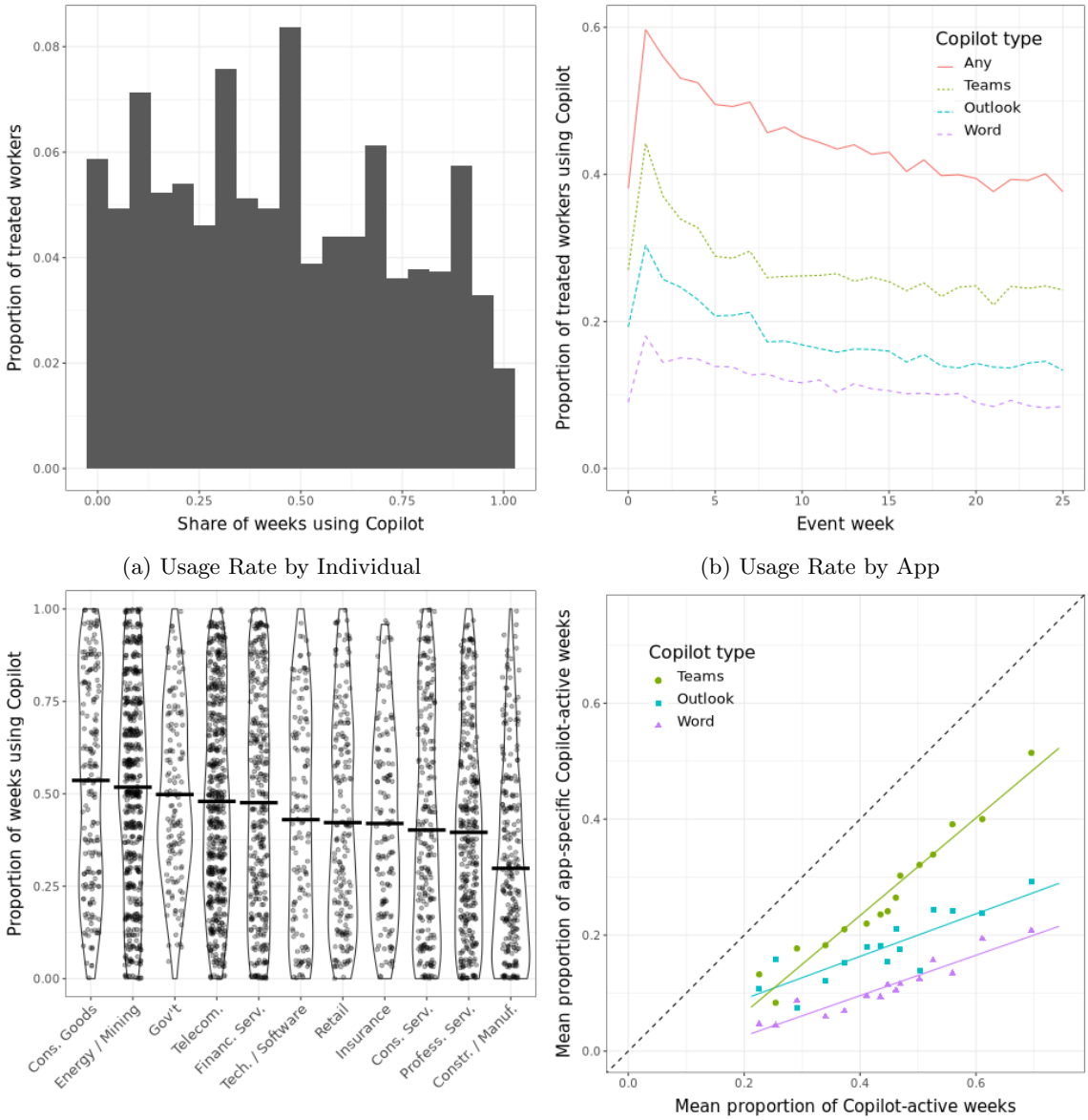
Table 1 shows the R-squared from regressions of each treated worker's average Copilot usage rate over the treatment period on various individual and firm metrics. The associated co-efficients are in Appendix Table 4. Individual pre-experiment activity in Outlook, Teams, and Word explains only 9.3% of the variation. The share of a worker's coworkers that have Copilot has a statistically significant impact on usage, but the effect is small and adds little explanatory power. Moving from no coworkers to the median of 12.5% would increase a person's usage by .66 percentage points. Firm fixed effects matter more, explaining 18.4% of the variation on their own and raising the R-squared from 0.094 to 0.235 when added to the other variables. Adoption rates vary meaningfully across industries (ranging from 30% in Construction and Manufacturing to 54% in Consumer Goods as shown in Figure 1c). Nevertheless, industry fixed effects explain less than a third as much of the variation in individual usage as firm fixed effects.[7]

Lastly, we look at how use of Copilot in different applications varied with overall firm usage. Figure 1d shows a binscatter of the usage of Copilot in the different apps by firms' all-up usage of Copilot. The Teams line is almost parallel to the 45° line, suggesting a constant share of people using Copilot

[5] Appendix Figure 4a shows how usage varied over time across firms.

[6] Weekly averages of: emails read, unique email threads replied to, Teams meeting time, Teams meetings attended, time in Word, documents edited or read before completion, number of unique email recipients, number of unique people met 1-on-1.

[7] Some of the firm variation could also come from time-of-year effects since firms joined the experiment at different times. In the Appendix we repeat this analysis with user-week level data where we can control for calender time. The R-squareds are overall much lower, since we are not averaging within a user, but firm fixed effects still add about twice as much explanatory power as individual pre-experiment work patterns.



(a) Usage Rate by Individual

(b) Usage Rate by App

(c) Usage Rate by Industry

(d) Usage Rates across Apps by Firm Usage

Figure 1: Copilot Usage in the Treatment Group

Note: Figures (a) and (c) calculate the share of weeks in the pre-period each worker used Copilot and show the distribution across (a) workers overall or (c) workers by industry. Figure (b) is average usage rate across all treated users over time. Figure (d) is a binscatter for the average share of users Copilot in a specific app in the average week grouped by the overall Copilot usage rate of the firm.

are using Copilot in Teams. However, while Outlook usage is relatively high amoung low-usage firms, it does not increase as much with firm usage. At firms with more active users, the additional usage seems to mainly come from the other apps.[8]

2 Methods

Since license assignment was randomized, comparing mean treatment and control group outcomes in the post-period would give an unbiased estimate. However, since our outcome metrics are high-variance, a difference-in-differences (DiD) approach greatly improves the precision of our estimates. Our intent-to-treat (OLS) specification estimates the effect of receiving a Copilot license on an outcome Y_{it} using a model with individual, calendar year-month, and firm-specific event month fixed effects:

$$Y_{it} = \alpha_i + \delta_t + \gamma_{f,\tau_{ft}} + \beta Z_i \mathbf{1}(\tau_{ft} \geq 0) + \epsilon_{it} \tag{1}$$

where f is the firm at which worker i works, τ_{ft} is the event month for firm f (month relative to when the firm started the experiment), and Z_i is the worker's experimental assignment. We include both calendar time and event time controls because firms entered and exited the study on a rolling basis over the 12 months. We also extend this two-way fixed effect setup to an event study specification where Z_i is interacted with a set of event month indicators rather than the single post-period indicator. We use Newey-West standard errors with a lag of 9 weeks to handle autocorrelation in outcomes. Since treatment was randomized and treatment and control workers in each firm have the same event-time, the potential biases from a rolling DiD analysis highlighted by (Goodman-Bacon, 2021) and others do not apply here.

As highlighted in the previous section, many workers who had access to Copilot did not use it regularly, and some workers who were assigned to the control group nonetheless obtained Copilot licenses. To adjust for low usage and non-compliance we also follow Duflo (2001) and Hudson, Hull and Liebersohn (2017) and estimate an instrumented difference-in-difference specification. In this case, we use treatment assignment as an instrument for whether a worker was an above-median Copilot user, using Copilot in at least 42% of the post-treatment weeks. The regression equations are[9]

$$\begin{aligned} Y_{it} &= \alpha_i + \delta_t + \gamma_{f,\tau_f} + \beta \hat{D}_{it} + \epsilon_{it} \\ \hat{D}_{it} &= \tilde{\alpha}_i + \tilde{\delta}_t + \tilde{\gamma}_{f,\tau_f} + \tilde{\beta} Z_i + \eta_{it}. \end{aligned} \tag{2}$$

These estimates capture the local average treatment effect of Copilot for the complying workers who actually use it. To extend this IV setup to the multi-period event studies, we instrument each $\hat{D}_{it}$ with interactions of Z_i and event time.

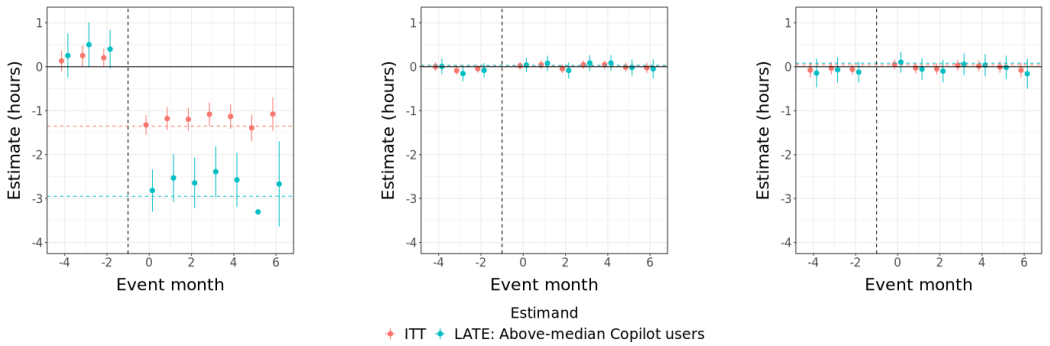
3 Effects of Generative AI on Patterns of Work

During this early adoption phase, the most prominent impact of Copilot on worker behavior was changes in how workers managed their email. In the pre-period, the average worker in our sample

[8] Appendix Figure 4b shows a similar pattern based on the average number of actions. As Copilot usage at a firm increases, the average number of Teams-Copilot actions among Copilot users increases, but the number of Outlook-Copilot actions is flat.

[9] Note that D_{it} does not vary (for an individual) across weeks within the pre- or post- period. It is one for all post-period weeks for individuals who were above-median Copilot users and zero otherwise.

spent over 11 hours per week reading and responding to emails[10], more than a quarter of a typical work week. For most workers, reading and writing emails does not represent their core job responsibilities but rather a communication overhead cost that supports their productive work. The task of managing email is also well-suited to generative AI tools. On average, workers with access to Copilot spent 1.4 hours less per week working on email in Outlook and workers who used Copilot regularly spent almost 3 fewer hours per week, a 25% decline from their pre-period average. Figure 2 illustrates that this reduction in Outlook time is apparent even in the first month after workers got access to the tool and persists throughout the study.



(a) Time in Outlook (b) Time in Word (c) Time in Teams Meetings

Figure 2: Event Study Plots for the Effect of Copilot on Time Allocation

Note: These plots show the coefficients from regressions similar to Equations (1) for the ITT (intent to treat) estimates and Equation (2) for the LATE estimates, where the β's are allowed to vary by firm's month from Copilot adoption. Vertical lines show 95% confidence intervals based on Newey-West standard errors with a lag of 9 weeks. Horizontal dashed lines give the average estimated effect from Table 2.

Figure 2 also illustrates that we find no similar shifts in the time workers spend in Teams meetings or working in Word, despite the fact that these tasks are also complemetary with the strengths of generative AI and, particularly in the case of Teams meetings, workers are regularly using Copilot through these applications. We can think of two hypothetical explanations for this difference. The first considers the demand for these tasks. When workers become more productive at a task, the effect on their total time spent doing it is ambiguous. Each unit of production in this task is now cheaper (in time) to produce, so workers may substitute towards it. For example, as generative AI improves workers' ability to capture and summarize the output of meetings, workers might choose to do more work such as brainstorming or outlining documents live in a meeting rather than asynchronously. Alternatively, if there is no demand for more of this task, workers will complete the same work as before in less total time. For example, we might see meeting times shorten or recurring meetings repeat less frequently. It could be the case that the second effect dominates in the case of email management, but the two effects roughly balance out for video meetings and document writing.

The second explanation concerns coordination across teams and organizations. During this study, the treated workers had access to Copilot, but very few of their colleagues did. In 2023, almost no companies had developed, let alone implemented, systemic plans of how generative AI could transform their workflows and allocations of responsibilities across workers. While emailing involves communicating with others, the tasks of reading emails, writing emails, and tracking conversations that require follow-ups are entirely solitary. Each worker is free to develop her own process for managing her inbox.

[10]Details on the construction of all worker behavior measures are in Appendix B.2.

In contrast, meetings are group activities. Shifting what work gets done in meetings or how long they last requires coordinating with colleagues and agreeing on new norms. Writing one's own contributions to a document is a solitary process, but taking on the responsibility of being the primary author of more documents or shifting work norms so that more documents are produced requires larger changes in organizational expectations.

Table 2: Effect of Copilot

Metric	**Pre-period mean (SD)**	**OLS (SE)**	**IV (SE)**	**Workers**	**n**
Outlook total session time	11.63	-1.35**	-2.95**	5,803	281,283
	(6.19)	(0.069)	(0.15)		
Total concentration time	24.60	0.90**	2.05**	5,620	247,907
	(6.00)	(0.060)	(0.14)		
Outlook out-of-hours session time	2.27	-0.30**	-0.65**	5,803	281,283
	(1.68)	(0.019)	(0.041)		
Unique email threads replied to	13.50	-0.23**	-0.47**	6,116	325,472
	(11.32)	(0.081)	(0.17)		
Time to reply from email delivery	14.76	-0.31**	-0.61**	6,111	285,332
	(4.89)	(0.092)	(0.19)		
Teams meeting time	5.56	0.041	0.081	5,698	208,754
	(3.43)	(0.034)	(0.066)		
Share of meetings attended late/early	0.29	0.006**	0.012**	5,645	178,218
	(0.11)	(0.002)	(0.004)		
Recurring Teams meetings time	2.37	0.013	0.025	5,698	254,319
	(1.71)	(0.016)	(0.031)		
Word total session time	1.34	0.018	0.032	2,150	114,863
	(0.91)	(0.023)	(0.041)		
Avg. document time to complete	179.20	-6.93	-11.65	2,150	36,523
	(80.71)	(5.37)	(9.03)		
Avg. collaborative document TTC	283.48	-35.83*	-55.77*	1,673	11,033
	(117.26)	(13.99)	(21.73)		
Avg. non-collaborative document TTC	145.70	-3.72	-6.32	2,140	30,975
	(74.40)	(5.20)	(8.83)		
Completed documents	0.79	0.028	0.051	2,150	91,213
	(0.70)	(0.017)	(0.031)		
Completed collaborative documents	0.15	0.011	0.019	2,150	91,213
	(0.21)	(0.006)	(0.010)		

Significance: ** $p < 0.01$, * $p < 0.05$

Note: For each outcome variable, this table shows pre-period mean and cross-user standard deviation (Column 1), the intent-to-treat coefficients for the effect of Copilot (Column 2) and the LATE coefficients (Column 3). Cross-user standard deviations are estimates of the standard deviation of the random intercept in the model $Y_{it} = \mu_i + \epsilon_{it}$. All time metrics are in hours. Share of meetings attended late/early refers to the share of Teams meetings that a worker joined more than 5 minutes after the meeting started or left more than 5 minutes before it ended. We assign documents to a worker if they were the 'primary editor', making the more most edits to the document.

Looking in more detail at how workers changed their behavior when using Copilot, we find more evidence for the second hypothesis: workers were more likely to adjust behaviors that could be changed unilaterally and less likely to shift behaviors that required coordination with colleagues. In the top panel of Table 2, we see that workers' responsiveness and intensity of email communication with colleagues is roughly unchanged. Workers who used Copilot replied to roughly the same number of

conversations per week (13 vs. 13.5) and replied to emails slightly faster than before. These workers are simply taking less time to do the same quantity of email work. We also illustrate some of the benefits of this shift in behavior. Workers who use Copilot are not only spending less time in Outlook each week, but condensing their work into fewer sessions. This opens up two more hours per week of concentration time, blocks of 30 minutes or more during the regular work day with no Outlook activity. Workers using Copilot are also disproportionaly cutting back in time spent in email outside of their regular work hours.[11]

If the null effect on total time spent in meetings masked offsetting shifts in how workers were using meetings, we would expect to see some changes in the types of meetings workers attended after gaining access to Copilot. However, as shown in Table 2, we find no more impact on time spent in recurring meetings, which we might have expected to fall, than in meetings overall.[12] The only impact of Copilot on meeting-related behavior that we find is a small, but precise, increase in the share of meetings where our studied worker joins late or leaves early, an individual choice. Teams is the most common application in which workers used Copilot, so they clearly found value in using generative AI in this context, perhaps in lower effort to retain information from meetings or more ability to focus on participating instead of taking notes. However, so far this value has not translated into any visible shift in the structure of work.

Focusing on workers who were regular users of Word in the pre-period, the bottom panel of Table 2 provides more details on how Copilot access and use affects workers' document production. We do not have the power to measure these effects very precisely because over half of study participants were not regular Word users. However, there are some suggestive patterns. Consistent with Noy and Zhang (2023), we find that workers with Copilot complete their documents more quickly, although the magnitude of the effect is far smaller than they found in the lab. Interestingly, the time savings are stongest, in both hours saved and percent reductions, in documents where our studied worker was the primary editor but at least one other coworker also contributed. These multi-author documents take longer to complete, nearly 12 days on average, so there may be more scope for time savings. Workers may also use generative AI to stay on top of colleagues' edits and outstanding tasks as well as to draft text. Consistent with workers changing the way they do their own work, but not taking on different responsibilities, we find no change in the number of documents where the studied workers are the primary editor.

3.1 Effects by Peer Copilot Use

To further explore the role of team and organizational coordination in shaping the early effects of generative AI, we consider whether the effects of Copilot access and use vary when workers have more close colleagues who also have access to Copilot. As discussed in Appendix B.3, we identify close colleagues as the coworkers our studied workers interact with most frequently through emails, meetings, and joint writing. Though we do not have random variation in the share of colleagues with Copilot, the treatment and control assignments are still random for people above (or below) the median, so we can get unbiased estimates for each subgroup.

Table 3 splits our sample by whether our studied workers have above or below median share (12.5%) of close coworkers with access to Copilot. We see almost no meaningful differences in behavior between these two groups, with one exception in document creation. Workers who have fewer colleagues with

[11] Appendix B.2 contains details on how we define work hours.

[12] In results not shown, we also find fairly precise zero effects on the number of meetings attended, ruling out a shift toward more shorter meetings, and on meetings by type (long meetings with few participants, short meetings with many participants, etc.), which previous research has shown to correlate with meeting purpose.

Table 3: Effect of Copilot - Split by Share of Coworkers with Copilot licenses

Outcome	Pre-period mean (SD)		OLS (SE)		IV (SE)		n	
	Above	Below	Above	Below	Above	Below	Above	Below
Outlook total session time	11.74	11.53	-1.50**	-1.27**	-3.04**	-2.97**	140,463	140,820
	(5.85)	(6.50)	(0.095)	(0.10)	(0.19)	(0.23)		
Total concentration time	24.60	24.61	1.05**	0.78**	2.22**	1.90**	128,272	119,635
	(5.71)	(6.28)	(0.084)	(0.087)	(0.18)	(0.21)		
Outlook out-of-hours session time	2.28	2.26	-0.33**	-0.28**	-0.67**	-0.64**	140,463	140,820
	(1.57)	(1.77)	(0.026)	(0.027)	(0.053)	(0.062)		
Unique email threads replied to	13.86	13.14	-0.26*	-0.26*	-0.49*	-0.58*	160,538	164,934
	(10.94)	(11.66)	(0.11)	(0.12)	(0.21)	(0.26)		
Time to reply from email delivery	14.94	14.58	-0.25	-0.35*	-0.47	-0.75*	144,189	141,143
	(4.84)	(4.92)	(0.13)	(0.13)	(0.24)	(0.29)		
Teams meeting time	6.02	5.09	0.009	0.067	0.017	0.14	105,376	103,378
	(3.35)	(3.45)	(0.047)	(0.048)	(0.087)	(0.10)		
% Teams meetings attended late/early	0.27	0.31	0.006*	0.006*	0.010*	0.013*	92,260	85,958
	(0.097)	(0.12)	(0.003)	(0.003)	(0.004)	(0.006)		
Recurring Teams meetings time	2.54	2.19	-0.008	0.031	-0.014	0.066	130,249	124,070
	(1.67)	(1.73)	(0.022)	(0.022)	(0.041)	(0.046)		
Word total session time	1.30	1.38	0.010	0.011	0.017	0.020	59,662	55,201
	(0.87)	(0.95)	(0.031)	(0.033)	(0.055)	(0.061)		
Avg. document time to complete	185.32	172.27	-1.95	-13.26	-3.23	-22.97	19,382	17,141
	(83.39)	(76.89)	(7.69)	(7.71)	(12.74)	(13.35)		
Avg. collaborative document TTC	286.03	279.82	-36.29	-24.95	-54.44	-42.13	6,507	4,526
	(118.51)	(115.26)	(19.21)	(22.51)	(28.83)	(37.96)		
Avg. non-collaborative document TTC	148.81	142.33	0.97	-9.62	1.64	-16.58	16,138	14,837
	(74.85)	(73.77)	(7.48)	(7.46)	(12.63)	(12.85)		
Completed documents	0.81	0.77	-0.005	0.057*	-0.009	0.11*	47,551	43,662
	(0.70)	(0.71)	(0.023)	(0.025)	(0.041)	(0.046)		
Others' documents read/edited	0.53	0.40	0.019	0.038*	0.033	0.072*	47,551	43,662
	(0.64)	(0.55)	(0.020)	(0.019)	(0.035)	(0.035)		
Completed collaborative documents	0.17	0.13	0.006	0.014	0.011	0.027	47,551	43,662
	(0.23)	(0.19)	(0.008)	(0.007)	(0.014)	(0.014)		

Significance: ** $p < 0.01$, * $p < 0.05$

Note: This table replicates the analysis in Table 2 separately for workers who had above or below median share of coworkers with Copilot licenses in the post-period. See that table not for explanations and variable definitions.

access to Copilot author 14% more documents when they use Copilot themselves. These documents are primarily solo authored, since we see no increase in collaborative documents, and these solo documents are completed more quickly. This pattern suggests that some teams adapted in the short run and reallocated writing work to the one team member with access to the new generative AI tool. However, we do not expect this pattern to persist in the long run as generative AI tools become more widespread. Overall, the lack of large differences suggest that the barriers to broader shifts in work patterns are primarily the time required to reimagine work flows and the institutional effort to drive these changes rather than simply coordination costs within small teams.

4 Conclusion

In order to capture the full productivity potential of new technologies, firms frequently have to undergo complex organizational transformations (Feigenbaum and Gross, 2024; Bresnahan, Brynjolfsson and Hitt, 2002). This institutional changes take time. Among recent technological innovations, generative AI is unique because it does not require new equipment or hardware to use, allowing individual workers to bring generative AI to work even without their employers' support. As Blandin, Bick and Deming

(2024) point out, the discrepency between the high rates of generative AI use reported in surveys of workers, such as theirs, and the lower rates reported in surveys of firms, as in Bonney et al. (2024), suggest that workers using these tools even before their employers develop formal adoption strategies.

This study provides some of the earliest signals of how this individual use of generative AI is beginning to shape the patterns of work for knowledge workers. We observed workers during the first year in which generative AI tools were beginning to gain broad use at work. While their colleagues may have been using other generative AI tools, the workers we studied were often the only members of their team with access to Copilot during the early roll-out period. The changing behaviors we see are consistent with workers independently exploring these new tools, but do not exhibit the larger shifts we might expect as teams and organizations begin larger transformations in response to this technology. We therefore expect that the patterns we find are only the beginning of a new era of work.

References

Acemoglu, Daron. 2025. "The simple macroeconomics of AI." *Economic Policy*, 40(121): 13–58.

Autor, David. 2024. "Applying AI to rebuild middle class jobs." National Bureau of Economic Research.

Blandin, Adam, Alexander Bick, and David Deming. 2024. "The Rapid Adoption of Generative AI." *Available at SSRN 4965142*.

Bonney, Kathryn, Cory Breaux, Cathy Buffington, Emin Dinlersoz, Lucia S Foster, Nathan Goldschlag, John C Haltiwanger, Zachary Kroff, and Keith Savage. 2024. "Tracking firm use of AI in real time: A snapshot from the Business Trends and Outlook Survey." National Bureau of Economic Research.

Bresnahan, Timothy F, Erik Brynjolfsson, and Lorin M Hitt. 2002. "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence." *The quarterly journal of economics*, 117(1): 339–376.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2021. "The productivity J-curve: How intangibles complement general purpose technologies." *American Economic Journal: Macroeconomics*, 13(1): 333–372.

Cambon, Alexia, Brent Hecht, Benjamin Edelman, Donald Ngwe, Sonia Jaffe, Amy Heger, Mihaela Vorvoreanu, Sida Peng, Jake Hofman, Alex Farach, Margarita Bermejo-Cano, Eric Knudsen, James Bono, Hardik Sanghavi, Sofia Spatharioti, David Rothschild, Daniel G. Goldstein, Eirini Kalliamvakou, Peter Cihon, Mert Demirer, Michael Schwarz, and Jaime Teevan. 2023. "Early LLM-based Tools for Enterprise Information Workers Likely Provide Meaningful Boosts to Productivity." Microsoft.

Campero, Andres, Michelle Vaccaro, Jaeyoon Song, Haoran Wen, Abdullah Almaatouq, and Thomas W. Malone. 2022. "A Test for Evaluating Performance in Human-Computer Systems."

Cui, Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. 2024. "The Effects of Generative AI on High Skilled Work: Evidence from Three Field Experiments with Software Developers." *Available at SSRN 4945566*.

Dell'Acqua, Fabrizio, Edward McFowland, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. 2023. "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." *Harvard Business School Technology & Operations Mgt. Unit Working Paper*.

Duflo, Esther. 2001. "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." *American economic review*, 91(4): 795–813.

Edelman, Benjamin G., Donald Ngwe, and Sida Peng. 2024. "Measuring the Impact of AI on Information Worker Productivity." *Available at SSRN 4648686*.

Edelman, Benjamin G., James Bono, Sida Peng, Roberto Rodriguez, and Sandra Ho. 2024. "Randomized Controlled Trials for Microsoft Copilot for Security." *Available at SSRN 4648700*.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. "GPTs are GPTs: Labor market impact potential of LLMs." *Science*, 384(6702): 1306–1308.

Feigenbaum, James, and Daniel P Gross. 2024. "Answering the Call of Automation: How the Labor Market Adjusted to Mechanizing Telephone Operation." National Bureau of Economic Research Working Paper 28061.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of econometrics*, 225(2): 254–277.

Hudson, Sally, Peter Hull, and Jack Liebersohn. 2017. "Interpreting instrumented difference-in-differences." *Metrics Note, Sept.*

Humlum, Anders, and Emilie Vestergaard. 2025. "The unequal adoption of ChatGPT exacerbates existing inequalities among workers." *Proceedings of the National Academy of Sciences*, 122(1): e2414972121.

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas. 2024. "AI adoption in America: Who, what, and where." *Journal of Economics & Management Strategy*, 33(2): 375–415.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science*, 381(6654): 187–192.

Otis, Nicholas, Rowan P Clarke, Solene Delecourt, David Holtz, and Rembrand Koning. 2023. "The uneven impact of generative AI on entrepreneurial performance." *Available at SSRN 4671369*.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. "The impact of AI on developer productivity: Evidence from GitHub copilot." *arXiv preprint:2302.06590*.

Spatharioti, Sofia Eleni, David M Rothschild, Daniel G Goldstein, and Jake M Hofman. 2023. "Comparing traditional and LLM-based search for consumer choice: A randomized experiment." *arXiv preprint:2307.03744*.

Toner-Rodgers, Aidan. 2024. "Artificial intelligence, scientific discovery, and product innovation." *arXiv preprint arXiv:2412.17866*.

Vaithilingam, Priyan, Tianyi Zhang, and Elena L Glassman. 2022. "Expectation vs. experience: Evaluating the usability of code generation tools powered by large language models." *Chi conference on human factors in computing systems extended abstracts*, 1–7.

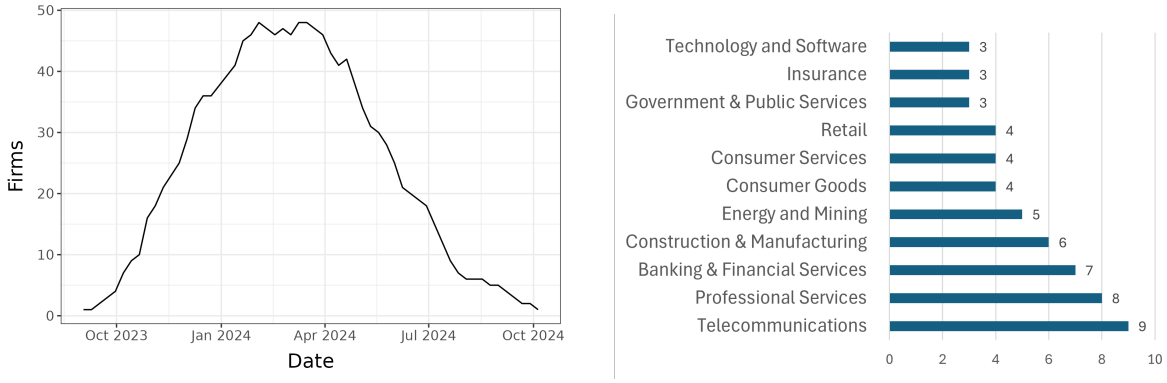
Appendix

A Experiment Details

We worked with firms to randomize the allocation of some of their 300 licenses within a pool of recruited workers. We asked all firms to allocate at least 50 licenses to the experiment and to give us a list to randomize of twice as many workers as licenses. Some firms did 200 or 250 licenses; a few did slightly less than 50. We asked firms to maintain the random assignment for at least six months. Our analyis sample has 54 firms. Additional firms were initially recruited into the experiment, but were dropped from the analysis because they failed to follow through with the randomization (many control group users got licenses or they dropped the assinments after fewer than 12 weeks), or selected workers for the experiment who were not well-suited to Copilot usage (less than 20% of treatment group workers ever used Copilot).

The data covers from May 05, 2023 to October 11, 2024. though many firms have earlier end dates when their 6-month post-period ended, they explicitly exited the experiment, or more than 20% of users in the control group got a Copilot license. We always drop all treatment and control users from a tenant, not just non-compliant individuals, so this removal from the experiment does not generate selection bias. In all analyses, we drop outcomes for the weeks ending December 29, 2023 and January 5, 2024 as work activity was low across all workers during this holiday period.

The study spanned one year, during which firms moved in and out of the sample. Most participating firms maintained the randomized allocation for the full six months. Figure 3a illustrates the number of partipating firms over the study window. Figure 3b shows the full list of firm industries. The experiment includes workers from many countries, with the majority based in the United States or Europe.



(a) Over Time

(b) Industry Mix

Figure 3: Participating Firms

Note: Figure (a) shows the number of firms active in the experiment over time. Figure (b) shows the number of participating firms in each industry.

B Data Details

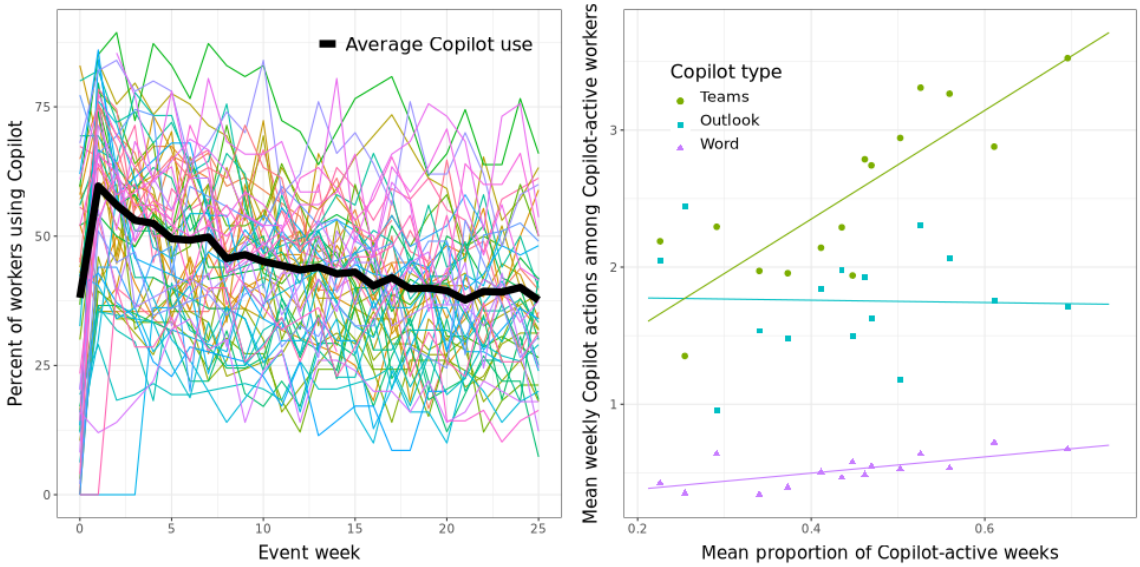
We analyzed millions of Microsoft 365 usage signals (telemetry), for both the treated and designated control workers, from product telemetry data that Microsoft already collects. To maintain workers'

and firms' privacy, we do not observe anything about the content produced by these workers, so our focus is on how workers allocate their time and interact with each other rather than on the quality of their work; moreover, we only analyze aggregate patterns, not individual activity.

For some analyses, we restrict our estimation to a subsample of workers who used the application in question. While virtually every worker in our sample used Outlook regularly, fewer workers used Word. We think it is reasonable that many workers do not have jobs that require document creation in Word. For our document results, we subset the analysis to users with higher Word use in the pre-period, indicated by above-median time spent in Word (.27 hours/week) and number of documents originated as the 'primary editor' (.19 documents/week). For Teams meetings, we drop five firms that were not regular users of Teams prior to the experiment because they relied on a different video meeting program.

B.1 Copilot Usage

We use the telemetry data to measure whether an individual has a license and their Copilot usage. All the usage data reflects usage only of Copilot, not other generative AI tools. Usage rates in this group may be lower than in other contexts since the technology was still new and most firms did not put their highest priority workers into this study (since the randomization meant all participants had only a 50% chance of receiving a license) or they may be higher if firms recruited people most excited about the new technology for the study. Figure 4 shows some additional Copilot usage patterns.



(a) Usage Rate by Firm

(b) Usage Intensity across Apps by Firm Usage

Figure 4: Copilot Usage

Note: Figure (a) shows usage over weeks of the experiment, with one line for each participating firm. Figure (b) shows a binscatter of the mean weekly Copilot actions taken in each app among users using some form of Copilot that week, along with the firm-level share of users that are active in an average week.

B.2 Outcome Metrics

We also construct weekly metrics of each worker's email, document editing, and meeting behaviors.

- **Outlook**

  For email, we aggregate the signals of various 'reading actions' individuals take into 'Outlook sessions': blocks of time where they don't go more than 15 minutes without reading an email.

  - **Outlook total session time** The time spent aggregated across all sessions in a week, measured in hours.
  - **Outlook out-of-hours total session time** The subset of Outlook session time that occured outside the individual's "working day" as defined as the 8 hours in which they were most active during one month of pre-period.
  - **Total concentration time** The total time spent during an individual's working day in non-Outlook blocks at least 30 minutes long.
  - **Unique email threads replied to** For emails received that week that are replied to within 7 days, the number of unique email threads/conversations.
  - **Time to reply from email delivery** For emails received that week that are replied to within 7 days, the average length of time it takes to reply to an email.

- **Teams**

  - **Teams meeting total session time** The time spent in scheduled Teams meetings (excluding spontaneous calls) in a week, measured in hours.
  - **Share of meetings attended 5 min late/early** The share of scheduled Teams meetings joined either $\geq$5 minutes after the Teams meeting was started or left $\geq$5 minutes before the meeting was ended.
  - **Recurring Teams meetings total session time** The time spent in Teams meetings marking as recurring meetings, measured in hours.

- **Word**

  - Across metrics, a worker is a primary editor if they make the most edits on a document; a document is completed when 90% of the total edits to the document have been completed.[13]
  - **Word total session time** The time spent in Word sessions, measured in hours.
  - **Avg. document time to complete** The average time between the creation (first edit) and completion of a document, for documents completed that week and where the target worker is the primary editor, measured in hours.
  - **Avg. collaborative document time to complete** The same as the average time to complete, but only for documents with multiple collaborators.
  - **Avg. non-collaborative document time to complete** The same as the "avg. document time to complete" but only for documents with no collaborators (one editor).
  - **Completed documents** The number of documents completed in a week where the target worker is the primary editor.

[13]Some documents show stray edits long after most active work is done, which may be accidental. This 90% cutoff aligns most often with other indicators of completion such as attaching to an email or printing to pdf.

– **Completed collaborative documents** The same as "completed documents" but only for documents with multiple collaborators.

– **Others' documents read/edited before completed** The number of documents read or edited in a week for which the target worker is not the primary editor.

B.3 Coworkers

We also use the telemetry data to identify the closest coworkers of each person in the experiment, based on a weighted sum of their joint meetings, calls, emails, chats, and document collaborations. For each metric, we divide by the firm-specific average. We add these normalized numbers together for an interaction score and call someone a "coworker" if the score is greater than 8.0 (If they add the average amount of interaction, the score would be 5.0 since there are five metrics.) If an individual has more than 15 coworkers, we take only the 15 with the highest interaction score. The average worker has 9 coworkers. We can then calculate what share of an individual's coworkers have a Copilot license. Figure 5 shows the distribution of that share.

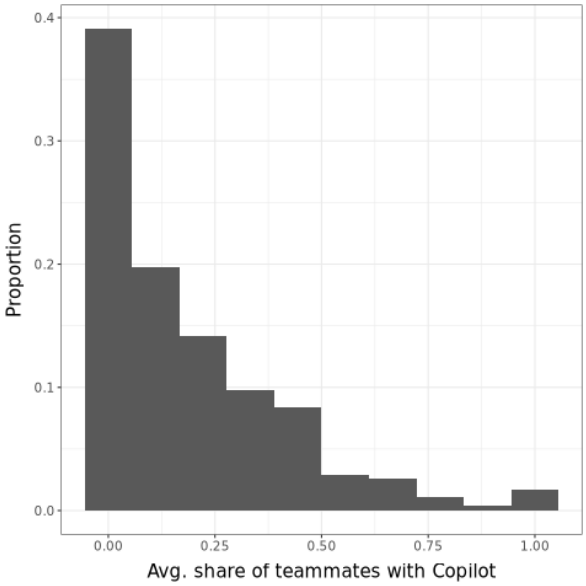


Figure 5: Distribution of share of coworkers with Copilot

C Additional Results Tables

Table 4 presents all the coefficients associated with the regressions in Table 1. Table 5 does the same analysis with user-week data instead of user-level averages. The R-squared's are lower since there is lots of unexplained week-to-week variation, however the pattern is similar.

Table 6 is similar to Table 3 in the paper, but instead of splitting individuals by the share of their coworkers who had a Copilot licenses, it splits by whether the average Copilot usage rate at the firm was above or below the median of 45%.

Table 4: Predicting Adoption - Coefficients

Constant	0.3800**	0.3705**					
	(0.0124)	(0.0134)					
Emails read	-0.0003**	-0.0003**	-0.0002**				-0.0004**
	(7.91×10^{-5})	(7.95×10^{-5})	(8.33×10^{-5})				(8.09×10^{-5})
Unique email threads replied to	-0.0014	-0.0014	-0.0011				-0.0010
	(0.0007)	(0.0007)	(0.0007)				(0.0007)
Teams meeting total session time	0.0109**	0.0108**	0.0107**				0.0127**
	(0.0030)	(0.0030)	(0.0032)				(0.0030)
Teams meetings attended	0.0068**	0.0067**	0.0040*				0.0050**
	(0.0018)	(0.0018)	(0.0020)				(0.0018)
Word total session time	0.0099	0.0107	0.0132				0.0213**
	(0.0068)	(0.0068)	(0.0073)				(0.0070)
Others' documents read/edited before completed	0.0590**	0.0574**	0.0340**				0.0570**
	(0.0121)	(0.0121)	(0.0123)				(0.0121)
Number of email recipients	-0.0009*	-0.0009*	-0.0008*				-0.0009*
	(0.0004)	(0.0004)	(0.0003)				(0.0004)
Number of unique people met 1-1 (total)	-0.0004	-0.0004	0.0003				-0.0004
	(0.0004)	(0.0004)	(0.0004)				(0.0004)
Share of coworkers with Copilot license		0.0439	0.0528*	0.0988**			0.0384
		(0.0239)	(0.0235)	(0.0226)			(0.0237)
Fixed-effects							
Firm			Yes	Yes	Yes		
Industry						Yes	Yes
Fit statistics							
Observations	2,718	2,718	2,718	3,167	3,167	3,167	2,718
R^2	0.09294	0.09412	0.23525	0.18860	0.18388	0.05045	0.13906
Within R^2			0.05276	0.00578			0.09072

Significance: ** $p < 0.01$, * $p < 0.05$

Note: This table shows OLS regressions results using individuals' pre-experiment behaviors, share of coworkers with a Copilot license, and firm and industry fixed effects to predict mean post-period Copilot weekly usage among treated workers. All variables are averages over the pre-period other than "number of unique people met 1-1," which is a total over a fixed pre-period.

Table 5: Predicting Weekly Copilot Use

Emails read	-0.0003**	-0.0003**	-0.0002**					-0.0004**
	(5.5×10^{-5})	(5.53×10^{-5})	(5.75×10^{-5})					(5.61×10^{-5})
Unique email threads replied to	-0.0013**	-0.0014**	-0.0010*					-0.0009
	(0.0005)	(0.0005)	(0.0005)					(0.0005)
Teams meeting total session time	0.0097**	0.0096**	0.0105**					0.0121**
	(0.0021)	(0.0021)	(0.0023)					(0.0022)
Teams meetings attended	0.0074**	0.0074**	0.0041**					0.0053**
	(0.0013)	(0.0013)	(0.0014)					(0.0013)
Word total session time	0.0127**	0.0131**	0.0134*					0.0218**
	(0.0047)	(0.0048)	(0.0051)					(0.0049)
Others' documents read/edited before completed	0.0498**	0.0491**	0.0341**					0.0534**
	(0.0084)	(0.0084)	(0.0087)					(0.0084)
Number of email recipients	-0.0007**	-0.0007**	-0.0008**					-0.0008**
	(0.0002)	(0.0002)	(0.0002)					(0.0002)
Number of unique people met 1-1 (total)	-0.0002	-0.0002	0.0002					-0.0003
	(0.0003)	(0.0003)	(0.0003)					(0.0003)
Share of coworkers with Copilot license		0.0236	0.0546**	0.1008**				0.0361*
		(0.0169)	(0.0172)	(0.0164)				(0.0169)
Fixed-effects								
Calendar week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Event month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm			Yes	Yes		Yes		
Industry							Yes	Yes
Fit statistics								
Observations	61,826	61,826	61,826	72,412	72,412	72,412	72,412	61,826
R^2	0.05026	0.05038	0.09285	0.07779	0.01986	0.07616	0.03432	0.06365
Within R^2	0.03130	0.03141	0.01595	0.00177				0.02973

Significance: ** $p < 0.01$, * $p < 0.05$

Note: This table shows OLS regressions results using individual pre-experiment behaviors, share of coworkers with a Copilot license, firm industry fixed effects, and calendar-week and event-month fixed effects to predict post-period Copilot weekly usage among treated workers. All variables are averages other than "number of unique people met 1-1," which is a total over a fixed pre-period.

Table 6: Effect of Copilot - above and below median firm usage of Copilot

Metric	Pre-period) mean (SD)		OLS (SE)		IV (SE)		n	
	Above	Below	Above	Below	Above	Below	Above	Below
Outlook total session time	11.45	11.86	-1.47**	-1.21**	-2.46**	-4.22**	156,648	124,635
	(6.00)	(6.41)	(0.096)	(0.099)	(0.16)	(0.35)		
Total concentration time	24.71	24.47	0.94**	0.86**	1.32**	3.19**	137,014	110,893
	(5.72)	(6.32)	(0.083)	(0.089)	(0.14)	(0.33)		
Outlook out-of-hours session time	2.22	2.34	-0.34**	-0.25**	-0.56**	-0.88**	156,648	124,635
	(1.35)	(1.70)	(0.026)	(0.027)	(0.043)	(0.094)		
Unique email threads replied to	13.37	13.66	-0.31**	-0.13	-0.50**	-0.41	181,898	143,574
	(10.82)	(11.92)	(0.11)	(0.12)	(0.18)	(0.37)		
Time to reply from email delivery	14.49	15.11	-0.38**	-0.21	-0.61**	-0.63	160,590	124,742
	(4.63)	(5.19)	(0.12)	(0.14)	(0.19)	(0.43)		
Teams meeting time	5.95	5.00	0.030	0.056	0.048	0.17	122,568	86,186
	(3.36)	(3.45)	(0.045)	(0.050)	(0.072)	(0.15)		
% Teams meetings attended late/early	0.28	0.29	0.010**	0.002	0.015**	0.005	107,111	71,107
	(0.100)	(0.13)	(0.002)	(0.003)	(0.004)	(0.009)		
Recurring Teams meetings time	2.57	2.09	0.018	0.006	0.028	0.019	148,836	105,483
	(1.73)	(1.35)	(0.021)	(0.023)	(0.033)	(0.068)		
Word total session time	1.42	1.21	0.044	-0.025	0.066	-0.068	71,076	43,787
	(0.94)	(0.86)	(0.030)	(0.034)	(0.045)	(0.093)		
Avg. document TTC	179.28	179.06	-9.05	-3.13	-12.85	-7.72	22,877	13,646
	(81.36)	(78.25)	(6.67)	(9.04)	(9.46)	(22.25)		
Avg. collaborative document TTC	277.53	295.74	-37.51*	-30.61	-52.44*	-65.23	7,429	3,604
	(110.45)	(130.26)	(16.59)	(26.14)	(23.14)	(55.66)		
Avg. non-collaborative document TTC	144.00	148.45	-5.77	-0.51	-8.18	-1.27	19,111	11,864
	(78.57)	(66.12)	(6.51)	(8.63)	(9.23)	(21.33)		
Completed documents	0.80	0.78	0.047*	-0.003	0.070*	-0.008	56,358	34,855
	(0.69)	(0.72)	(0.021)	(0.028)	(0.032)	(0.076)		
Completed collaborative documents	0.17	0.13	0.007	0.017	0.011	0.045	56,358	34,855
	(0.22)	(0.18)	(0.007)	(0.008)	(0.011)	(0.023)		

Significance: ** $p < 0.01$, * $p < 0.05$

Note: For each outcome variable, this table shows pre-period mean and cross-user standard deviation (Columns 1 & 2), the intent-to-treat coefficients for the effect of Copilot (Columns 3 & 4), and the LATE estimates (Columns 5 & 6) seperately for workers at firms who had above or below median Copilot usage in the post-period. All time metrics in hours. A worker is a primary editor if they make the most edits on a document. Time to complete (TTC) and completed document counts are calculated only for documents where the target worker is the primary editor.