Generative AI and the Reorganization of Work*

Lindsey Raymond
MIT & Microsoft Research

Eleanor Dillon
Microsoft Research

March 27, 2026

PRELIMINARY & INCOMPLETE

Abstract

While there is considerable interest in the capabilities of generative AI, causal evidence on its labor market effects remains limited. We study how the adoption of GitHub Copilot, an AI coding assistant, affects employment and workforce composition at large multinational firms. We exploit a natural experiment arising from frictions in enterprise software procurement: adding Copilot during a scheduled contract renewal requires only an amendment to an existing agreement, but adoption outside the renewal window requires a new procurement process. Firms with a renewal closer to the Copilot launch in January 2023 were therefore more likely to adopt early. We find no evidence that Copilot adoption causes large reductions in total employment or technical headcount. However, adoption reorganizes the workforce along two margins. First, adopting firms reduce the domestic share of their software engineering workforce, driven by a shift in the composition of new hires toward international locations. This effect is concentrated among US-headquartered IT services firms. Second, hiring of junior software engineers declines by 20% while senior hiring is unchanged, consistent with AI substituting for entry-level coding tasks. Unlike the geographic effects, the junior hiring decline is broad-based across firm types. These findings suggest that the early employment effects of generative AI operate through the reallocation of hiring, where firms recruit and at what seniority level, rather than through aggregate displacement.

JEL Classifications: D80, J24, M15, M51, O33

Keywords: Generative AI, Large Language Models, Technology Adoption, Worker Productivity, Worker Learning, Experience of Work, Organizational Design.

---

*Correspondence to lraymond@mit.edu and eleanor.dillon@microsoft.com. We are enormously grateful to Brian Xu, Kevin Xu, Lisa Simon, and Ben Zweig for invaluable data assistance, to Bee Xiong for helpful conversations about the GitHub sales process, to Namrata Kala and Kevin Xu for helpful comments, and to Will Wang and Ian Pittman for excellent research assistance. Raymond acknowledges support from the Schmidt Sciences AI2050 Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of Microsoft Research or the Massachusetts Institute of Technology.

1 Introduction

Technological progress has played a central role in how the nature of work has evolved over the past century. Successive waves of innovation, such as electrification, mechanization, computerization, reshaped the occupational structure of the economy, displacing workers in some tasks while creating demand for new ones (Autor, Levy and Murnane, 2003; Acemoglu and Autor, 2011). While living standards improved dramatically over the long run, technological transitions produced winners and losers, and the adjustment process was sometimes slow and uneven (Autor, 2015; Hicks, 1932).

Recent progress in large language models (LLMs) raises concerns that the next technological transition may be faster and broader in scope. Large language models are general-purpose AI systems trained on vast corpora of text and code that can generate, summarize, and transform natural language and programming languages. These models have demonstrated the ability to pass professional licensing exams, produce competent legal and financial analysis, and write functional software (Bubeck et al., 2023). What distinguishes this moment is not only the current capabilities of these systems but the speed at which they have been advancing and broad set of tasks AI can capably perform (Eloundou et al., 2024; Maslej et al., 2025). Some commentators argue that AI could displace a substantial share of entry-level white-collar employment within years. Others suggest LLMs are still unreliable and of limited utility, and point to the history of technological adoption, which suggests even transformative technologies like electrification took decades to diffuse through the economy and reshape labor markets (Rosenberg and Trajtenberg, 2004; David, 1990). Distinguishing between these possibilities requires causal evidence on if and how firms actually deploy AI tools and adjust their workforces in response.

Producing empirical answer to these questions has been complicated by two challenges. First, the release of generative AI tools in late 2022 coincided with the post-pandemic labor market rebalancing and a rapid tightening interest rate cycle by the Federal Reserve, masking the effects of this new technology. Second, data on firm-level AI adoption is scarce. Instead, imputed measures of AI exposure based on occupational task content are often used as proxies. While these proxies can identify firms whose workers could potentially be affected by AI, such an approach makes it difficult to distinguish the causal effect of adoption itself, as opposed to the average outcome across firms with similar workforce compositions.

In this paper, we provide early causal evidence of the impact of LLM-based coding assistants on employment at large, multinational, technology-intensive firms. We study the adoption of GitHub Copilot, an LLM-based AI programming tool initially built on OpenAI's Codex, among customers of GitHub's enterprise source code management software. Software engineering is a compelling setting for studying the labor market effects of AI for two reasons. First, it is a high-paying and fast-growing white-collar occupation and the information technology sector has been a central engine of productivity growth and employment expansion in the US economy. If AI tools reshape employment even in this sector, where workers are highly skilled, labor markets are tight, and firms tend to be well-resourced, the implications for other knowledge-work occupations are significant.

Second, the structure of enterprise software procurement creates a natural experiment that creates quasi-exogenous variation in the timing of firms' AI adoption.

GitHub Enterprise contracts follow predictable annual renewal cycles. During a scheduled renewal, the customer and GitHub already have an open procurement process with legal, security, and compliance stakeholders convened to negotiate the terms of the agreement. Adding Copilot during this window requires only an amendment to a negotiation already underway. Outside the renewal window, adoption requires reassembling these stakeholders and initiating a separate procurement process, raising the organizational cost of adoption. We exploit this friction: firms whose scheduled renewal fell close to the Copilot launch in January 2023 faced a lower cost of early adoption. Early-renewal firms are 9.6 percentage points more likely to have adopted Copilot in any given post-period month ($F = 84$). Renewal timing is balanced on pre-treatment firm characteristics and uncorrelated with pre-period employment trajectories, supporting the exclusion restriction.

We combine proprietary data from GitHub, including firm-level Copilot adoption dates and seat counts, Enterprise contract renewal dates, and subscription terms, with monthly employment records from Revelio Labs covering headcount, hires, and separations by role, seniority level, and geography from January 2021 to December 2025. Our sample consists of about 6,100 firms that were existing enterprise customers of GitHub in the third quarter of 2022, before the launch of GitHub Copilot. These firms are disproportionately large, with an average employment of 8,900 workers, and technology-intensive, employing about 400 software engineers. About half are headquartered in the United States, and almost all are multinationals and employ workers in multiple countries.

We have three sets of findings. First, we find no evidence that Copilot adoption causes large reductions in total employment, software engineering headcount, or other technical roles. The effect on aggregate employment is theoretically ambiguous: if AI coding tools substitute for developer labor, firms may reduce headcount; if instead the tools raise productivity and lower the cost of software development, firms may expand output and hire more workers. The 2SLS point estimate for log headcount is 0.014 with a standard error of 0.08, and the 95% confidence interval $[-0.14, 0.17]$ rules out declines larger than 14%. Estimates for software engineering and technical employment are similarly small and imprecise. These results suggest that, over the medium run, AI coding tools have not led to dramatic workforce reductions at adopting firms, but neither have they generated detectable employment gains. However, the reduced-form point estimates for domestic software engineering employment, SWEs located in the firm's headquarters country, are negative, with early-renewal firms employing roughly 3% fewer domestic SWEs. This indication that aggregate stability may mask geographic reallocation within firms motivates our second set of results.

Second, we examine whether Copilot adoption reshapes the geography of software engineering work. Technology has historically shifted the location of production as coordination costs fall (Grossman and Rossi-Hansberg, 2008; Blinder, 2006) and AI coding tools could accelerate this process by standardizing code quality across locations or reducing the costs of managing distributed teams. Alternatively, if the tools disproportionately raise the productivity of onshore developers, firms may consolidate work domestically. Our sample is well-suited to test between these channels: three quarters of firms in our sample employ software engineers in both their home country and abroad, with the United States and India accounting for the largest shares. Early-renewal firms reduce the domestic share of their software engineering workforce by 0.6 percentage points relative to late-renewal firms. The corresponding 2SLS estimate implies a 7 percentage point decline among compliers, from a mean of 70 percent. This effect is driven by the composition of new hires rather than differential separations: the domestic share of quarterly SWE inflows falls by 1 percentage points in the reduced form, or roughly 11 percent among compliers. The geographic recomposition is concentrated among US-headquartered IT services firms, who shift the composition of their software engineering workforce toward international locations. Firms outside IT or headquartered outside the US show smaller or null effects. This heterogeneity is consistent with the offshoring channel: firms with high domestic labor costs and existing international operations are best positioned to substitute offshore for onshore developers when AI tools compress quality differences across locations.

In our third set of results, we explore whether Copilot is a complement or substitute for experience. AI coding assistants could automate tasks disproportionately performed by junior developers, such as simple code, refactoring and debugging, which should reduce demand for entry-level hires. On the other hand, evidence from worker-level studies of the productivity effects of AI tools often show that less experienced workers see a proportionally larger benefit from access to AI tools (Brynjolfsson, Li and Raymond, 2025; Noy and Zhang, 2023; Dell'Acqua et al., 2023). We test this by examining how Copilot adoption affects the seniority composition of new hires. Quarterly inflows of junior SWEs (seniority levels 1–2) decline by 3 workers per quarter among early-renewal firms, a 20 percent reduction from the sample mean. Senior SWE hiring is unchanged. The combination of fewer juniors and stable seniors, suggests that AI substitutes for entry-level tasks without displacing experienced workers, consistent with a reallocation of the task distribution within teams rather than a reduction in the overall demand for software engineering labor. Unlike the geographic recomposition, which is concentrated among US IT firms, the junior hiring decline is broad-based: all four industry-by-headquarters cells are negative and but not statistically different from one another.

Together, these findings provide early evidence on how AI tools reshape the labor market at adopting firms. Copilot adoption is not associated with the large-scale displacement of technical workers that some forecasts predicted. Instead, the effects operate at the margins of hiring: firms shift where they recruit and at what seniority level, while leaving aggregate headcount unchanged. Because these effects work through inflows rather than separations, they are easy to miss in real time; Nonetheless, if sustained, they will compound and eventually alter the stock of employment. The capabilities of AI coding tools continue to expand rapidly. Whether the patterns we identify accelerate, stabilize, or reverse as the technology matures is a central question for future research.

This paper contributes to a growing literature on the labor market effects of AI. Several recent studies examine how firms' AI investments affect employment, generally finding positive associations between AI adoption and firm-level revenue and employment growth (Babina et al., 2024; Hampole et al., 2025). However, Acemoglu et al. (2023) show that AI-adopting firms grew faster and had higher labor productivity even before making these investments, raising concerns that selection drives the observed correlations.

Evidence on the composition of employment is more nuanced. Hampole et al. (2025) find that employment falls for workers in AI-exposed roles at adopting firms, and Acemoglu et al. (2022) find that firms with more AI-exposed workers shift their vacancy postings toward machine learning skills and away from other competencies. Using Danish administrative data, Humlum and Vestergaard (2025) find no effects of AI adoption. In the US, Brynjolfsson, Chandar and Chen (2025) document lower recent employment growth for young workers in exposed occupations, while Dillon et al. (2025) find null effects of Copilot office tools. In a prior study of GitHub Copilot among smaller firms, Baird et al. (2024) find largely null effects on hiring at firms that invest in GitHub Copilot.

Our paper advances this literature in two ways. First, we provide causal estimates using an instrument for adoption timing, a central identification challenge. In our data we show adopters were growing faster before Copilot launched, and use renewal timing to isolate exogenous variation in when firms adopt. Second, we observe actual adoption directly rather than imputing AI exposure from occupation or vacancy descriptions, and we can trace the workforce response along specific margins, such as geography and seniority, documenting composition effects.

We also contribute to a literature on skill-biased technological change and the task content of work (Katz and Murphy, 1992; Autor, Levy and Murnane, 2003; Acemoglu and Autor, 2011). Historical studies of specific technologies illustrate how automation reshapes the demand for different types of labor: Feigenbaum and Gross (2024) study AT&T's automated switchboards, and McElheran et al. (2025) document the effects of AI and automation in manufacturing. This work has primarily used the task framework to understand which workers are affected by technological change, for instance, those performing routine, codifiable tasks are most exposed to displacement. Our findings suggest the task framework may also help explain where work is performed. If the task composition of an occupation varies across locations, then technologies that automate specific tasks will shift not only the skill composition but also the geographic distribution of employment. The geographic reallocation we document is consistent with this channel.

2 Setting

2.1 AI Coding Assistants

The use of AI-assisted coding for professional developers predated the emergence of LLMs. Early machine learning-based coding support, such as IntelliSense which was integrated into the Visual Studio IDE in 1997 and Tabnine, primarily offered line-by-line code completion and error flags. As early as 2006, Java developers were using this line completion functionality as often as they took cut or paste actions (Murphy, Kersten and Findlater, 2006). Code completion remained one of the most popular features of integrated development environments (IDE) (Aye, Kim and Li, 2021), but it was primarily useful for providing information and reducing time spent on documentation rather than meaningfully reducing the number of keystrokes required to complete code (Jiang and Coblenz, 2024).

LLMs, who often include large amounts of open source code in their training corpus, immediately improved the ability of AI tools to suggest coding solutions. Even general purpose LLM tools like ChatGPT, released in November 2022, were useful for answering coding questions. GitHub Copilot, released in October 2021 as an extension to Visual Studio Code IDE, with additional IDEs introduced later in 2021, was one of the first coding-specific tools powered by an LLM. GitHub Copilot was originally presented as an evolution of code completion tools, highlighting its ability to suggest multiple lines of code or create whole functions from context including natural language descriptions in comments. GitHub Copilot was powered by OpenAI Codex, a language model trained on billions of lines of publicly available source code. The initial release emphasized that Copilot worked best in the languages that are common in this public source code such as Python and JavaScript.

LLM-powered coding assistants continued to expand the capability to infer user intent, reason over multiple files, and generate and debug large sections of code from conversational prompts. GitHub Copilot released its chat function in March 2023. Other developer-focused coding assistants were released in 2023, including Cursor and Amazon CodeWhisperer. 2025 marked another leap in capabilities with the release of command-line interfaces including Claude Code and GitHub Copilot CLI.

By 2023, the first year that the Stack Overflow Developer Survey[1] included a section on AI, 83% of professional developers reported using ChatGPT for AI-powered search and 56% reported using GitHub Copilot as an AI-powered developer tool (Stack Overflow, 2023). Other LLM-powered developer tools were still emerging with 5% of professional developers reporting use of Amazon CodeWhisperer, the next most popular tool. The benefits of AI-assisted coding are also illustrated in the way professional developers wrote code. In 2018, three development environments shared roughly equal market share and only one was a full-featured IDE with code completion options (Stack Overflow, 2018). By 2023, two of the top three developments were IDEs and 74% of professional developers used Visual Studio Code, the first IDE to offer GitHub Copilot (Stack Overflow, 2023). Use of AI coding assistants continued to increase as the tools became more powerful. 81% of professional developers who responded to the 2025 Stack Overflow Developer Survey reported using AI tools in their development process, up from 44% in 2023 (Stack Overflow, 2023, 2025).

Unlike earlier code completion tools, LLM-powered tools show some potential to help software developers produce usable code faster. Using the results of a randomized trial, Cui et al. (2025) find that use of GitHub Copilot allowed professional developers to complete 26% more tasks. Importantly, AI coding assistants seem for now to be primarily useful as supports to attentive human coders rather than autonomous artificial coders. In other words, the augment software developers rather than replacing them. Pearce et al. (2025) review the limitations of these AI assistants, with a particular focus on security vulnerabilities that can be introduced with AI-generated code.

2.2 GitHub and Its Role in Software Development

GitHub is a web-based platform for storing, managing, and collaborating on software code. It is the dominant platform for code management across businesses and individual coders. 6sense estimates it

[1] An annual survey of roughly 75,000 developers in 185 countries recruited through Stack Overflow, a popular question and answer website for programmers.

had an overall market share of 87% as of January 21, 2026. 81% of professional developers reported using GitHub at work in the 2020 Stack Overflow Developer Survey (Stack Overflow, 2020), a share which held steady through 2025.

GitHub Enterprise is a GitHub product aimed at large organizations that allows them to bundle licenses for all employees. In addition to these scale advantages, GitHub Enterprise facilitates work on internal projects and data by allowing top-down security controls and custom permissions. Microsoft reports that 90% of Fortune 100 companies use GitHub; globally, large firms overwhelmingly use GitHub Enterprise. Our sample of GitHub Enterprise customers is therefore largely representative of large firms, within and beyond the United States.

Most large firms negotiate custom agreements for GitHub Enterprise, working with an assigned account team within GitHub. GitHub is sold on a subscription model, with a common annual renewal cycle. Sales teams advise firms on how to integrate GitHub products into their workflows, help determine the right scale and mix of products for each firm, and settle pricing. Renewal dates are staggered throughout the calendar year for different firms. While mid-year adjustments in subscriptions are possible, many of the changes in firms' subscriptions, including the addition of Copilot, take place around this annual renewal.

2.3 GitHub Copilot: Product Development

Understanding why enterprise procurement governs access to AI coding tools is central to our identification strategy. GitHub Copilot was initially released in 2021 as a product for individual developers. Microsoft launched GitHub Copilot for Business in December 2022, with the earliest enterprise subscriptions beginning in January 2023, followed by GitHub Copilot for Enterprise in November 2023. For the large firms in our sample, enterprise procurement is the primary channel through which developers gain access to AI coding assistants. Corporate security policies typically restrict which software can be installed on managed devices, and many organizations explicitly prohibit unapproved tools from transmitting proprietary code to third-party services. Even if an individual developer could install the consumer product, it would lack access to internal repositories—a key limitation, since the enterprise product integrates with an organization's proprietary code base to generate context-aware suggestions. Enterprise adoption therefore requires a firm-level procurement decision: seats are purchased at the organization level, security and access controls are configured centrally, and the tool is deployed within the firm's existing development infrastructure. Copilot for Enterprise further allows organizations to fine-tune models on internal code and offers more powerful models with higher usage limits. Our sample includes subscriptions to both the Business and Enterprise tiers.

As noted earlier, GitHub Copilot dominated the early stage of LLM-powered coding assistants. It remains the market leader even with the emergence of more competitors. In 2025, 68% of professional developers who used agentic AI assistants reported using GitHub Copilot, vs. 41% for Claude Code, the next most popular developer tool (Stack Overflow, 2025). Because of the security options and the integration in the Microsoft ecosystem (including GitHub and Visual Studio Code), GitHub Copilot remains particularly popular with large enterprises. As of 2025, Microsoft reports that over 77,000 organizations use GitHub Copilot for Business, including 90% of Fortune 100 companies. While workers at enterprise firms may have worked with individual GitHub Copilot licenses prior to the release of these enterprise products, the added ability to reference internal code bases and organization-wide availability likely made the tool substantially more useful for workers.

3 Data

Our analysis combines three data sources: GitHub Enterprise licensing records, GitHub's customer relationship management (CRM) system, and Revelio Labs employment data.

3.1 GitHub Enterprise Licensing and Sales Data

GitHub provides monthly firm-level snapshots of enterprise product licenses, measured on the first of each month. For each firm-month, we observe the number of GitHub Enterprise seats a firm has purchased. This number reflects the total block of licenses purchased by the firm, including both activated licenses and seats not currently in use. We also observe the number of GitHub Copilot Enterprise and Business seats in use by a firm, which reflect active licenses only. GitHub Business and GitHub Enterprise are slightly different products for business customers; for the remainder of the analyses we refer to both as Copilot. We will define Copilot adoption as a firm having non-zero Enterprise or Business Copilot seats. These records cover the universe of GitHub's enterprise customer base from January 2021 through June 2025. Multiple subsidiaries of a parent firm sometimes purchase licenses separately; we use GitHub's internal firm hierarchy data to link sub-accounts to their ultimate parent firms, allowing us to construct firm-level measures of GitHub Enterprise and Copilot adoption at the parent level, along with metadata such as firm name and country. This dataset allows us to track the timing and intensity of GitHub Copilot adoption at the firm-month level.

GitHub's Salesforce CRM system tracks enterprise customer contracts for the GitHub Enterprise sales team over the same period. For each monthly snapshot, we observe whether a firm holds an active GitHub Enterprise contract and the firm's next scheduled contract renewal date—the date at which the contract is up for renegotiation. On average, renewal cycles are annual, although there is variation in the contract renewal cycle. Contract renewal dates are updated as new contracts are renegotiated, allowing us to reconstruct each firm's renewal cycle history. As documented in Section 5.2, prior to the launch of GitHub Copilot, renewal dates change in 88% of renewal months and only 3% of off-cycle months, validating these dates as accurate measures of contractual renewal timing. We link the renewal data to the licensing data using internal GitHub firm identifiers, constructing a firm-month panel of product adoption and contract renewal timing.

3.2 Revelio Labs Employment Data

We merge the GitHub data to firm-level employment data from Revelio Labs, a workforce analytics provider that collects the near-universe of LinkedIn profiles and online job postings and harmonizes them into a structured employer–employee panel. The data contain information on workers' employment spells, job titles, educational background, county of employment, and skills. Revelio also provides company-level metadata, including headquarters country, year founded, NAICS codes, industry classification, and other firm identifiers such as CUSIP numbers.

We construct firm-month panels of overall employment, as well as employment by job category (e.g., software engineering), seniority level, and country of employment. Country of employment is drawn from the location reported on each worker's online resume data. Job categories are constructed by applying a clustering algorithm to job descriptions from resumes, online profiles, and the responsibilities sections of job postings, grouping the tens of millions of unique job titles in the data into a standardized occupational taxonomy. For example, titles such as "Software Developer," "Backend Engineer," and "Full Stack Developer" are all clustered together under a single "Software Engineering" category. A representative title is assigned to each cluster. At the most granular level, the taxonomy contains 1,700 distinct job categories, which are aggregate up to 150 groups. Seniority is assigned using an ensemble model that combines three inputs: (i) the worker's current title, company, and industry; (ii) career history features such as tenure and prior seniority; and (iii) the worker's age. The model produces a continuous seniority score for each individual. We use these panels to construct firm-level hiring flows (new worker–firm matches) and separations (end of employment spells) over time, incorporating Revelio sample weights that account for the probability that a given worker's profile appears in the data and for the time delay between a worker starting a new job and updating their online profile (Revelio Labs, 2026).

Revelio data are primarily drawn from LinkedIn, which over-represents workers in professional occupations, younger cohorts and larger firms. Revelio coverage is strongest among large, digitally intensive firms that employ software engineers, which overlaps with precisely the population of GitHub Enterprise customers. Other work has validated that Revelio provides reliable coverage of the white-collar labor market in developed countries. Tambe (2026) show that the distribution of education, occupation, and age in Revelio closely tracks the CPS and ACS, particularly among managerial and professional roles. Hampole et al. (2025) show that Revelio tracks both the levels and five-year changes of employment in their sample of AI developers and AI-exposed firms. We further validate the Revelio data in our sample by comparing employment levels and dynamics to external benchmarks, and GitHub measures of software engineering employment.

3.3 GitHub–Revelio Firm Matching

We link the GitHub licensing and renewal data to Revelio by matching on firm name. Revelio's internal algorithm uses LinkedIn company pages as a common identifier, linking approximately 60% of GitHub enterprise firms algorithmically. For the remaining firms, we use an AI agent-assisted procedure: a large language model searches for each firm's LinkedIn company page, using firm name, country, and industry to identify the correct profile. Appendix Section A.2 describes both matching procedures in detail. We are able to match 83% of our sample. Firms we are unable to match tend to be Japanese firms, where LinkedIn coverage is low and non-Latin characters complicate name matching. As a final filtering step, we drop firms where the average Revelio-provided total employment is less than the average number of GitHub Enterprise licenses purchased by the firm in the pre-period. Because GitHub seats are purchased for individual developers, a firm's total workforce should weakly exceed its seat count. Firms that violate this condition likely reflect entity-matching errors between the GitHub and Revelio databases, or cases where Revelio's LinkedIn-based employment estimates have sparse coverage of the firm's workforce. This filter drops 986 firms, leaving a final sample of 6,745 firms.

We validate each pipelines by comparing random samples of 100 firms each against the match a human reviewer would have selected. The algorithmic matches have an agreement rate of 83%, with most errors involving subsidiaries matched to parent companies rather than entirely wrong firms. The agent-assisted matches achieve a precision of 0.97 and recall of 0.91 (F1 = 0.94). Appendix Tables A.1 and A.2 report detailed error classifications for each pipeline. Because our sample is primarily large, established, tech-intensive firms, we are able to match 93% of GitHub enterprise firms to a Revelio company profile.

3.4 Revelio Data Validation

Our analysis relies on Revelio to measure employment levels, hiring and separation flows, and workforce composition, particularly among software engineers, whom we identify using Revelio's job taxonomy. We validate these measures against external benchmarks along two dimensions.

For the 19% of firms in our sample with 10-K filings, we compare average annual Revelio employment to total employment reported in 10-K filings from Compustat. Two sources of discrepancy are worth noting: 10-K employment generally covers only W-2 employees, while Revelio includes all workers who self-report employment on LinkedIn, including contractors and freelancers; and 10-K figures are measured at fiscal year-end, while our Revelio measure is an annual average. Despite these differences, the two measures are strongly correlated (Pearson's $r = 0.827$, Spearman's $r = 0.853$ for the natural log of employment). Appendix Figure A.3 plots 10-K reported employment against Revelio employment.

Because our identification strategy relies on within-firm changes over time, we also validate that Revelio captures employment *dynamics*. We compare year-over-year changes in average annual employment reported by Revelio to year-over-year changes in 10-K employment for the Compustata-matched subsample and find a strong correlation (Pearson's $r = 0.530$, Spearman's $r = 0.576$). Appendix Figure A.4 plots employment flows reported in yearly 10-K reports against changes in Revelio employment over the same period.

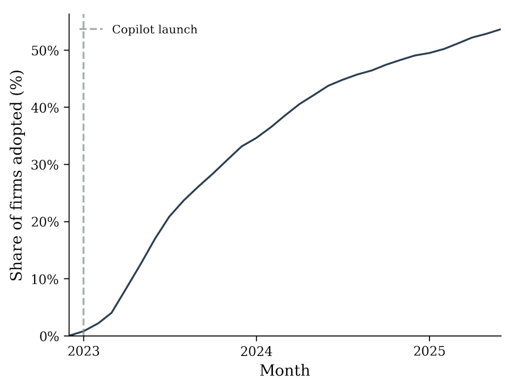
4 Empirical Strategy

4.1 Sample Construction

Our sample comprises private-sector firms that are customers of GitHub Enterprise and are served by the GitHub sales team, enabling us to observe contract renewal timing. We restrict to firms matchable to Revelio Labs employment data, which provides our primary outcome measures. The baseline sample includes all firms with an active GitHub Enterprise contract as of September 2022, one quarter before the announcement and launch of GitHub Copilot Enterprise, so that we observe each firm's position in its renewal cycle at the time of Copilot's introduction. Note that we do not restrict our sample to forms that adopt Copilot. In our sample period, 60% of firms eventually adopt, while 40% do not. Out of our original sample, 93% can be matched to a firm in the Revelio database. 68% are matched algorithmically, and the remaining 32% are manually matched.

We apply three exclusion criteria. First, we drop 558 public-sector entities: 377 identified by NAICS code 92 ("public administration"), and an additional 181 identified by organization name where NAICS code was missing.[2] Public procurement processes differ sufficiently from private-sector contracting to make these firms non-comparable. Second, we drop firms with no employment, no software engineering employment or where Revelio's pre-period headcount falls below the firm's GitHub Enterprise seat count. These conditions indicates an erroneous or incomplete entity match, since these firms have confirmed nonzero headcounts in GitHub's purchasing data, the GitHub product is targeted at companies who code and employ SWEs, and GitHub seats are generally purchased for a subset of technical workers within a firm's total employment. The final sample consists of 6,653 firms, forming a balanced panel of monthly employment and GitHub data spanning January 2021 through June 2025.

Figure 1: Cumulative Copilot Adoption Among Sample Firms



As shown in Table 1, the sample is 52% U.S.-headquartered, with an average of 8,929 employees as of September 2022. Software engineers account for 5.5% of employment, and technical roles more broadly account for 12.6%[3]. Information Technology is the most prevalent sector (23%), followed by Financial Services (14%) and Healthcare (7%). The vast majority of firms employ workers in multiple countries, as expected of our sample of large enterprise firms. Figure 1 plots the cumulative adoption curve. Adoption begins immediately after the launch of Copilot for Business in January of 2023 and rises steeply through 2023 before gradually leveling off. By June 2025, 60% of the sample has adopted GitHub Copilot, with an average of 143 Copilot seats among adopters. The remaining 40% of firms do not adopt during our sample period.

[2]We classify an organization as public sector if its name contains any of the following terms: *government, county, city of, state of, department of, ministry, municipality, public sector, federal, province,* or *prefecture.*

[3]We define technology workers as employees in the following job categories: software engineer, application developer, data engineer, QA engineer, infrastructure engineer, systems engineer, cybersecurity engineer, and design engineer.

Table 1: Summary Statistics at Baseline (September 2022)

	Full sample	Adopters	Non-adopters
Employment (Revelio)	8,929	11,180	5,543
	(33,597)	(40,088)	(19,761)
Firm Age	46	44	50
	(44)	(44)	(44)
SE employment	389	530	177
	(3,173)	(3,747)	(2,004)
SE share of employment (%)	5.5	6.8	3.6
	(7.2)	(7.4)	(6.6)
Technology share of employment (%)	12.6	14.2	10.1
	(11.6)	(11.5)	(11.3)
Data/analytics share (%)	1.8	2.0	1.5
	(2.7)	(2.7)	(2.6)
Analyst share (%)	7.6	8.1	6.9
	(7.3)	(7.5)	(6.9)
Junior share of employment (%)	29.4	27.2	32.6
	(15.8)	(15.3)	(15.8)
Domestic share of employment (%)	71.1	69.0	74.4
	(27.5)	(26.8)	(28.2)
Hiring rate, monthly avg. (%)	2.40	2.61	2.08
	(1.28)	(1.32)	(1.13)
Separation rate, monthly avg. (%)	1.75	1.86	1.58
	(0.73)	(0.72)	(0.70)
Net flow rate, monthly avg. (%)	0.65	0.75	0.49
	(0.82)	(0.87)	(0.71)
Headcount growth, 2021–2022 (%)	8.1	10.0	5.4
	(14.6)	(15.3)	(13.1)
Avg. renewal cycle (months)	12.1	11.9	12.3
	(1.8)	(1.4)	(2.1)
Multi-country workforce (%)	99.1%	99.5%	98.4%
U.S. headquartered (%)	52.3%	55.0%	48.2%
U.K. headquartered (%)	6.3%	6.7%	5.7%
Canada headquartered (%)	5.0%	4.2%	6.2%
Information Technology (%)	23.2%	28.3%	15.5%
Financial Services (%)	13.9%	13.1%	15.1%
Healthcare (%)	6.8%	5.5%	8.9%
N firms	6,654	3,997	2,657

Adopters differ from non-adopters along several dimensions. Adopters have a roughly double the share of software engineers (6.8% vs. 3.6%) and technology workers (14.2% vs. 10.1%), Adopters also have a more senior workforce; junior employees account for 27.2% of adopters' headcount compared to 32.6% for non-adopters. Adopters also hired more workers in the pre-period, had greater separation rates (1.86% vs. 1.58%) in the pre-Copilot period, and were growing faster in the year before launch (10.0% vs. 5.4%). The average contract renewal cycle is 12 months in both groups.

4.2 Instrumental Variables Design

Copilot adopters differ systematically from non-adopters along dimensions likely correlated with employment trajectories. To isolate exogenous variation in the timing of adoption, we exploit a feature of how enterprise software procurement works. GitHub Enterprise contracts are typically renegotiated on an annual cycle. The renewal is a natural decision point: it is when the GitHub sales representative is scheduled to contact the firm. It is also when lawyers, security and compliance experts, and other stakeholders are already convened to negotiate the terms of the enterprise agreement such as what activities are tracked by the software, what data on repository access is collected, who can access that data, and where it is stored. Adding Copilot during this window requires only an amendment to an agreement already under negotiation. Outside the renewal window, adoption requires initiating a separate procurement process, reassembling the same stakeholders, and negotiating a standalone addendum. These frictions imply that scheduled renewal dates generate plausibly exogenous variation in the timing of Copilot adoption.

Pre-Copilot GitHub Enterprise renewal activity demonstrates how sticky the renewal cycle is in firm decision-making. The renewal cycle is highly persistent: of the 30,000 observed changes to firm's scheduled renewal, indicating an update to their contract, only 12% happen off-schedule, confirming that firms rarely renegotiate their enterprise agreement outside the renewal cycle. This rigidity shapes how firms adjust their GitHub usage. In the pre-Copilot period, 33% of renewal months involve a change in Enterprise seat counts, compared with 10% of non-renewal months, as shown in Table 2. Together, these patterns confirm that enterprise contract renewals represent discrete procurement events rather than a continuous decision margin: firms wait for the scheduled renewal to adjust both the terms of their agreement and the scale of their usage.

Table 2: GitHub Enterprise Seat Changes at Renewal vs. Off-Renewal (Before 2023)

	Renewal months (± 1)	Non-renewal months
Firm-months	27,215	249,747
Months with seat change	4,266	25,912
Pr(seat change)	0.16	0.10
Within Seat Changes		
Share increases	0.59	0.82
Share decreases	0.41	0.18

The key identifying assumption is that the timing of a firm's renewal cycle, which depends on when the firm first signed its GitHub Enterprise contract and the dates of subsequent negotiations, is as good as random with respect to firm employment outcomes, conditional on firm and time fixed effects. Under this assumption, renewal timing serves as an instrument for the timing of Copilot adoption: it shifts *when* firms adopt, but does not directly affect employment except through its effect on adoption. We emphasize that we are not instrumenting for whether a firm adopts Copilot, but rather for the *timing* of adoption. The annual nature of the renewal cycle mean that nearly all firms in our sample eventually have the opportunity to adopt. The instrument exploits variation in how quickly they do so.

We construct our instrument from a firm's position in its renewal cycle, measured as months to next scheduled renewal (MTNR). MTNR is the predetermined date, set at the time of the last contract negotiation, at which the current contract expires and renegotiation is triggered. To avoid simultaneity, since Copilot adoption may itself trigger a contract renegotiation that updates the renewal date, we compute MTNR using the *lagged* renewal date: the next scheduled renewal date on file as of the prior month. For instance, let MTNR_i denote the number of months between January 2023 and firm i's next scheduled renewal date as recorded in December 2022. We define:

$$Z_i(W) = \mathbf{1}\left[|\text{MTNR}_i| \leqslant W\right], \tag{1}$$

where W is the window width in months. The first approach exploits this cross-sectional variation at the time of launch: firms with $Z_i = 1$ are firms who had a contract renewal coming up within W months after the Copilot launch, giving them an earlier opportunity to adopt. Firms with $Z_i = 0$ are "late renewal" firms whose next renewal was further away. We present results for $W \in \{1, 2, 3\}$ months.

4.3 Estimating Equations

Cross-sectional approach. The cross-sectional instrument Z_i generates a single shift in the cost of adoption at the time of Copilot's launch. The first stage is:

$$D_{it} = \alpha_i + \lambda_t + \pi \cdot Z_i \times \text{Post}_t + \varepsilon_{it}, \tag{2}$$

where $D_{it} = \mathbf{1}(\text{firm } i \text{ has adopted Copilot by month } t)$ is a cumulative adoption indicator, α_i and λ_t are firm and year-month fixed effects, and $\text{Post}_t = \mathbf{1}(t \geqslant \text{January 2023})$. The coefficient π measures the differential increase in cumulative adoption for early-renewal firms after Copilot's launch.

The reduced form replaces the single post indicator with event-time interactions to trace out the dynamic relationship between the instrument and employment:

$$Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k \, \mathbf{1}[t - \text{Launch} = k] \times Z_i + \varepsilon_{it}, \tag{3}$$

where k indexes months relative to the Copilot launch. Because the Copilot launch date is common to all firms, there are no staggered-adoption complications: the coefficients β_k trace out the differential path of outcomes for early- versus late-renewal firms in calendar time. We show that these instruments have a strong first-stage relationship with Copilot adoption and that our employment results are robust to various definitions of close to renewal.

The 2SLS estimand combines the first stage and reduced form. The system is:

$$\text{First stage:} \quad D_{it} = \alpha_i + \lambda_t + \pi \cdot Z_i \times \text{Post}_t + \varepsilon_{it}, \tag{4}$$

$$\text{Second stage:} \quad Y_{it} = \alpha_i + \lambda_t + \beta \cdot \hat{D}_{it} + u_{it}, \tag{5}$$

where $\hat{D}_{it}$ is the predicted value from the first stage. The coefficient β is the local average treatment effect of Copilot adoption for compliers, firms whose adoption timing is shifted by renewal proximity, over the post-period. The estimating sample is a balanced firm-by-month panel, which we aggregate to the quarterly level for outcomes where monthly observations are sparse. All specifications include firm fixed effects, which absorb time-invariant differences across firms, and calendar-month (or calendar-quarter) fixed effects, which absorb common shocks such as interest rate changes or sector-wide hiring cycles. Standard errors are clustered at the firm level throughout.

4.4 Identifying Assumptions

Our estimation strategy relies on three assumptions (Imbens and Angrist, 1994; Angrist and Imbens, 1995). Relevance requires that renewal proximity shifts the probability of Copilot adoption; we document this in Section 5.2. The exclusion restriction requires that renewal timing affects employment only through Copilot adoption, conditional on firm and time fixed effects, which we assess in Section 5.3. Monotonicity requires that renewal proximity weakly increases adoption for all firms; this is economically plausible because the renewal window reduces the transaction cost of adding Copilot, but does not increase the cost of doing so. We also show in Section 5.10 that the first stage is strong across observable subgroups. Under these assumptions, the estimates identify the causal effect for compliers, firms at the margin of adoption for whom out-of-cycle procurement frictions are binding, whom we characterize in Section 5.9.

Figure 2: Discrete Hazard of First Copilot Adoption by Months to Next Scheduled Renewal

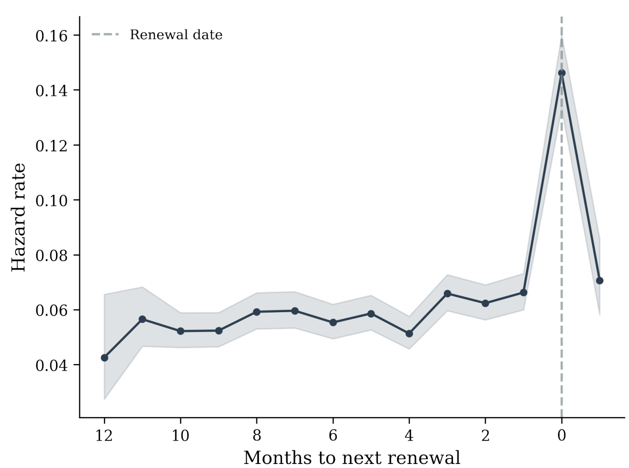


Figure 2 provides visual evidence for the instrument. For each position in the renewal cycle, we compute the discrete hazard rate of Copilot adoption: the share of at-risk firms who have not yet adopted Copilot in any month after its launch that adopt, by renewal cycle position. Renewal cycle position is shown as months to next scheduled renewal, so moving from left to right is getting closer to an upcoming contract renewal date. The hazard spikes sharply at the next scheduled renewal date (MTNR = 0), approximately tripling to 14% from a baseline of roughly 5%. Away from renewal, the hazard is relatively flat across the cycle, confirming that it is the renewal event itself, not secular trends in cycle position, that drives adoption timing.

5 Results

5.1 OLS Event Studies

We begin by documenting the association between Copilot adoption and employment outcomes using OLS event studies that compare adopters to non-adopters relative to the Copilot launch date:

$$Y_{it} = \alpha_i + \lambda_t + \sum_k \beta_k \mathbf{1}[t - \text{Launch} = k] \times \text{EverAdopt}_i + \varepsilon_{it}, \tag{6}$$

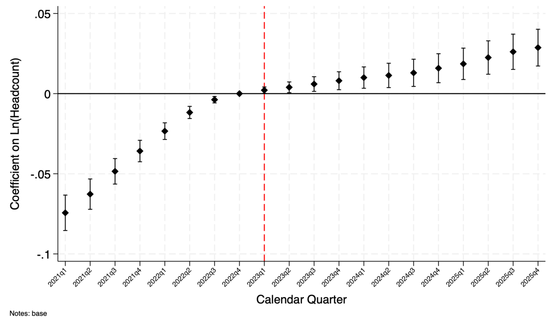
where EverAdopt_i is an indicator for firms that adopt Copilot at any point during the sample period. The coefficients β_k measure the differential outcome for adopters versus never-adopters at each month k relative to the Copilot launch, after absorbing firm means and common year-month shocks. These estimates are descriptive: they reflect both the causal effect of Copilot adoption and selection into treatment.

Figure 3 presents the results for four outcomes. Panels (a) and (b) show log headcount and log headcount of software engineers. Adopters exhibit a strong positive pre-trend: they were growing faster than non-adopters throughout the two years before Copilot launched. Post-launch, the coefficients are small and statistically insignificant, hovering near zero. This pattern could be consistent with an Ashenfelter's dip, where firms that adopt Copilot are those whose headcount was already converging toward non-adopters, and the flat post-period may reflect mean reversion rather than a causal effect (Ashenfelter, 1978).

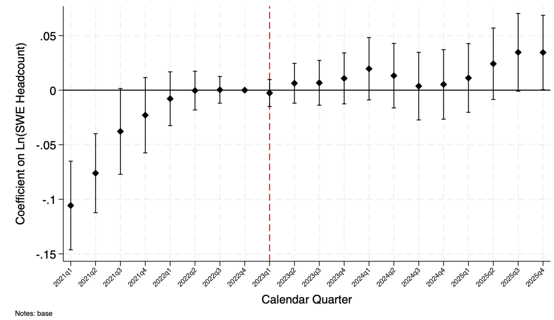
Panel (c) shows the hiring rate. Adopters had persistently higher hiring rates than non-adopters throughout the pre-period, suggesting prior growth is correlated with prior hiring intensity. Panel (d) shows the separation rate. Adopters have lower separation rates than non-adopters throughout the sample, although these are much noisier.

Two features of these patterns are important for interpreting the IV results that follow. First, the pre-trends are large and statistically significant for headcount and hiring, confirming that a naïve comparison of adopters and non-adopters is confounded by selection. Even controlling for six-digit NAICS×year, firm size quartile×year, and country×year interactions, the pre-trends persist, suggesting these patterns are not fully explained by differential industry or size trends. Second, the direction of the omitted variable bias is ambiguous. On one hand, in 2023 adopters could be mean-reverting from a temporary growth spurt. If so, their post-period employment was slowing regardless of Copilot adoption and OLS estimates will be biased downwards. On the other hand, adoption could be positively correlated with continued faster growth if firms that are willing to adopt new technologies early are better managed firms. In this case, adopters would have continued to grow anyways and the OLS coefficients will be upwardly biased. The data cannot distinguish between these stories, which is precisely why an instrument is needed.

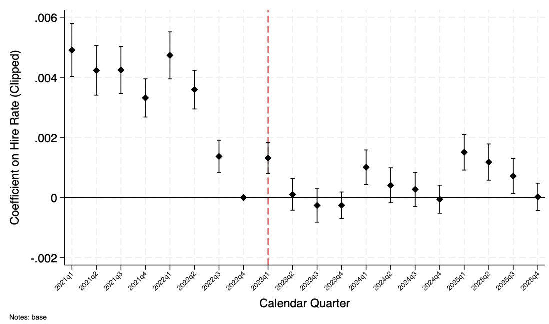
Figure 3: OLS Event Study: Adopters vs. Never-Adopters Relative to Copilot Launch



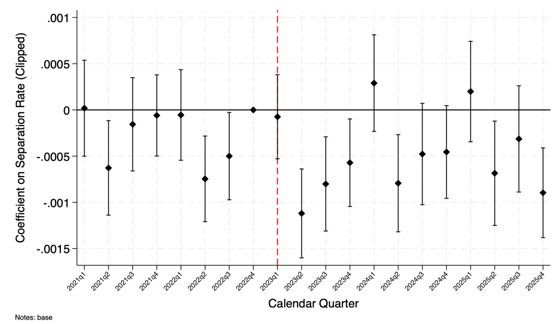
(a) Log Headcount



(b) Log SWE Headcount



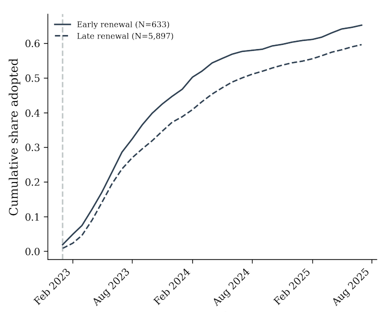
(c) Hire Rate



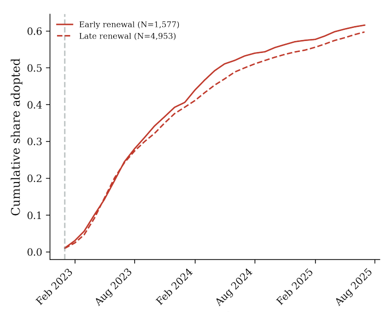
(d) Separation Rate

5.2 First Stage: Renewal Timing and Copilot Adoption

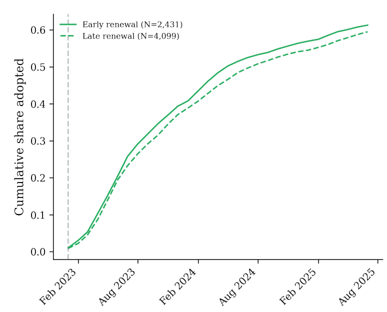
Figure 4: Cumulative Copilot Adoption by Renewal Proximity at Launch



(a) W = 1



(b) W = 2



(c) W = 3

Figure 4 provides visual evidence for the first stage. Each panel compares firms whose renewal date fell within W months of the Copilot launch (solid) to firms further from renewal (dashed), for

$W \in \{1, 2, 3\}$. Firms closer to renewal at launch adopt Copilot faster: the gap between early- and late-renewal firms is largest for the narrowest window ($W = 1$) and narrows as the window widens.

Table 3 reports the first-stage estimates. Columns 1 and 2 present the cross-sectional specification, in which the dependent variable is cumulative Copilot adoption A_{it}, following equation 2. Firms within one month of renewal at the Copilot launch are 12 percentage points more likely to have adopted in any given post-period month, relative to a dependent variable mean of 20.0%, with an effective F-statistic of 64. Within two months of renewal, firms are 10% more likely to have adopted in any given post-period month, with an effective F-statistic of 84. Similar results are true for a window of three months. All instruments are well above the Stock and Yogo (2002) threshold of 16.38 for a single instrument. Column 4 presents the time-varying specification, in which the dependent variable is the monthly adoption hazard a_{it} among at-risk firms—those that have not yet adopted. Being in the month of renewal increases the adoption hazard by 2.2 percentage points relative to a base rate of 2.4%, with an effective F-statistic of 96.1. For the remainder of the paper, we use the cross-sectional instrument with $W = 2$ as our preferred specification, but show that results are robust to alternative window definitions in the appendix.

Table 3: First-Stage Estimates

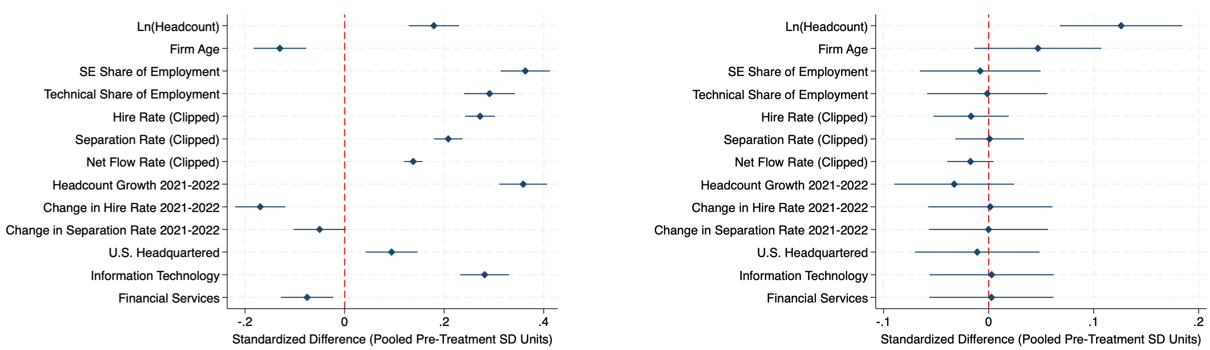
	(1) Cross-Sectional (W=1)	(2) Cross-Sectional (W=2)	(3) Cross-Sectional (W=3)	(4) Time-Varying
$Z_i \times$ Post	0.1240***			
	(0.0155)			
$Z_i \times$ Post		0.0958***		
		(0.0104)		
$Z_i \times$ Post			0.1086***	
			(0.0090)	
$Z_{it} \times$ Post				0.0222***
				(0.0023)
DV Mean	0.2071	0.2071	0.2071	0.0244
F-stat	63.81	84.32	144.45	96.12
Observations	364,841	364,841	364,841	154,157
Clusters	5,981	5,981	5,981	6,070
Firm FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Within R^2	0.0047	0.0060	0.0103	0.0011

5.3 Balance Checks

For the renewal cycle to be a valid instrument, the timing of a firm's renewal must be uncorrelated with firm characteristics that independently affect employment outcomes, conditional on firm and time fixed effects. We assess this by regressing standardized pre-treatment firm characteristics, on group indicators, and plotting the resulting coefficients.

Figure 5 presents these balance checks for firm characteristics spanning employment levels, workforce composition, pre-period trends, geography, and industry. Panel (a) confirms the pattern documented in the summary statistics and OLS event studies: adopters differ significantly from non-adopters across most characteristics. Panel (b) repeats the exercise for the cross-sectional instrument, comparing firms with $Z_i = 1$ to firms with $Z_i = 0$ using a window of $W = 2$ months. In contrast to Panel (a), none of the fourteen coefficients is statistically distinguishable from zero at the 10% level, except for headcount, and all differences are small in magnitude. Because the main regressions contain firm and year-month fixed effects, which absorb level differences across firms, differences in pre-period trends are a greater threat to identification that pre-period level differences. The coefficients on pre-period headcount growth, changes in hiring rates, and changes in separation rates are all close to zero, with no evidence of differential pre-trends between early- and late-renewal firms. The lack of significant differences across this broad set of characteristics supports the idea that renewal timing is plausibly exogenous with respect to employment outcomes.

Figure 5: Balance of Pre-Treatment Characteristics



(a) Adopter vs Non-Adopter

(b) $Z_i = 1$ vs $Z_i = 0$ ($W = 2$)

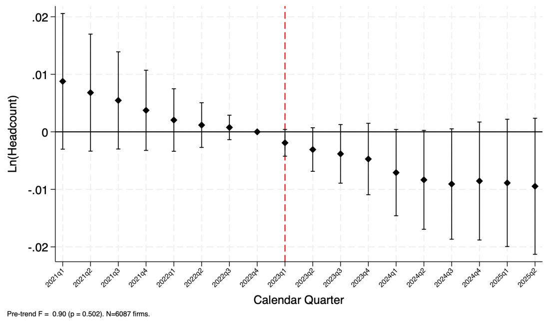
5.4 Reduced Form Employment Effects

Having established that renewal timing predicts Copilot adoption and is balanced on pre-treatment firm characteristics, we now examine whether renewal timing affects employment outcomes. These reduced-form estimates are intention-to-treat effects: the differential change in outcomes for early-renewal firms relative to late-renewal firms after Copilot's launch.

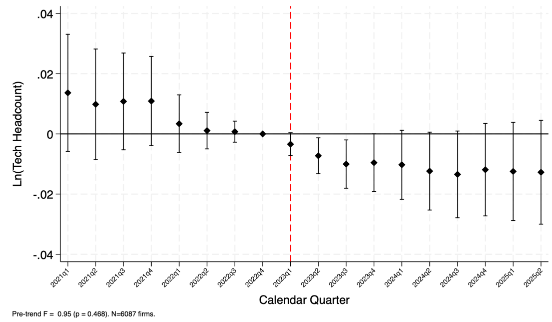
Figure 6 presents the quarterly reduced-form event study, plotting $\hat{\beta}_q$ from Equation (3). These figures are the direct analog of the OLS event study in Figure 3, replacing the endogenous EverAdopt_i indicator with the instrument Z_i. Whereas the OLS specification exhibited large, significant pre-trends in headcount, the reduced-form coefficients are centered near zero throughout the pre-period for all four outcomes, confirming that early- and late-renewal firms were on parallel employment trajectories before Copilot launched.

Post-launch, point estimates drift negative for total, SWE, and technology worker headcount, but confidence intervals include zero at nearly every quarter. The imprecision is not surprising: the reduced form outcomes look at the stock of workers employed by a firm, which is about 9,000 workers on average. Even dramatic changes in hiring practices would only slowly impact such a large stock of workers. However, panel B suggests a drop in the number of domestic software engineers.

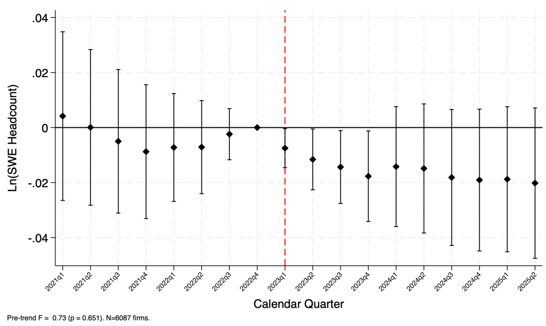
Figure 6: Reduced Form Event Study: Employment by Role



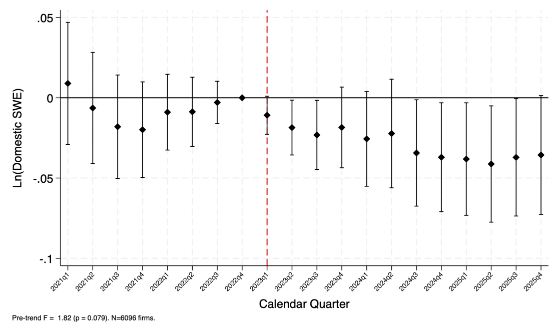
(a) Log Headcount



(b) Log Tech Headcount



(c) Log SWE Headcont



(d) Log Domestic SWE Headcount

Table 4 presents the main employment results, comparing OLS and 2SLS estimates of the effect of Copilot adoption on total headcount and technology worker headcount. Columns 1–2 report OLS estimates with the endogenous adoption indicator D_{it} entered directly. Firms that have adopted Copilot have 8.2% higher total headcount and 10.3% higher technical headcount than non-adopters, conditional on firm and month fixed effects.

Columns 3–4 instrument cumulative adoption with the cross-sectional renewal instrument $Z_i(W{=}2)\times$ Post$_t$. The effective F of 69 exceeds the Montiel Olea and Pflueger (2013) critical value of 37.4 for 5% worst-case size distortion. The 2SLS estimates differ markedly from OLS. The point estimate on total headcount falls from 0.082 under OLS to 0.014 under 2SLS; for technical headcount, from 0.103 to 0.007. Neither 2SLS estimate is statistically significant, with standard errors of 0.080 and 0.109, respectively.

The divergence between OLS and 2SLS is consistent with the selection documented above: Copilot adopters are positively selected on prior growth, and the positive OLS estimates reflect this upward bias rather than a causal effect. The 2SLS estimates on headcount are noisy: the 95% confidence interval for total headcount is approximately $[-0.14, +0.17]$, so we cannot rule out headcount swings below that in either direction.

Figure 4: OLS and 2SLS Estimates: Effect of Copilot Adoption on Employment

	Ln(Headcount) (1) OLS	Ln(Tech Headcount) (2) OLS	Ln(Headcount) (3) 2SLS (W=2)	Ln(Tech Headcount) (4) 2SLS (W=2)
Copilot Adopted	0.0820***	0.1032***	0.0140	0.0067
	(0.0053)	(0.0075)	(0.0801)	(0.1090)
Observations	591312	591310	580157	580155
DV Mean	7.4796	4.9451	7.4653	4.9312
Effective F-stat			69.34	69.35
Clusters	6,096	6,096	5,981	5,981
Firm FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes

Together, these results rule out large aggregate employment effects: Copilot adoption does not produce dramatic declines in either total headcount or technical employment. The firms in our sample are large multinationals with geographically dispersed workforces spanning a range of seniority levels. Stable aggregate headcount may therefore mask reallocation along other margins. In the next sections, we examine two such margins: where firms hire (geography) and whom they hire (seniority).

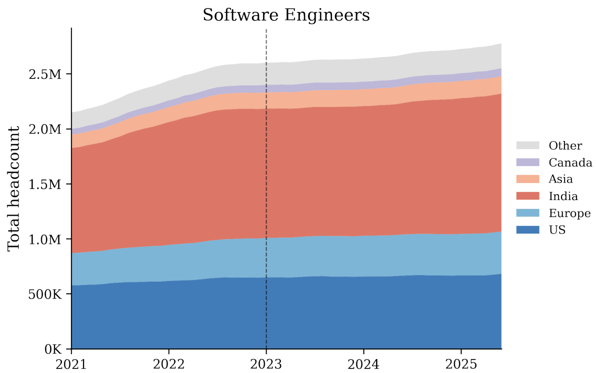
5.5 Geographic Composition

Technology has historically reshaped the geographic distribution of work. Reductions in communication and coordination costs enabled firms to fragment production across locations, shifting tasks to lower-cost labor markets (Grossman and Rossi-Hansberg, 2008; Blinder, 2006). Software development has been at the center of this trend: programming is a tradable task that requires no physical presence, and the global IT services industry grew rapidly as firms offshored development work, particularly to India (Blinder, 2006).

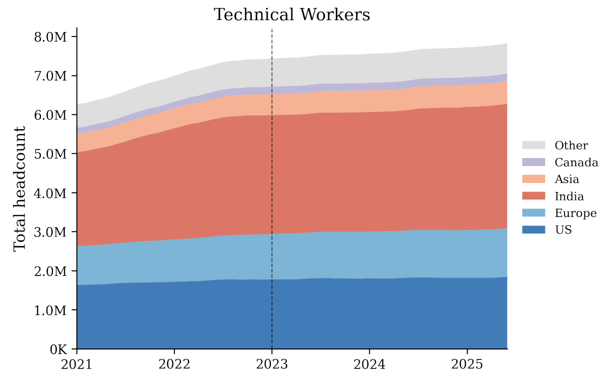
AI coding tools could shift the location of technical work in either direction. If the tools disproportionately raise the productivity of onshore developers, or automate some of the tasks done by offshore developers, firms may consolidate work domestically. Several channels push in the opposite direction. First, if the tools compress quality differences across developers by bringing lower-cost workers closer to the frontier, the penalty for offshoring falls. Second, by standardizing code style, documentation, and development practices, AI assistants may reduce the coordination costs of managing geographically dispersed teams. Both channels lower the effective cost of offshore relative to onshore labor. If workers in the same job do different tasks, the same technology could complement workers in one country and substitute for workers in another (Autor and Thompson, 2025).

Figure 7 shows the geographic distribution of employment for the two role categories most directly affected by AI coding tools. Panel (a) plots total software engineer headcount by country group over time. The U.S. accounts for the largest share, followed by India and Europe, illustrating the geographic distribution of software engineering employment in our sample. Panel (b) shows the same decomposition for data and analytics workers (data scientists and data analysts). Both panels reveal substantial international dispersion of technical talent, emphasizing the scope for geographic reallocation in response to AI-driven productivity changes.

Figure 7: Geographic Distribution of Employment by Role Category



(a) Software Engineers



(b) Technology Workers

Table 5 presents reduced-form and 2SLS estimates of the effect of Copilot adoption on the domestic share of employment for all workers, software engineers, and technical workers. We define domestic relative to the firm's headquarters country: for a US-headquartered firm, domestic workers are those based in the United States; for a UK-headquartered firm, domestic workers are those based in the United Kingdom.

The domestic share of the overall workforce declines by 0.38 percentage points in the reduced form ($p = 0.001$); the 2SLS implies Copilot adoption reduces the domestic share by 4.3 percentage points among compliers, from a base of 70%. The effect is about double for software engineers (-0.62 pp, $p = 0.013$), with the technical workers (-0.37 pp, $p = 0.00$) category more broadly. The decline is not concentrated in any single international destination: country-specific share regressions for India, Europe, Asia, Canada, and Israel are all individually insignificant.

Table 5: Effect of Copilot Adoption on Domestic Share of Employment

	All Dom Share (1)	SWE Dom Share (2)	Tech Dom Share (3)
Panel A: Reduced Form			
$Z_i \times$ Post	-0.0038***	-0.0062**	-0.0037**
	(0.0011)	(0.0025)	(0.0019)
Panel B: 2SLS			
Copilot Adoption	-0.0431***	-0.0705**	-0.0423**
	(0.0131)	(0.0291)	(0.0211)
Eff. F-stat	67.6	67.6	67.6
DV Mean	0.701	0.698	0.704
Firm FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Clusters	5,959	5,959	5,959

Notes: Each column reports effects on the domestic share of employment within the indicated role category (domestic headcount / total headcount in that role). The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The geographic recomposition primarily operates through hiring flows, not separations. Table 6 reports the domestic share of quarterly new hires overall, for software engineers and technical workers. The domestic share of SWE inflows declines by 1.0 percentage point ($p = 0.004$); the 2SLS implies Copilot adoption reduces the domestic share of SWE hiring by 11.3 percentage points among compliers. The overall domestic share of inflows also declines (-0.38 pp, $p = 0.065$), while the tech-specific effect goes in the same direction but is not statistically significant. Together, this suggests that changes in the domestic share of software engineers are moderated by the other technology roles. Appendix Table A.3 confirms that outflows do not change—the geographic shift is driven entirely by the reallocation of new hires, not by increased separations.

Table 6: Effect of Copilot Adoption on Domestic Share of Quarterly Inflows

	All (1)	SWE (2)	Tech (3)
Panel A: Reduced Form			
$Z_i \times$ Post	-0.0038*	-0.0102***	-0.0045
	(0.0021)	(0.0036)	(0.0031)
Panel B: 2SLS			
Copilot Adoption	-0.0417*	-0.1127***	-0.0498
	(0.0226)	(0.0407)	(0.0339)
Eff. F-stat	66.8	65.8	66.4
DV Mean	0.715	0.697	0.706
Firm FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Clusters	5,959	5,943	5,959

Notes: Each column reports effects on the domestic share of quarterly inflows within the indicated role category (domestic inflows / total inflows). Data collapsed to firm-quarter. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.6 Seniority Composition of Hiring

A second margin where reorganization can occur is the experience of the workers hired. Figure 8 plots the average firm-level junior share of employment, the fraction of workers that are either entry or junior level, by role category. Descriptively, the junior share of software engineers declines steadily from 73% to 68% over the sample period, with the decline accelerating after Copilot's launch, with a similar pattern amount other non-SWE technology workers. The junior share across all workers is flat, indicating that the seniority shift is specific to technical roles.

Figure 8: Average Firm-Level Junior Share of Employment by Role Category

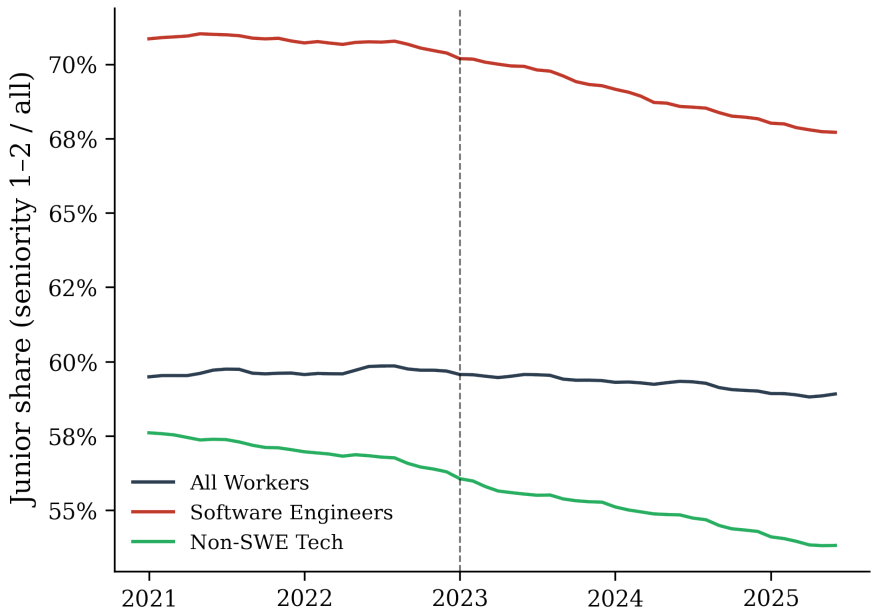


Table 7 tests whether Copilot adoption is associated with a slowdown in junior hiring. Junior inflows (seniority levels 1–2) decline across technical and SWE categories: by 3.3 workers per quarter for SWE ($p = 0.042$), 7.8 for technical workers ($p = 0.054$), and 21.0 for all workers ($p = 0.088$). Senior inflows, defined as levels above 5 and above, who focus primarily on management, technical architecture and high level decision making, are unchanged in every specification. The 2SLS results are noisier, but support the large fall in junior SWE inflows. The reduction in new hiring falls entirely on junior workers, consistent with AI tools performing tasks that would otherwise be allocated to entry-level engineers. Unlike the geographic effects, the seniority shift is broad-based: it is not concentrated among IT Services firms or U.S.-headquartered firms (Appendix Table A.5).

Table 7: Effect of Copilot Adoption on Quarterly Inflows by Seniority

	All Workers		SWE		Tech	
	Junior (1)	Senior (2)	Junior (3)	Senior (4)	Junior (5)	Senior (6)
Panel A: Reduced Form						
$Z_i \times$ Post	-20.97*	-0.99	-3.26**	-0.05	-7.81*	-0.35
	(12.28)	(1.76)	(1.61)	(0.23)	(4.05)	(0.28)
Panel B: 2SLS						
Copilot Adoption	-227.35*	-10.71	-35.39**	-0.50	-84.62*	-3.76
	(134.75)	(19.17)	(17.86)	(2.53)	(44.60)	(3.09)
Eff. F-stat	68.7	68.7	68.7	68.7	68.7	68.7
DV Mean	294.4	57.4	14.9	1.2	36.0	4.3
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	5,981	5,981	5,981	5,981	5,981	5,981

Notes: Each column reports effects on the count of quarterly inflows at the indicated seniority level and role category. Junior = seniority levels 1–2; Senior = seniority levels 5–7. Data collapsed to firm-quarter. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.7 Occupational Composition

If AI coding tools were displacing software engineers, we would expect the SWE share of employment to fall. Table 8 reports reduced-form and 2SLS estimates for the share of employment in each role category. The SWE, tech, and non-tech shares are all unchanged: every reduced-form estimate is close to zero with t-statistics below 0.5. Firms still employ the same mix of occupations after adopting Copilot. The reorganization documented above operates within the existing occupational structure, changing *where* firms hire and *at what seniority level*, but not *which roles* they fill.

Table 8: Effect of Copilot Adoption on Occupational Composition

	SWE Share (1)	Tech Share (2)	Non-Tech Share (3)
Panel A: Reduced Form			
$Z_i \times$ Post	0.0000	0.0002	-0.0002
	(0.0005)	(0.0006)	(0.0006)
Panel B: 2SLS			
Copilot Adoption	0.0003	0.0018	-0.0018
	(0.0052)	(0.0065)	(0.0065)
Eff. F-stat	69.3	69.3	69.3
DV Mean	0.058	0.130	0.870
Firm FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Clusters	5,981	5,981	5,981

Notes: Each column reports effects on the share of employment in the indicated role category. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5.8 Heterogeneity

The geographic reorganization is not uniform across firms. Table 9 decomposes the effect on the domestic SWE share by industry (IT Services vs. non-IT) and firm headquarters location (U.S. vs. non-U.S.). The effect is concentrated among U.S.-headquartered IT Services firms, which reduce their domestic SWE share by 1.1 percentage points ($p < 0.10$). The remaining three cells are small and statistically insignificant, suggesting that the aggregate domestic share decline is driven primarily by U.S. IT firms shifting SWE positions offshore.

Table 9: Heterogeneity: Effect on Domestic SWE Share by Industry and HQ Location

	US-HQ	Non-US-HQ
IT Services	-0.0108*	0.0054
	(0.0062)	(0.0065)
Non-IT	-0.0030	-0.0016
	(0.0044)	(0.0035)
Firm FE	Yes	Yes
Month FE	Yes	Yes
Clusters	5,911	

Notes: Each cell reports the reduced-form coefficient on $Z_i \times$Post interacted with industry and HQ location indicators. Dependent variable: domestic share of SWE employment. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Unlike the geographic results, there is no significant heterogeneity in junior hiring effects across industry and headquarters groups (F-test $p > 0.4$ for all pairwise comparisons). While the effects are negative for all four subgroups in Appendix Table A.5, and the largest effect is in the IT $\times$ US cell (-3.9, $p = 0.016$), we cannot reject that the declines are equal in magnitude.

There is no heterogeneity by firm size: above and below median size firms reduce their domestic share at identical rates ($p = 0.90$; Appendix Table A.6). However, the firms in our sample are extremely large, so even the below median size firms still employ thousands of workers.

5.9 Complier Characterization

The 2SLS estimates identify the local average treatment effect for compliers: firms whose Copilot adoption timing is shifted by renewal proximity. We characterize this population using the κ-weighting method of Abadie (2003), computing $E[X_i \mid \text{complier}] = E[\kappa_i X_i]/E[\kappa_i]$ where κ_i is the Abadie compliance weight constructed from the time-varying treatment D_{it} and instrument $Z_i \times \text{Post}_t$ in the panel.

Table 10 reports pre-period means of firm characteristics for the full sample, always-takers, compliers, and never-takers. Compliers are younger firms (41 years vs. 47 in the full sample) with slightly smaller headcounts (log headcount 7.32 vs. 7.41). They are of similar technically intensity to the average firm: the SWE share of employment among compliers is 6.7%, compared to 5.8% in the full sample, and the broader technical share is similar. Compliers are somewhat more likely to be U.S.-headquartered (57% vs. 54%) and to be in the IT industry. Hiring and separation rates are similar across groups.

Compliers have shorter average renewal cycles (8.7 months vs. 12.1 months in the full sample). This difference is partly mechanical: the instrument selects firms whose renewal falls close to the Copilot launch date, and firms with shorter cycles, on average, have a greater likelihood of being closer to renewal.

Always-takers (firms that adopt in all post-periods even with $Z_i = 0$) are the largest firms in the sample (log headcount 8.03), consistent with large enterprises adopting Copilot irrespective of renewal timing. Never-takers ($N = 1{,}076$) are similar to the full sample across most characteristics, suggesting that non-adoption reflects firm-level preferences rather than observable differences.

Table 10: Complier Characteristics

	Full Sample	Always-Takers	Compliers	Never-Takers
Ln(Headcount)	7.41	8.03	7.32	7.27
SWE Share	0.0577	0.0628	0.0673	0.0387
Tech Share	0.1285	0.1288	0.1383	0.1051
Hire Rate	0.0239	0.0264	0.0259	0.0203
Sep Rate	0.0177	0.0194	0.0186	0.0162
Firm Age	46.91	52.15	41.32	52.85
Ln(GH Seats)	4.41	5.06	4.64	3.59
US HQ	0.5370	0.4754	0.5720	0.4667
IT Industry	0.2448	0.2623	0.2788	0.1751
Finance	0.1380	0.0820	0.1401	0.1537
Cycle Length	12.07	.	8.69	12.13
Complier share		0.3672		
N (firms)		6,096		

Notes: Complier means computed using κ-weights from Abadie (2003) with the panel first stage (D_{it} on $Z_i \times \text{Post}_t$, $W = 2$). Always-takers: $D_{it} = 1$ in all post-periods with $Z_i = 0$. Never-takers: $D_{it} = 0$ in all post-periods with $Z_i = 1$. Complier share $= E[\kappa] = 0.36$.

5.10 Robustness

We explore the sensitivity of our results to alternative instrument windows, compute the first stage strength in subgroups used in our heterogeneity analysis and test for complier stability at alternative window definitions.

Instrument window sensitivity. Our main results use an instrument window of $W = 2$ months. Appendix Table A.7 reports the domestic share results for $W \in \{1, 2, 3\}$. All four domestic share outcomes are significant at $W = 1$ and $W = 2$. The reduced-form coefficients declines monotonically as the window widens, -from -0.76 pp ($W = 1$) to -0.31 pp ($W = 3$) for the overall domestic share, consistent with wider windows including firms less strongly affected by renewal timing. The $W = 1$ specification, which uses the most conservative instrument definition, produces the largest and most significant effects. The IT $\times$ HQ heterogeneity pattern is also stable across windows: the US IT cell ranges from -5.2 pp ($W = 1$) to -4.7 pp ($W = 3$), and all four cells maintain their sign and significance (Appendix Table A.8).

The quarterly inflow results are similarly robust. The domestic share of SWE inflows is significant at all three windows (-1.67 pp at $W = 1$, -1.02 pp at $W = 2$, -0.71 pp at $W = 3$; Appendix Table A.9). Junior SWE inflows decline significantly at all windows, with the $W = 1$ estimate (-7.3 workers per quarter, $p < 0.01$) more than double the $W = 2$ estimate.

Pre-trends. The domestic share outcomes used in the main results all pass the joint pre-trend F-test at the 5% level (Appendix Table A.10).

First stage in subgroups. The first stage effective F is above 14 in seven of eight heterogeneity subgroups (Appendix Table A.11). The exception is IT headquartered outside the US, where the renewal instrument does not predict Copilot adoption, likely due to the small number of Non-US IT firms (around 500). The null coefficient for this cell in Table 9 is therefore uninformative and should not be interpreted as evidence of no effect.

Complier stability. Complier characteristics are stable across instrument windows (Appendix Table A.12). The complier share ranges from 0.37 to 0.39, and complier means for key characteristics such as headcount, SWE intensity, firm age, and headquarters location, are similar across windows.

6 Conclusion

Advances in large language models have opened a broad set of possibilities for automating knowledge work. This paper provides early empirical evidence on the employment effects of generative AI tools for software engineers. Using quasi-exogenous variation in the timing of enterprise procurement cycles, we find that adoption of GitHub Copilot among large, technology-intensive multinationals reduces the domestic share of software engineering employment and decreases the hiring of entry-level and junior SWEs. These results are driven by changes in hiring inflows rather than separation rates, and are largest among US-headquartered firms in the IT services industry. We do not find evidence of large declines in total headcount, software engineering employment, or other technical roles.

Our analysis is subject to some qualification and raises some additional questions.

First, our sample consists of large multinational firms adopting a particular type of AI tool. These firms employ hundreds of technical workers across multiple countries and can more readily shift the geographic distribution of work than smaller or purely domestic employers. The effects we document may not generalize to firms without existing international operations or to other classes of AI tools.

Second, our sample period covers the first generation of AI coding assistants, which operated primarily as suggestion-based autocomplete tools that can write small sections of code. Beginning in late 2025, a new generation of agentic coding tools emerged that can autonomously execute complicated, multi-step programming tasks without human intervention. These tools may have qualitatively different effects on the demand for software engineers. Our estimates capture the short and medium-run response to the earlier technology.

Third, we do not observe wages. The employment reallocation we document could be accompanied by wage increases for retained domestic workers, wage compression across geographies, or both. More broadly, we estimate partial equilibrium effects for adopting firms; general equilibrium responses, including wage adjustments across the broader labor market, remain unmeasured.

Finally, our findings raise questions about the nature of automation in white-collar occupations. The firms in our sample did not dramatically reduce their technical workforce; instead, they reorganized it without dramatic changes to the composition of their workforce. To understand how this technology affects workers, future research must trace the outcomes of the workers who would have been hired absent these tools, and whether they find comparable employment elsewhere, shift to adjacent occupations, or face prolonged displacement, as well as how the jobs of workers who are retained change. Given the speed of advancement of generative AI tools, these topics are worthy of further exploration.

References

Abadie, Alberto. 2003. "Semiparametric instrumental variable estimation of treatment response models." *Journal of Econometrics*, 113(2): 231–263.

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics.* Vol. 4, , ed. David Card and Orley Ashenfelter, 1043–1171. Amsterdam, NL:Elsevier.

Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. 2022. "Artificial intelligence and jobs: Evidence from online vacancies." *Journal of Labor Economics*, 40(S1): S293–S340.

Acemoglu, Daron, Gary Anderson, David Beede, Catherine Buffington, Eric Childress, Emin Dinlersoz, Lucia Foster, Nathan Goldschlag, John Haltiwanger, Zachary Kroff, et al. 2023. "Advanced technology adoption: selection or causal effects?" Vol. 113, 210–214, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90(430): 431–442.

Ashenfelter, Orley. 1978. "Estimating the Effect of Training Programs on Earnings." *The Review of Economics and Statistics*, 60(1): 47–57.

Autor, David, and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association*, 23(4): 1203–1271.

Autor, David H. 2015. "Why are there still so many jobs? The history and future of workplace automation." *Journal of economic perspectives*, 29(3): 3–30.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics*, 118(4): 1279–1333.

Aye, Gareth Ari, Seohyun Kim, and Hongyu Li. 2021. "Learning autocompletion from real-world datasets." 131–139, IEEE.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics*, 151: 103745.

Baird, M, Mar Carpanelli, Brian Xu, and Kevin Xu. 2024. "Early Evidence on the Impact of GitHub Copilot on Labor Market Outcomes for Software Engineers." Working Paper.

Blinder, Alan S. 2006. "Offshoring: The next industrial revolution?" *Foreign Affairs*, 85(2): 113–128.

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the coal mine? six facts about the recent employment effects of artificial intelligence." *Digital Economy.*

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly Journal of Economics*, 140(2): 889–942.

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. "Sparks of Artificial General Intelligence: Early experiments with GPT-4."

Cui, Z. K., M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz. 2025. "The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers." Available at SSRN 4945566.

David, Paul A. 1990. "The dynamo and the computer: an historical perspective on the modern productivity paradox." *The American economic review*, 80(2): 355–361.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, Francois Candelon, and Karim Lakhani. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." Harvard Business School Working Paper 24-013.

Dillon, Eleanor W., et al. 2025. "Shifting Work Patterns with Generative AI." National Bureau of Economic Research Working Paper w33795.

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." *The Quarterly Journal of Economics*, 139(3): 1879–1939.

Grossman, Gene M., and Esteban Rossi-Hansberg. 2008. "Trading tasks: A simple theory of offshoring." *American Economic Review*, 98(5): 1978–1997.

Hampole, Menaka, Dimitris Papanikolaou, Lawrence D. W. Schmidt, and Bryan Seegmiller. 2025. "Artificial Intelligence and the Labor Market." National Bureau of Economic Research w33509.

Hicks, John R. 1932. *The Theory of Wages.* London:Macmillan.

Humlum, Anders, and Emilie Vestergaard. 2025. "Large language models, small labor market effects." National Bureau of Economic Research.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467–475.

Jiang, Shaokang, and Michael Coblenz. 2024. "An analysis of the costs and benefits of autocomplete in ides." *Proceedings of the ACM on Software Engineering*, 1(FSE): 1284–1306.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics*, 107(1): 35–78.

Maslej, Nestor, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, et al. 2025. "Artificial intelligence index report 2025." *arXiv preprint arXiv:2504.07139*.

McElheran, Kristina, Mu-Jeung Yang, Zachary Kroff, and Erik Brynjolfsson. 2025. *The Rise of Industrial AI in America: Microfoundations of the Productivity J-curve (s).* US Census Bureau, Center for Economic Studies.

Montiel Olea, José Luis, and Carolin Pflueger. 2013. "A Robust Test for Weak Instruments." *Journal of Business & Economic Statistics*, 31(7): 358–369.

Murphy, Gail C, Mik Kersten, and Leah Findlater. 2006. "How are Java software developers using the Eclipse IDE?" *IEEE software*, 23(4): 76–83.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence." *Science*, 381(6654): 187–192.

Pearce, Hammond, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri. 2025. "Asleep at the keyboard? assessing the security of github copilot's code contributions." *Communications of the ACM*, 68(2): 96–105.

Revelio Labs. 2026. "Revelio Labs Data Dictionary :: Methodologies."

Rosenberg, Nathan, and Manuel Trajtenberg. 2004. "A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century US." *Journal of Economic History*, 64(1): 61–99.

Stack Overflow. 2018. "Stack Overflow Developer Survey 2018." https://survey.stackoverflow.co/2018/, Accessed March 23, 2026.

Stack Overflow. 2020. "Stack Overflow Developer Survey 2020." *https://survey.stackoverflow.co/2020/*, Accessed March 23, 2026.

Stack Overflow. 2023. "Stack Overflow Developer Survey 2023." *https://survey.stackoverflow.co/2023/*, Accessed March 23, 2026.

Stack Overflow. 2025. "Stack Overflow Developer Survey 2025." *https://survey.stackoverflow.co/2025/*, Accessed March 23, 2026.

Stock, James H., and Motohiro Yogo. 2002. "Testing for Weak Instruments in Linear IV Regression."

Tambe, Prasanna B. 2026. "Reskilling the Workforce for AI: Domain Expertise and Algorithmic Literacy." *Management Science*, 72(1): 515–537.

Appendix Materials – For Online Publication

A Data Construction

A.1 GitHub–Revelio Firm Matching Procedure

The GitHub data include firm name, country, and other identifiers that we link to Revelio's employment data, which is constructed from LinkedIn profiles. We use Revelio's internal firm-name matching algorithm, which strips common legal suffixes (LLC, Inc., GmbH) and junk patterns (e.g. "– Parent") from company names, then performs an exact match to cleaned firm names in Revelio's database. The algorithm excludes LinkedIn pages not associated with any employees and, when a subsidiary is matched, aggregates employment to the parent firm.

This procedure yields a high-confidence match for approximately 60% of firms. The remaining firms are unmatched, typically because their names are very common, very short, or differ between the GitHub and LinkedIn databases. Name discrepancies arise for several reasons: firms may be referenced differently across systems (e.g. "ABC Inc." vs. "ABC Inc. Science"), character encoding may differ (particularly for Japanese firms), or the GitHub account may be held by a subsidiary whose name does not appear in Revelio's cleaned database. For these firms, we develop an AI agent-assisted matching procedure that uses additional information—industry, country, and firm size—to identify the correct LinkedIn company page. We describe this procedure in the next subsection.

A.2 AI Agent-Assisted Matching Procedure

For the approximately 40% of firms not matched by Revelio's algorithm, we develop an AI agent-assisted procedure to identify the correct LinkedIn company page. The agent (Claude Sonnet 4.6) executes the following steps for each unmatched firm:

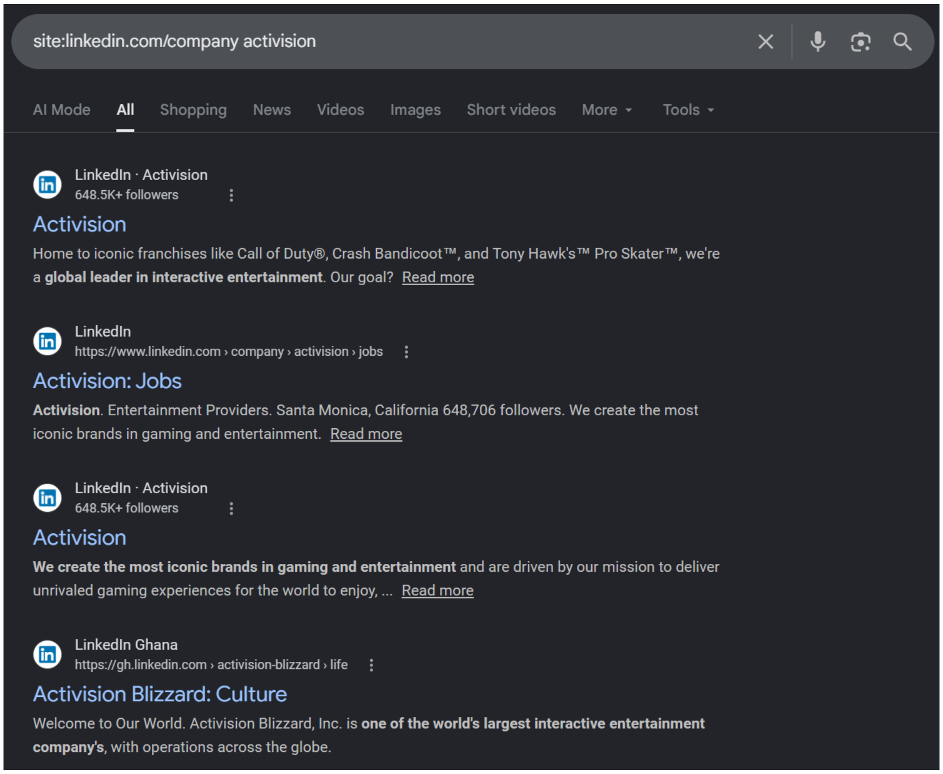
1. **Search.** Use Google's site-specific search to search LinkedIn for the firm's Salesforce firm name and country: `site:linkedin.com/company ''COMPANY_NAME'' COUNTRY`. If results are poor, search using less restrictive queries, including dropping exact search, removing company suffixes, reordering words, and using name stubs instead of full names.

2. **Filter.** Among the results, identify the company page whose name closely matches the full Salesforce account name and whose LinkedIn-reported location matches the firm's account country. Additional suggestive criteria include whether the company page is "claimed" by an owner (and therefore not automatically generated) and larger follower and employee counts.

3. **Record.** Record the page's LinkedIn slug from the URL (e.g., `linkedin.com/company/activision`; Figure A.2), and note possible issues with the match, such as company renames, parent-subsidiary relationships, and mismatched countries.

Because this procedure is a well-defined sequence of steps, we prompt the agent to execute it for each of the approximately 3,200 unmatched firms. A second agent then scores the confidence of

Figure A.1: Using site-specific search for "activision"



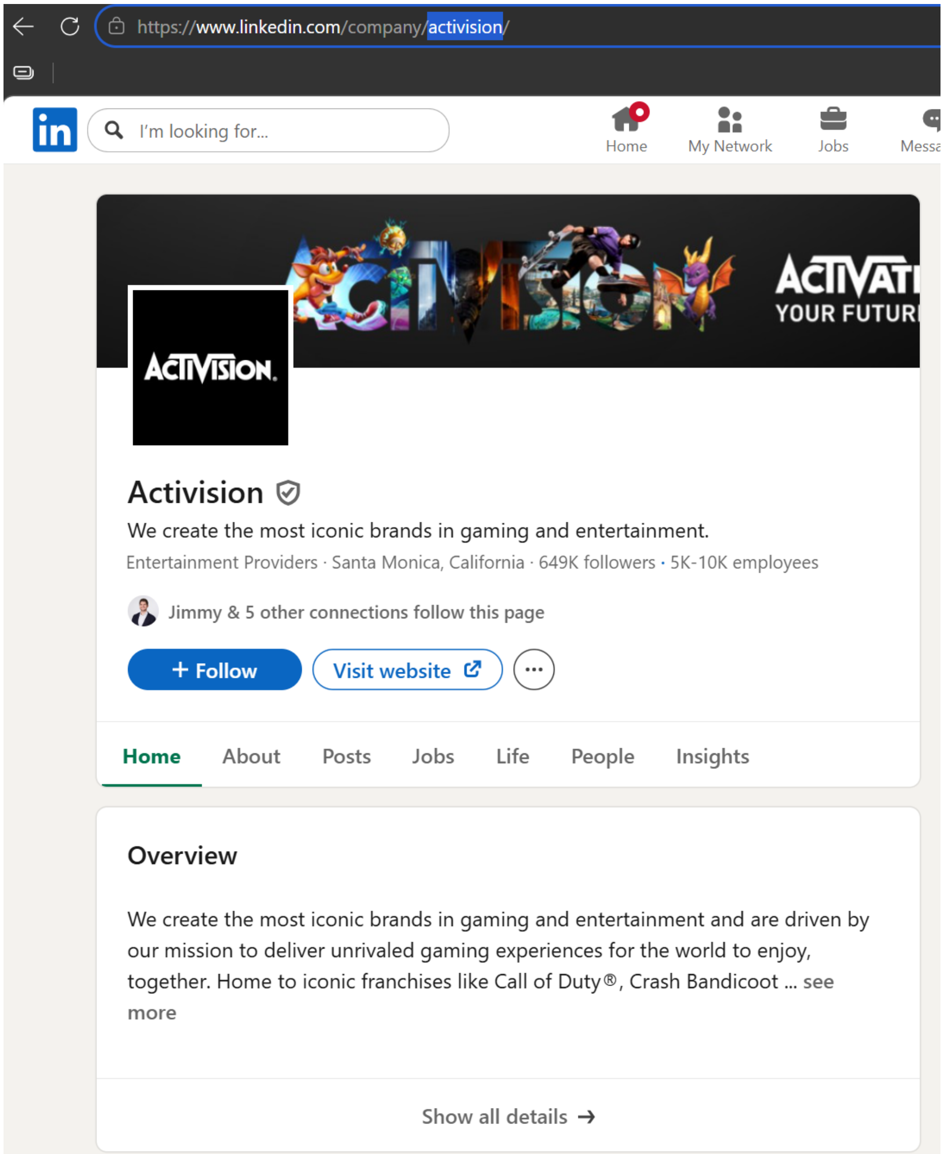
each proposed match as *high*, *medium*, *low*, or *no_match* by evaluating the similarity between the firm's Salesforce name and the LinkedIn page name, whether the country matches, and whether the LinkedIn page has a plausible number of employees. We retain only high- and medium-confidence matches for the final sample.

RCID lookup. Each matched LinkedIn slug is converted to a Revelio company identifier (RCID) by querying Revelio's `company_mapping` table on WRDS, which maps LinkedIn URLs to RCIDs. Of the 2,803 high- and medium-confidence agent matches, 2,679 (96%) yield an RCID, confirming that the matched LinkedIn pages correspond to firms in Revelio's employment database.

Duplicate resolution. In some cases, multiple GitHub firms (typically a parent and subsidiary) match to the same LinkedIn page. We identify 82 such duplicate matches. These are retained: each GitHub firm is linked to the same Revelio RCID, and employment is allocated at the parent level using GitHub's internal firm hierarchy.

URL repair. A second human reviewer checked firms in four chunks whose LinkedIn slugs were dead, redirected, or otherwise inaccessible. The reviewer provided corrected slugs where possible, recovering matches for 46 of 59 firms with broken URLs. This URL repair step corrects link rot but does not independently re-evaluate the agent's firm identification.

Figure A.2: LinkedIn company profile page showing the slug `activision` in the URL



A.3 Validation Sample and Match Accuracy

We validate both matching pipelines by drawing random samples of 100 firms from each and comparing the algorithmic or AI-proposed match to the firm a human reviewer would have selected. The human reviewer identified the correct LinkedIn company page by comparing firm name, country, and industry using the same algorithm as given to the agent.

For the Revelio algorithmic match validation, the reviewer compared the Revelio-assigned company (identified by RCID) to the firm they identified independently, recording whether the match was correct, incorrect (wrong firm), or a false positive. For the agent match validation, the reviewer compared the agent's proposed LinkedIn slug to the slug they identified independently, recording the correct slug or marking the firm as unmatchable.

Table A.1 reports the precision audit of Revelio's algorithmic matches. Because the validation sample is drawn from firms that Revelio matched, this is a precision audit rather than a full confusion matrix. Of the 100 firms reviewed, 83 were correctly matched, 16 were matched to the wrong firm (typically a subsidiary matched to its parent company, or vice versa), and 1 was a false positive that should not have been matched.

Table A.1: Revelio Algorithmic Match: Precision Audit (N = 100)

Classification	Count	Share
Correct match	83	0.83
Incorrect match	16	0.16
False positive	1	0.01
Agreement rate	0.83	

Table A.2 reports the error classification for the AI agent-assisted matches. The agent correctly identified the LinkedIn page for 85 of 100 firms, missed a match that the human found for 8 firms, correctly classified 4 firms as unmatchable, and proposed the wrong firm for 3. These counts imply a precision of 0.97 and a recall of 0.91, for an F1 score of 0.94. Of the 85 correct agent matches, 80 (94%) yielded a Revelio company identifier (RCID), confirming that the matched LinkedIn pages correspond to firms in Revelio's employment database. Both validation exercises used a single rater; no inter-rater reliability assessment was conducted.

Table A.2: AI-Assisted Match: Error Classification (N = 100)

Classification	Count	Share
Correct match (TP)	85	0.85
False negative (FN)	8	0.08
True negative (TN)	4	0.04
Incorrect match (IM)	3	0.03
False positive (FP)	0	0.00
Precision	0.97	
Recall	0.91	
F1	0.94	

B Additional Mechanism and Heterogeneity Results

Table A.3 reports effects on SWE outflows by seniority level. Neither junior nor senior SWE separations change significantly, confirming that the geographic recomposition in the main text operates through the reallocation of new hires, not through increased separations.

Table A.3: Effect of Copilot Adoption on Quarterly SWE Outflows by Seniority

	Junior SWE Outflow (1)	Senior SWE Outflow (2)
Reduced Form		
$Z_i \times$ Post	-0.08	0.33
	(0.68)	(0.21)
DV Mean	11.9	1.1
Firm FE	Yes	Yes
Quarter FE	Yes	Yes

Notes: Each column reports effects on quarterly SWE outflow counts. Data collapsed to firm-quarter. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.4 reports effects on SWE outflows and net flows by geography.

Table A.4: Effect of Copilot Adoption on Quarterly SWE Outflows and Net Flows by Geography

	SWE Outflow		SWE Net Flow	
	Domestic (1)	Intl (2)	Domestic (3)	Intl (4)
Reduced Form				
$Z_i \times$ Post	-0.06	2.19**	-1.81	-2.47**
	(0.69)	(1.05)	(1.19)	(1.21)
DV Mean	11.3	8.1	2.5	2.5
Firm FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes

Notes: Columns (1)–(2) report effects on quarterly domestic and international SWE outflows. Columns (3)–(4) report net flows (inflows minus outflows). Data collapsed to firm-quarter. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5 decomposes the junior and senior SWE inflow results by industry and HQ location. Unlike the geographic effects, the junior hiring decline is broad-based—all four cells are negative, and there is no significant heterogeneity by industry or HQ.

Table A.5: Heterogeneity: Junior and Senior SWE Inflows by Industry and HQ Location

	US-HQ	Non-US-HQ
Junior SWE Inflow (Quarterly Count)		
IT Services	-3.91**	-0.63
	(1.62)	(2.20)
Non-IT	-4.29	-2.55
	(2.66)	(3.09)
Senior SWE Inflow (Quarterly Count)		
IT Services	-0.47**	-0.04
	(0.21)	(0.16)
Non-IT	-0.21	0.29
	(0.43)	(0.42)
Firm FE	Yes	Yes
Quarter FE	Yes	Yes

Notes: Each cell reports the reduced-form coefficient on $Z_i \times$ Post interacted with industry and HQ location indicators. Dependent variable: quarterly SWE inflow count. Data collapsed to firm-quarter. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.6 reports the firm size heterogeneity in the domestic share result. Large and small firms reduce their domestic share at identical rates (F-test $p = 0.90$).

Table A.6: Heterogeneity: Domestic Share of Employment by Firm Size

	Large (Q3–Q4)	Small (Q1–Q2)
$Z_i \times$ Post	-0.0037***	-0.0039**
	(0.0012)	(0.0019)
Firm FE	Yes	Yes
Month FE	Yes	Yes

Notes: Each column reports the reduced-form coefficient on $Z_i \times$ Post interacted with firm size indicators. Large = pre-period size quartiles 3–4; Small = quartiles 1–2. Dependent variable: domestic share of total employment. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C Robustness

C.1 Instrument Window Robustness

Table A.7 reports the domestic share results for $W = 1, 2, 3$. All outcomes are significant at $W = 1$ and $W = 2$; the effect is monotonically declining in magnitude as the window widens, consistent with wider windows including more marginally affected firms.

Window Robustness: Domestic Share of Employment

	$W = 1$	$W = 2$	$W = 3$
Domestic Share	-0.0076***	-0.0038***	-0.0031***
	(0.0018)	(0.0011)	(0.0010)
Share SWE Domestic	-0.0108***	-0.0062**	-0.0046**
	(0.0035)	(0.0025)	(0.0022)
Share Tech Domestic	-0.0087***	-0.0037**	-0.0026
	(0.0028)	(0.0019)	(0.0016)
Share Analyst Domestic	5.3e+13	6.1e+13	7.1e+13
	(5.3e+13)	(6.1e+13)	(7.1e+13)
Firm FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes

Notes: Each cell reports the reduced-form coefficient on $Z_i \times$ Post for the indicated instrument window W. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.8 reports the IT × HQ heterogeneity for each window. The IT × US cell is stable across all windows (-5.2 to -4.7 pp).

Table A.8: Window Robustness: IT × HQ Heterogeneity on Domestic SWE Share

	$W = 1$	$W = 2$	$W = 3$
DV: Domestic SWE Share			
IT × US	-0.0516***	-0.0468***	-0.0466***
	(0.0067)	(0.0056)	(0.0050)
IT × Non-US	-0.0062	-0.0018	0.0018
	(0.0106)	(0.0059)	(0.0050)
Non-IT × US	-0.0098*	-0.0131***	-0.0120***
	(0.0059)	(0.0041)	(0.0033)
Non-IT × Non-US	0.0152***	0.0171***	0.0184***
	(0.0051)	(0.0033)	(0.0028)
Firm FE	Yes	Yes	Yes
Month FE	Yes	Yes	Yes

Notes: Each cell reports the reduced-form coefficient on $Z_i \times$ Post interacted with industry and HQ indicators, for the indicated instrument window W. Dependent variable: domestic share of SWE employment. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.9 reports the quarterly inflow results across windows. Both the domestic share of SWE inflows and junior SWE inflow counts are significant at all three windows, with effects strengthening at $W = 1$.

Table A.9: Window Robustness: Quarterly Inflow Outcomes

	$W = 1$	$W = 2$	$W = 3$
Dom Share SWE Inflow	-0.0167***	-0.0102***	-0.0071**
	(0.0054)	(0.0036)	(0.0031)
Junior SWE Inflow	-7.30***	-3.26**	-2.49**
	(2.57)	(1.61)	(1.16)
Senior SWE Inflow	-0.14	-0.05	-0.01
	(0.27)	(0.23)	(0.16)
Firm FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes

Notes: Each cell reports the reduced-form coefficient on $Z_i \times$ Post for the indicated instrument window W. Data collapsed to firm-quarter. Standard errors clustered by firm. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C.2 Pre-Trend Tests

Table A.10 reports joint pre-trend F-tests for the monthly domestic share outcomes. All four outcomes pass at the 5% level, with p-values ranging from 0.060 (SWE) to 0.213 (analyst).

Table A.10: Pre-Trend Tests: Monthly Domestic Share Outcomes

Outcome	Pre-trend F	p-value
Domestic Share (All)	1.74	0.095
Domestic Share (SWE)	1.94	0.060
Domestic Share (Tech)	1.84	0.075
Domestic Share (Analyst)	1.37	0.213

Notes: Joint F-test on all pre-period (before 2023Q1) quarterly interaction coefficients from the event study specification $Y_{it} = \alpha_i + \delta_t + \sum_q \gamma_q Z_i \times \mathbf{1}(q) + \varepsilon_{it}$, with $W = 2$.

C.3 First Stage by Subgroup

Table A.11 reports the first stage coefficient and F-statistic for each heterogeneity subgroup. All subgroups have $F > 14$ except IT $\times$ Non-US ($F = 0.2$, $N = 519$), where the renewal instrument does not predict Copilot adoption. The null coefficient for this cell in the heterogeneity analysis (Table 9) is therefore uninformative—it reflects a weak instrument, not a true null effect.

Table A.11: First Stage Strength by Subgroup

Subgroup	N Firms	FS Coef	SE	F-stat
Full Sample	5,981	0.0892	(0.0107)	69.3
IT Services	1,466	0.0766	(0.0200)	14.7
Non-IT	4,515	0.0925	(0.0125)	55.1
US-HQ	3,185	0.1069	(0.0145)	54.5
Non-US-HQ	2,747	0.0645	(0.0159)	16.5
IT $\times$ US	942	0.0971	(0.0235)	17.0
IT $\times$ Non-US	519	0.0165	(0.0382)	0.2
Non-IT $\times$ US	2,243	0.1037	(0.0180)	33.3
Non-IT $\times$ Non-US	2,228	0.0802	(0.0174)	21.3

Notes: Each row reports the first stage regression of D_{it} (treat_baseline) on $Z_i \times \text{Post}_t$ within the indicated subgroup. F-statistic is the squared t-statistic on the excluded instrument. The instrument is $Z_i \times$ Post with $W = 2$. Standard errors clustered by firm.

C.4 Complier Characteristics: Window Comparison

Table A.12 compares complier characteristics across instrument windows. The complier share is stable (0.37–0.39), and complier characteristics vary only modestly. $W = 1$ compliers are slightly smaller and less SWE-intensive, consistent with the tightest window selecting firms with shorter renewal cycles.

Table A.12: Complier Characteristics by Instrument Window

	$W = 1$	$W = 2$	$W = 3$
Ln(Headcount)	7.27	7.32	7.37
SWE Share	0.0500	0.0673	0.0663
Tech Share	0.1189	0.1383	0.1365
Firm Age	42.22	41.32	40.85
US HQ	0.5102	0.5720	0.5882
IT Industry	0.2040	0.2788	0.2726
Cycle Length	9.79	8.69	8.48
Complier share	0.3897	0.3672	0.3797

Notes: Complier means computed using κ-weights from Abadie (2003) with the panel first stage (D_{it} on $Z_i \times \text{Post}_t$) for each instrument window W. Complier share = $E[\kappa_i]$.

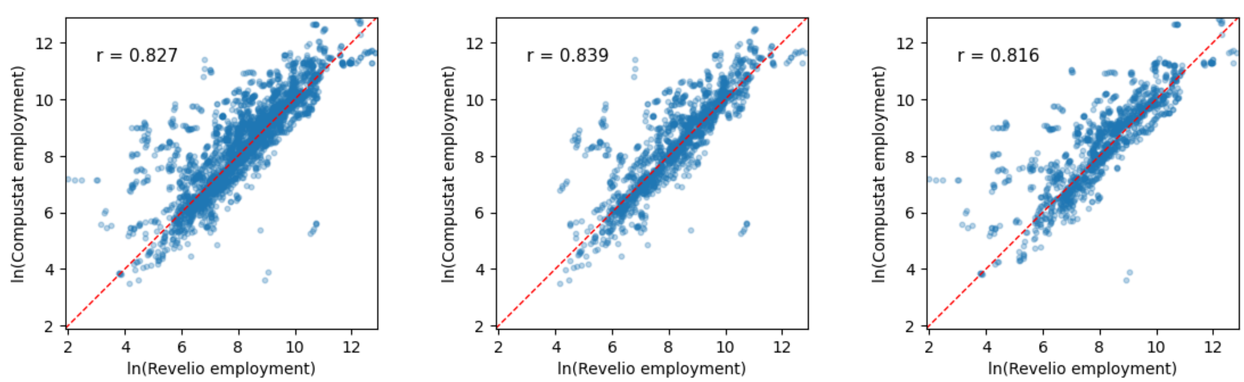
D Revelio Data Validation

D.1 Correlations with Compustat data

Compustat collates financial and other business information on companies across the world. A product of S&P Global, a financial services firm that specializes in market analytics, it focuses on publicly traded companies and uses investor reports and filings with government agencies for its data.

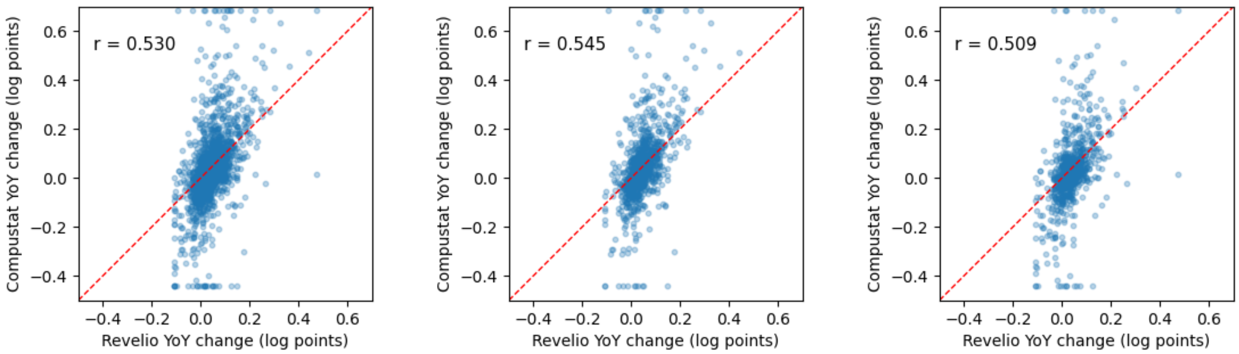
Figure A.3 displays the tight correlations between employment levels reported in Compustat and Revelio. Figure A.4 shows the positive correlation between year-over-year employment flows in Compustat and Revelio.

Figure A.3: Correlations with reported employment



(a) Full sample (b) Adopters (c) Non-adopters

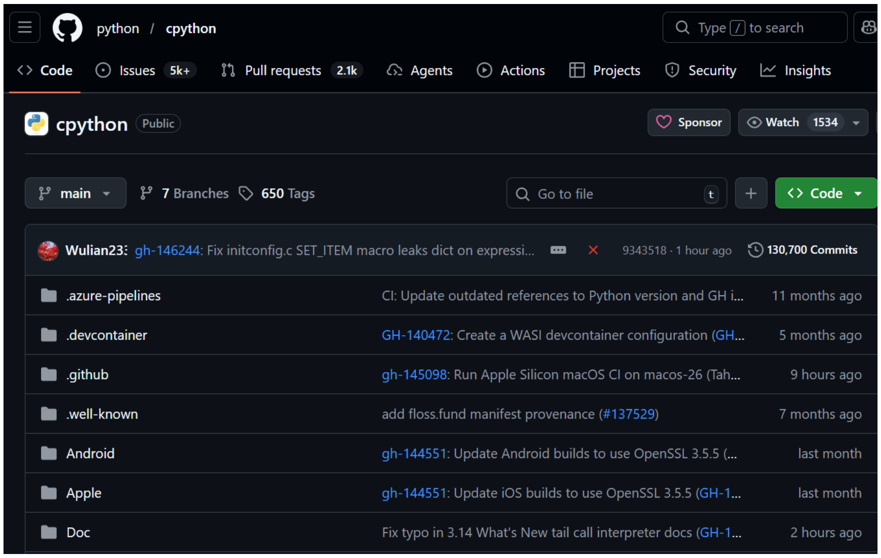
Figure A.4: Correlations with reported yearly employment flows



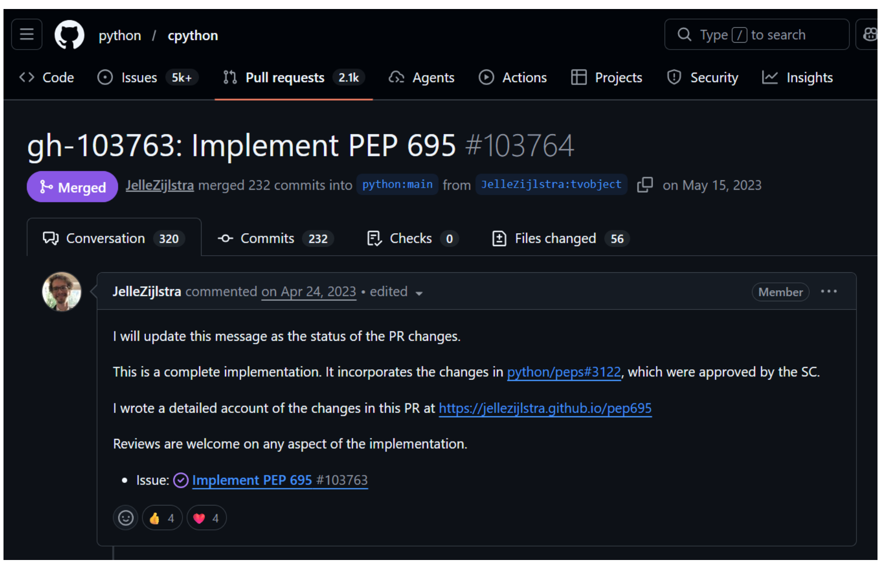
(a) Full sample (b) Adopters (c) Non-adopters

E GitHub Copilot Enterprise

Figure A.5: GitHub



(a) Repository for the standard implementation of the Python language



(b) "Pull request" seeking to update the Python repository

GitHub is a platform that allows users to share code bases, or "repositories," with each other and collaborate on software development. Users can download repositories to their local computers, make changes, and upload them back to GitHub. GitHub tracks all changes made to files unless otherwise directed, meaning every version of a tracked file is available on the platform.

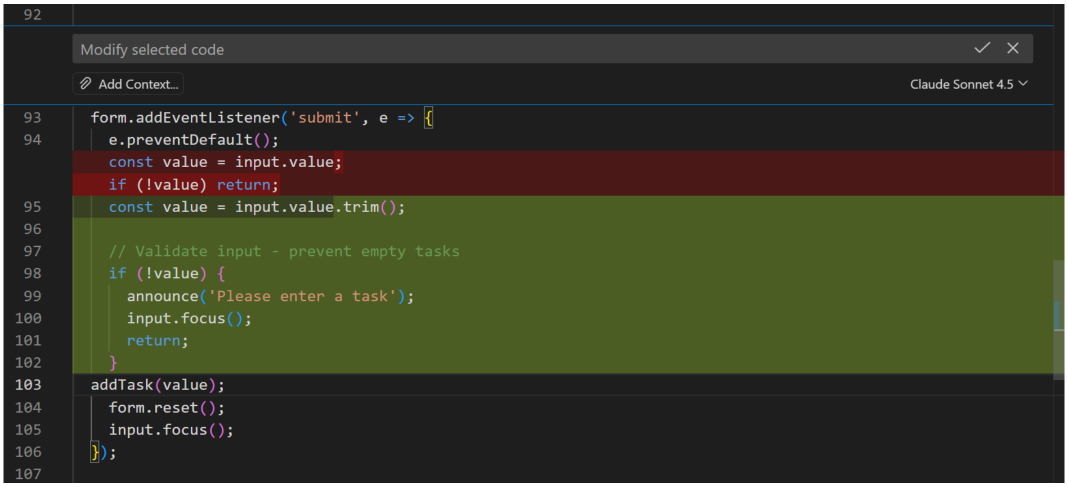
Importantly, many users can make changes to a repository in parallel. In an organizational context, developers attempt to avoid creating conflicting file changes by making "branches" of the main code base. These branches must be "merged" back to the main code base for changes to be reflected in the final product; merging is the key point at which developers review code to ensure functionality, correctness, and functionality. GitHub's "pull request" feature implements the act of merging changes from one branch into another and allows users to describe and discuss proposed changes.

GitHub also includes features for project management and code deployment. GitHub Enterprise bundles these features, allows heavier usage, and gives organizational administrators control over repository access and security.

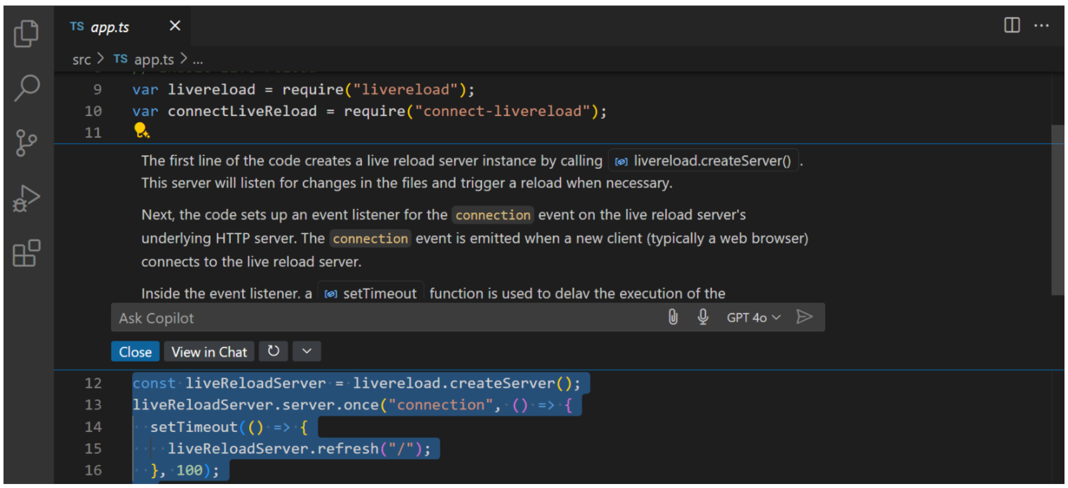
GitHub Copilot is a programming tool that uses large language models to process, interpret, and write code. At its waitlist-only, free launch in June 2021, its primary feature was code completion, or the ability to suggest blocks of code from surrounding file context and natural language comments. It was made available to general audiences in June 2022 as a paid product for individual developers. In December 2022, organizations became able to purchase GitHub Copilot for Business, which included commitments not to use customer code for model training.

Chat functionality, which allowed users to converse with GitHub Copilot about a code base, and Copilot for Pull Requests, which automated the process of opening pull requests and describing changes, entered private preview in March 2023. In November 2023, GitHub announced that Copilot Chat would become generally available. In February 2024, the new GitHub Copilot Enterprise combined GitHub Copilot Business with Copilot for Pull Requests, as well as the ability to get chat recommendations personalized to an organization's code base. Releases of agentic capabilities — where GitHub Copilot takes on more complex tasks and edits multiple files at once — began in April 2024 with Copilot Workspace and expanded in February 2025 with private previews of "agent mode," though public preview did not open until almost the end of our sample period in May 2025.

Figure A.6: GitHub Copilot



(a) Code generation



(b) Chat