

WORKING PAPER

Early Evidence on the Impact of GitHub Copilot on Labor Market Outcomes for Software Engineers

Matthew Baird

LinkedIn Economic Graph Research Institute

Mar Carpanelli

LinkedIn Economic Graph Research Institute

Brian Xu

LinkedIn Economic Graph Research Institute

Kevin Xu

GitHub

September 2024

EG WP No. #5

Any opinions expressed in this paper are those of the author(s) and not those of LinkedIn or GitHub. This working paper represents work in progress distributed to encourage discussion and feedback. Citation should account for the provisional character.

Early Evidence on the Impact of GitHub Copilot on Labor Market Outcomes for Software Engineers

Matthew Baird¹, Mar Carpanelli², Brian Xu³, and Kevin Xu⁴

September 2024

ABSTRACT⁵

This paper examines the impact of GitHub Copilot (GHC), a generative AI (GAI)-powered coding assistant, on labor market outcomes for software engineers (SWE). Using data from LinkedIn and GitHub corporate licenses, we analyze how GHC adoption affects skills and labor demand, supply, and hiring. We find evidence that, contrary to some fears regarding AI, companies adopting this augmentation tool hire more SWEs. Specifically, GHC adoption leads to a 3.2 percentage point (pp) higher probability of hiring new SWEs each month, led primarily by more entry-level individual contributor (IC) SWE hires (6.6 pp higher likelihood, with 3.2% higher count hired monthly), as well as 4.9% higher probability of hiring at least one senior IC each month. GHC leads to 13.3% more non-programming skills among new SWE hires. High-concentration GHC firms also see an increase in the non-programming skills of existing SWEs without slowing the rate of new programming skills. Additionally, these firms increase their job postings for SWEs, including some evidence of increase in postings for SWEs without degrees.

Keywords: labor, human capital, artificial intelligence, productivity

JEL classifications: J24, J20, O33

¹ Corresponding author. Senior Staff Economist, LinkedIn Corporation. mdbaird@linkedin.com

² Staff Data Scientist, LinkedIn Corporation. mcapanelli@linkedin.com

³ Staff Data Scientist, LinkedIn Corporation. brxu@linkedin.com

⁴ Staff Software Engineer, GitHub Inc. khxu@github.com

⁵ Acknowledgements: we would like to thank several people for their useful feedback on this paper, including Jon Adams, Gorki De Los Santos, Chris Grant, Kory Kantenga, Karin Kimbrough, Bob Liu, Sida Peng, Sharat Raghavan, Michael Schwarz, and participants at the Artificial Intelligence, Cognition, and Economy Spring Workshop, LinkedIn Economic Seminar series, and AI at Wharton 2nd Annual Business and Generative AI Workshop.

1. Introduction

In recent years, artificial intelligence (AI) has advanced rapidly, especially in the realm of generative AI (GAI). The release of OpenAI's ChatGPT in November 2022 marked a pivotal moment, showcasing the potential of AI to respond to queries posed in natural language, and to produce human-like outputs, performing tasks traditionally done by humans, such as writing, content creation, and data analysis. As GAI capabilities increasingly intersect with human expertise, there is a growing need to examine the implications of these new technologies on the workforce.

The impact of emerging technologies on the workforce has been extensively studied in academic literature. However, the different nature of this specific technology – GAI – creates new questions on the scope and mechanisms behind technology's potential impact on the workforce. GAI is a transformational technology with virtually unprecedented speed of development and potentially adoption. There is new research estimating that roughly one in two jobs may see half of their tasks impacted by GAI (Eloundou et al., 2024) and that 84% of professionals stand to see at least one quarter of the way they do tasks impacted (Kimbrough & Carpanelli, 2023). Moreover, there is recent evidence that GAI can boost worker productivity (Brynjolfsson et al., 2023a; Dell'Acqua et al., 2023), ideation and innovation (Doshi & Hauser, 2023; Girotra et al., 2023), and, to some extent, problem solving (Boussioux et al., 2023; Otis et al., 2023).

However, little has been done to evaluate the effects of GAI on labor market outcomes for impacted workers. These workers, along with business leaders, policymakers, and educators, benefit from understanding the potential impacts of GAI on the workforce, such as may arise from increases in worker productivity via augmentation of tasks complemented by GAI, as well as the risks of its adoption, including the possibilities of job displacement as tasks are replicated by this technology.

This paper bridges this knowledge gap by investigating the impact of GAI on labor market outcomes. We do so by examining one of the earliest GAI-powered technologies broadly available to the software development community: GitHub Copilot (GHC). GHC is an AI coding assistant which embeds directly into coding Integrated Development

Environments (IDEs).⁶ Within a developer’s IDE – the software in which they write and manage their code – GHC provides auto-code suggestions and auto-completions (at times several lines long) as well as a sophisticated query-based chat functionality. The chat function of GHC can answer questions, such as the inputs and syntax of functions, how a given task would be coded, and the meaning of and solution to different bugs. GHC was launched in June 2021 as a plugin for Visual Studio Code, and in June 2022, it moved out of the technical preview period. It has seen relatively quick adoption across many companies starting in 2023.

We explore various labor market outcomes for software engineers (SWEs), including the level and skills required by job postings and job seeker behavior related to job titles searched – by combining LinkedIn’s Economic Graph data with GHC corporate licenses. By examining proprietary data both from LinkedIn and GitHub, we are able to explore the impact of GHC, one implementation of a major technological advancement, on one segment of the economy – software engineers – during the process of adoption. Such understanding is essential for workers, business leaders, policymakers, businesses, and educators to navigate potential and actual changes and prepare the workforce for a future where AI plays a more prevalent role.

1.1 Related literature

While easily accessible GAI technologies are relatively new, the economic literature has been focusing for decades on the impact of various technological advancements on the workforce.

Historically, technological advancements have primarily automated lower-skilled jobs that involve routine and repetitive tasks (Autor & Dorn, 2013; Goldin & Katz, 1998). These jobs have traditionally been most easily codified and mechanized, leading to some job displacement in specific sectors, such as manufacturing and agriculture (Acemoglu & Restrepo, 2019; Katz & Murphy, 1992). While low-skill service jobs have grown due to the types of skills and tasks involved, the overall effect of technological change has been

⁶ <https://docs.github.com/en/copilot/about-github-copilot/what-is-github-copilot>

an increased demand for skilled labor and a rise in wage inequality (Autor & Dorn, 2013).

A relatively new literature on the impact of “primal” forms of AI on the labor market has consistently found that AI is predominantly capable of automating routine tasks (Brynjolfsson et al., 2018; Felten et al., 2018; Furman & Seamans, 2019; Webb, 2019). These tasks are generally rule-based and predictable, making them suitable for machine learning algorithms. In contrast, non-routine tasks (e.g., those that require people skills, strategy, and problem-solving), which are integral to many jobs, involve complexity, creativity, and human judgment that primal AI was yet to master effectively at the time of the prior studies. Consequently, this initial wave of research agreed that the displacement effects of early AI technologies on high-skilled employment remained somewhat limited—as these roles typically require sophisticated cognitive and interpersonal skills. Relatedly, the potential for the primal AI technology to reduce wage inequality was found to be uncertain. As one example, assuming the historical pattern of long-run substitution observed in other technologies, Webb estimated that AI may reduce wage inequality, but may not affect the top 1% (Webb, 2019).

The release of OpenAI’s ChatGPT in November 2022 marked a pivotal moment for AI and, particularly, GAI technologies. GAI tools such as GHC, ChatGPT, Microsoft Copilot, and Gemini introduced a new type of automation: these technologies are capable of reproducing many cognitively routine tasks as well as many non-routine tasks, such as writing, content creation, and data analysis. These tools have even outperformed humans in standardized academic and professional tests in the fields of economics, medicine, and others (Geerling et al., 2023; Kung et al., 2023).

As GAI stands to be a technology with the potential to automate tasks that have been traditionally exclusively performed by humans, there is growing interest in understanding the scope of the potential impact, and the mechanisms behind it. The most recent analyses estimate that roughly half of jobs and workers could be significantly impacted by GAI. For example, a recent analysis using a task-based framework suggests that 46% of jobs could have over half of their tasks affected by GAI (Eloundou et al., 2024). Similarly, using a skills-based framework, LinkedIn researchers find that 84% of professionals in the US are in occupations that could leverage GAI to automate at least

25% of their (primarily routine) tasks (Kimbrough & Carpanelli, 2023).

While research on the actual impact of GAI on the workforce is still new, there is some evidence that GAI can enhance productivity across multiple fields. For example, a randomized experiment on customer support agents found that access to an AI-based conversational assistant increases productivity, as measured by issues resolved per hour, by 14% on average, and by 34% for novice and low-skilled workers (Brynjolfsson et al., 2023a). Moreover, a joint study between academic researchers and the management consulting firm Boston Consulting Group found that consultants using an AI tool were significantly more productive than those not using it: they completed 12% more tasks on average, completed tasks 25% faster, and their results were 40% higher quality as compared to the control group (Dell'Acqua et al., 2023).

When it comes to ideation and innovation, core human aptitudes, researchers have found some early evidence that GAI technologies can generate ideas much faster and cheaper than humans and that the GAI-generated ideas are on average higher quality (Girotra et al., 2023). However, in fields like creative writing, GAI-generated stories tended to converge to each other, suggesting limitations for the impact on overall creativity (Doshi & Hauser, 2023).

Similarly, in problem solving, another core human aptitude, results are mixed. A study on a crowdsourcing challenge focused on circular economy business ideas found that while solutions generated through human-AI collaboration matched the creativity of those from the human solvers, the human-AI solutions provided more value (Boussioux et al., 2023). On the other hand, an experiment providing AI-generated advice to Kenyan entrepreneurs' business problems found that the causal effect of GAI access varied with the baseline business performance of the entrepreneur: high performers improved their performance by roughly 20% when leveraging AI advice, whereas low performers did roughly 10% worse (Otis et al., 2023). These early studies suggest that new GAI technologies can be promising tools to augment human capabilities across several fields.

2. Context and Conceptual Framework

One powerful capability of GAIs is the ability to write and debug code, which are core skills for software engineers, and, consequently, has fueled discussions around the potential impact of GAI on SWE roles (Senz, 2023). According to Eloundou et al. (2024), SWEs (Computer Software Engineers / Architects under O*NET) face high exposure to GAI, as these tools overlap with multiple core tasks, and could decrease the time required to complete most of such tasks by at least 50% (Eloundou et al., 2024). Similarly, Kimbrough & Carpanelli find that SWEs' skills and tasks stand to be assisted by GAI, since more than half of their most representative skills could be replicated by GAI tools, while the role would still require a good share of complementary or "exclusively-human" skills (Kimbrough & Carpanelli, 2023).

GHC is one of said GAI tools. It is a code completion tool developed by GitHub in collaboration with OpenAI, and it uses machine learning algorithms to provide real-time code suggestions and autocompletions directly within the user's IDE, as well as chat functionality to respond to questions (GitHub, 2024). By analyzing the context of the code being written, GHC can suggest entire lines or blocks of code, making coding more relevant and efficient. The query-based features allow SWEs to ask how to write certain parts of code, inquire about functions and syntax, debug based on error codes, and more. Since its launch in 2022, it has been widely adopted by software developers and programmers across various industries, from solo developers to large enterprise teams (Zhao, 2023). Our data, as shown below, show evidence in favor of widening adoption of GHC.

The implications of GHC on the software development community are potentially significant: it can greatly enhance productivity by reducing the amount of repetitive coding work and by allowing developers to focus on more complex and creative aspects of their projects. A survey ran by GitHub researchers found that 88% of the developers surveyed have a higher perceived productivity when using GHC, with 96% reporting that this tool makes them work faster on repetitive tasks and 74% saying it enables them to focus on more satisfying work (Kalliamvakou, 2022). More recently, a controlled experiment where software developers were asked to implement an HTTP server in

JavaScript as quickly as possible found that the treatment group, with access to GHC serving as a pair programmer, completed the task 56% faster than the control group (Peng et al., 2023).

Survey and experimental research results on the impact of GHC on software developers are very much aligned with general results in the broader AI and GAI literature. However, there is limited evidence on the impact of exposure to this technology on the labor market. It is unclear whether these productivity gains translate into changes in labor market outcomes for these workers, such as skilling, career growth, or hiring.

In this paper, we explore six groupings of outcomes: skills listed in job postings for SWEs; skills of existing SWEs at GHC customers; skills of new hires at GHC firms; labor supply decisions of SWEs at GHC firms; labor demand for SWEs from GHC firms; and hiring outcomes at GHC firms. These outcomes are enumerated in the data section.

With regards to skills of existing SWEs at GHC firms, we hypothesize that exposure to GAI technology will allow them to more rapidly develop programming skills. GAI acts as a tutor and increases the confidence of SWEs with respect to programming languages. Meanwhile, this same complementarity of programming skills may lead hirers at GHC firms to hire SWEs who have more nonprogramming skills under the expectation that their productivity with regards to programming skills will be assisted by GHC (e.g., code completion, help with errors) and thus increase the comparative advantage of non-programming skills.

The hypotheses regarding hiring are less clear. As prior research has pointed out, increased productivity for a group of workers may increase demand for these workers given their increased efficiency and capabilities—their marginal product of labor has increased. On the other hand, the number of SWEs desired by a firm may decrease as a result of the increased productivity, as fewer SWEs can do the same amount of work as more SWEs prior to productivity gains (Acemoglu & Autor, 2011; Acemoglu & Restrepo, 2019; Webb, 2019).

3. Data

This research utilizes two key data sources to connect GHC adoption to labor market outcomes. The first source is company-level data from GitHub identifying timing of both corporate GitHub and GHC accounts as well as the number of licenses purchased. GitHub is the largest repository of code globally and provides users (including teams within firms) the version-control Git applications and hosting services. The data we leverage, containing the timing of a firm using GitHub as well as GHC provides crucial information on the treatment status and sample for evaluating the labor market outcomes from GHC adoption.

The second data source we examine comes from LinkedIn's Economic Graph. LinkedIn is the largest professional network in the world, as well as a leading large online job search market. LinkedIn also has particularly strong market penetration in the information and technology sector as well as the professional, scientific, and technical activities sector, all of which are key markets for SWEs (Zhu et al., 2020). We leverage key data from both job postings as well as members' profiles over time to understand the evolution of labor market outcomes with respect to skills, job posts, and hiring. We aggregate the data to the firm-by-month level to correspond with the GitHub data.

The goal of this paper is to provide an exploratory look into how GHC may be impacting the labor market for SWEs. Given the exploratory nature, we investigate a broad set of outcomes.⁷

- Labor demand skills: average number of programming and nonprogramming skills listed in job posting.
- Labor supply skills (skills of existing SWEs)⁸: average number of skills (overall and by skill group), share of skills in a given skill group.
- New hires' skills: average number of programming and nonprogramming skills held by new hires.

⁷ For count outcomes, such as count of new SWE hires, we estimate the model for three outcomes: the actual number, the Winsorized number (at 99th percentile), and the extensive margin ($n=0$ vs $n>0$).

⁸ Skill groupings examined: programming vs. nonprogramming; soft vs. business vs. disruptive tech. vs other tech. vs. specialized industry skills. We test overall and limiting the sample of SWEs at the company to those who had at least one skill listed on platform before GHC rollout.

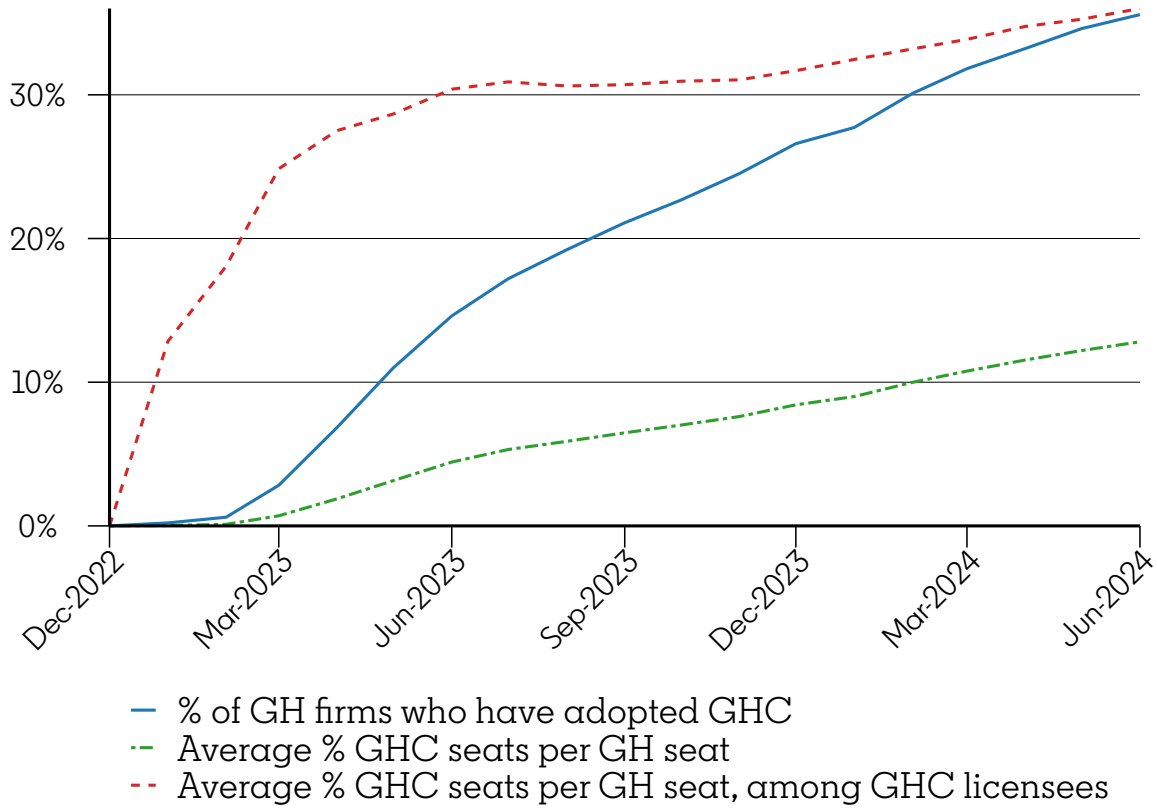
- Labor supply behavior of SWEs: among existing SWEs at company, number and share of job views and applications for other job postings, both overall and for positions more senior than their current position.
- Labor demand for SWEs: number and share of job postings for SWEs; number and share of job postings for SWEs with no degree requirements (US only for the last outcome).
- SWE hiring outcomes⁹: Number and share of new hires, overall and by seniority level.

To join the two data sets, we do exact and fuzzy matching on company name with manual inspection. The Appendix describes the methodology for this matching procedure and the related statistics (Tables A.1-A.3, Figure A.1). In our primary specification, we retain all firms which we can match between the two data sets. In sensitivity analysis, we limit the sample to the strongest matches as defined in the appendix methodology.

Figure 1 presents three trends in GHC adoption over time in our matched sample. Adoption started in January 2023. There was rapid adoption thereafter. This is true on the extensive margin (the percent of GH firms who have at least one GHC license, increasing from 0.2% and 0.6% of GH firms in January and February 2022 respectively, to 35.6% in June 2024). It is also true on the intensive margin (the average percent of GHC licenses per GH license, among GHC licensees), which increased from 12.8% among the first adopters in January 2022 to 36.0% in June 2024. The net effect then, driven by both the extensive and intensive margin of adoption (that is, inclusive of the non-GHC-adopting GitHub firms), reveals that for the average GH licensee, there are 12.8 GHC licenses per GH license. Each of these growth rates shows no immediate sign of slowdown either.

⁹ Seniority levels: intern, entry, senior non-management, and executive

Figure 1: GHC adoption over time



Note: Sample is limited to GitHub accounts that are matched within LinkedIn data

Table 1 provides descriptive statistics for treatment and covariates. It also shows comparisons between GHC and non-GHC firms and months. Across months, we find that 4% of GH firms have GHC. This is lower than suggested by Figure 1 because it averages from the start of our data (February 2019), relevant pre-treatment data for the difference-in-difference model to incorporate. Table 1 also shows the importance of the covariates, which are used to help with the matching. Compared to firms/months without GHC, GHC firm/months have on average more SWEs employed, more GH licenses, more hires, more job postings, more skills per job post and per new hire. Each of these is statistically different between firm/months with and without GHC. Appendix Table A.4 further provides the same measures for each of the outcomes.

Table 1: Descriptive statistics on treatment assignment and covariates

	All firms	Firm/months without GHC	Firm/months with GHC
Treatment status			
Any GHC licenses	0.040 (0.196)	0 N/A	1 N/A
GHC / GH (GHC penetration)	0.013 (0.084)	0 N/A	0.327 (0.274)
Covariates			
Log number SWE	2.899 (1.327)	2.877 (1.318)	3.427 (1.419)
Log number GH licenses	1.685 (1.890)	1.586 (1.838)	4.047 (1.559)
Log overall hires	30.212 (224.056)	29.740 (221.114)	41.511 (285.354)
Number of job postings	18.626 (413.838)	17.306 (372.182)	50.212 (976.816)
Number of SWE job postings	2.145 (62.983)	2.057 (63.465)	4.244 (50.034)
Avg. number of programming skills for job post	0.362 (1.485)	0.353 (1.472)	0.574 (1.741)
Avg. number of nonprogramming skills for job post	0.226 (1.006)	0.216 (0.984)	0.444 (1.422)
Average number of programming skills per new hire	2.982 (5.566)	2.948 (5.538)	3.792 (6.137)
Average number of nonprogramming skills per new hire	1.132 (2.506)	1.116 (2.492)	1.501 (2.794)

Note: mean and standard deviation (in parentheses) reported. log number SWE and log number GH licenses are always included as covariates. The remaining covariates listed may or may not be included, depending on the model. Appendix Table A.4 lists each outcome explicitly.

4. Empirical Strategy

We evaluate the impact of GHC on labor market outcomes using Callaway & Sant'Anna (2021)'s Difference-in-Difference (CSDID) for scenarios with multiple time periods and staggered adoption. Intuitively, this statistical technique estimates the causal effect of corporate adoption of GHC on labor market outcomes by comparing changes in

outcomes over time between companies that adopt GHC and observationally similar companies that have not. The method uses the trajectory changes in the outcomes of the control group to proxy for the magnitude of the change in the outcome for the treated group (i.e. firms that adopt GHC). We use the R package “did”.

The treatment group is defined by those who adopt GHC. Treated periods include both the months they had GHC as well as any months after adoption, even if they were not using it anymore. This is one of the modeling assumptions of CSDID, and is appropriate in our context for two reasons. First, we find that it is relatively uncommon for companies to stop using GHC. If a company has GHC in a given month, 95.6% of the time they still have GHC in the following month. Put another way, 86.5% of firms that have GHC at some point in our sample still have it in the final month of our data. Figure A.2 shows the distribution of months a GHC firm had GHC in the year after first adoption: only a very small fraction of firms drop GHC after a few months, and most have GHC for all months. Second, we would expect that SWEs who have had exposure to GHC, even if turned off, will have benefitted from the upskilling component that may arise from treatment (i.e., enduring impacts of learning additional coding approaches), and with more familiarity with using GAI tools generally, may be more likely to turn to substitutes. This was found for example by Brynjolfsson et al. (2023) in a different context (persistent treatment effects when there were outages of GAI tools for customer support agents).

We chose as the control group GH company/ months who in the given month have not yet been GHC customers. This includes firms who never adopted GHC within our data frame, and firms in months prior to their adoption of GHC. Inclusion of those firms which had not yet adopted GHC but did so later in the data helps with enabling better comparisons between firms. As shown in Table 1, non-GHC firms tend to be smaller than GHC firms. However, as shown in Figure A.3, early adopters of GHC (before the median start date in our data, September 2023) have very similar distributions of number of SWEs as late adopters (after the median start date). Figure A.4 shows the distribution of GH licenses by adoption timing; here we find a slightly larger separation, with late adopters having slightly more GH licenses than early adopters. Further, we implement the doubly

robust version of CSDID. This allows us to target comparison firms (both never-treated and not-yet-treated) in the treatment assignment probabilities stage according to the covariates described in Table 1, allowing us to reweight in a way reflecting the differences in these observable characteristics.

CSDID estimates average treatment effects on the treated $ATT(g, t)$ for every group g in every period t using the doubly robust, semi-parametric estimator comparison of the treated and weighted control groups. These $ATT(g, t)$ can carefully be aggregated up to different levels, as outlined by Callaway & Sant’Anna (2021). We primarily focus on the overall aggregation which uses the simple aggregation weights across g and t . We also report the event studies, which produce estimates for each period by time before and after start of treatment $e = t - g$.

The primary assumptions of CSDID are that conditional on covariates, the never- and not-yet-treated firms can provide a proper counterfactual for the change in the outcome for a treated firm. While this is an untestable assumption, the doubly robust approach and controls help, and we see in the event studies that the pre-treatment trends are parallel (the point estimates are not statistically different from zero in pre-treatment periods). This approach allows for a robust analysis of the impact of GitHub Copilot on labor market outcomes, providing valuable insights into how AI tools influence job requirements and employment trends.

While we first focus on treatment as a binary condition of any GHC licenses, there is variation in penetration within the company. Appendix Figure A.5 presents the histogram of the fraction of GH licenses that have a GHC license within-firm; we leverage this in sensitivity analysis to compare high and low penetration firms.

Given the exploratory nature of this endeavor, we primarily present results without adjusting for the multiple hypotheses beyond the adjustments the “did” package does within an outcome across periods in the event study. Nonetheless, for completeness we additionally adjust inference using Benjamini-Hochberg’s correction for false discovery rate at the most conservative level across all outcomes examined, both across types of outcomes (demand vs. supply vs. hiring, skills vs. employment) as well as type (extensive margin, count, Winsorized) (Benjamini & Hochberg, 1995). We then report

which estimates are still statistically significant at the 5% level after this broad application of multiple hypothesis correction.

5. Results

5.1. Main results

Table 2 presents an overview of the results from the primary sample specification. This research provides exploratory investigations into how GHC may impact various labor market outcomes. As such, we cast a wide net across many outcomes. Additionally, for count variables we examine different versions (unadjusted, Winsorized, and binary/extensive margin). This leads us to calculate the model for 70 outcomes.

Without making adjustments for multiple hypotheses (yet), only six of the 70 outcomes' ATT estimates are statistically significantly different from zero at the 5 percent level. None of the outcomes for skills demanded in job postings, skills of existing SWEs, or demand for workers in terms of job postings are statistically significant. Meanwhile, we find evidence in support of GHC increasing the non-programming skills of new hires as well as increasing the likelihood of hiring at least one new SWE, overall and for entry and senior individual contributor workers. There is also an increase in the number of new SWE hires at the entry level (Winsorized), and a decrease in the number of current SWEs in a GHC firm who view job postings for more senior positions.

Table 3 provides the statistics for the six statistically significant outcomes from Table 2. Appendix Table A.5 provides the same statistics for all 70 outcomes. Focusing on these statistically significant findings, the outcome with the largest percent increase over the control group mean is for the average number of non-programming skills new hires have at time of hiring. The control group's SWEs have on average 1.5 skills listed on profile that were non programming, while the treatment leads to an increase of 0.203 non-programming skills, for a 13.3% increase over the control group. The second largest proportional increase is for the probability of hiring at least one entry-level SWE in a given month. We observe a 6.6% higher rate (a 2.9 percentage points higher from treatment over the 44.1% average in the control group,). These two outcomes remain statistically significant after correcting for multiple hypotheses across all 70 outcomes.

Table 2: Overview of findings for ATT across outcomes

Category	Negative (p<0.05)	Not statistically significant	Positive (p<0.05)
Skills demanded in job postings		Average number of programming and non-programming skills associated with each job post	
Skills of existing SWEs		Average number of skills (overall and by skill group), share of skills in a given skill group.	
Skills of new hires		Average number of programming skills new hires have	(1) Average number of non-programming skills new hires have
Demand		Number and share of job postings for SWEs; number and share of job postings for SWEs with no degree requirements (US only)	
Supply	(2) Number of people who viewed SWE job postings that are more senior than their current role, extensive margin	Among existing SWEs at company, number and share of job views and applications for other job postings, both overall and for positions more senior than their current position for all but outcome (2) listed	
New hires		Number and share of new hires, overall and by level for all unlisted levels (those not in outcomes 3-6).	(3) Number of new SWE hires, extensive margin (4) Number of new SWE hires who are entry-level, extensive margin (5) Number of new SWE hires who are senior level, extensive margin (6) Number of new SWE hires who are entry-level, Winsorized

Note: Number of firms=24,517. Number of total time periods=65. Number of groups (starting months)=18. Full results in Appendix Table A.5. Skill groupings examined: programming vs. nonprogramming; soft vs. business vs. disruptive tech. vs other tech. vs. specialized industry skills. We test overall and limiting the sample of SWEs at the company to those who had at least one skill listed on platform before GHC rollout. Note, for all count outcomes, we evaluate the overall outcome, the Winsorized (at 99th percentile), and an extensive margin version taking on values of 0 (count=0) and 1 (count>0).

Table 3: Statistically significant findings, primary specification

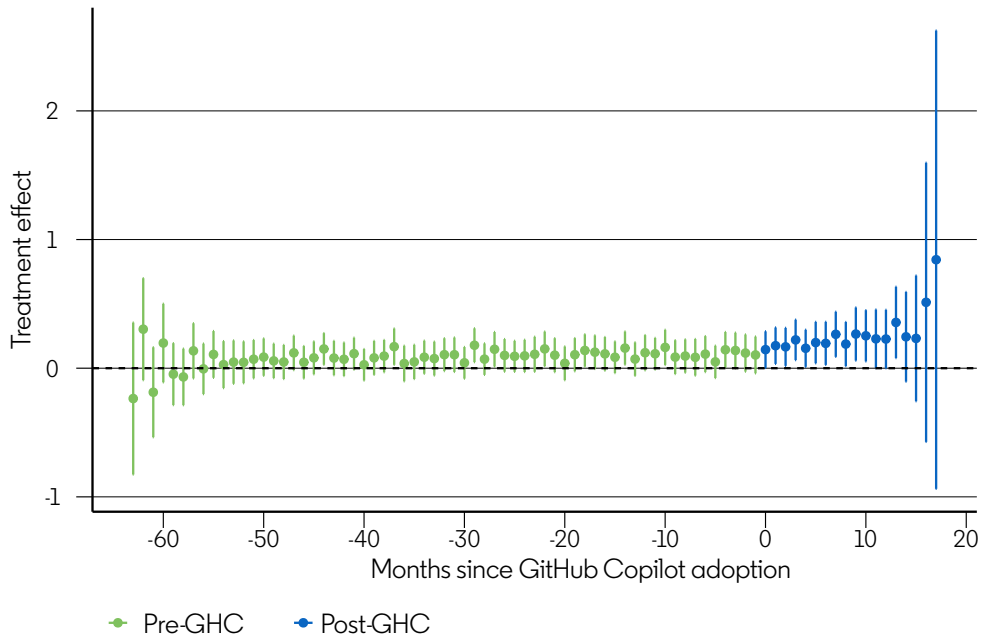
Outcome	Coef.	Control mean	% increase over control	Std. error	p-value
Skills of new hires					
Average number of non-programming skills new hires have	0.203	1.527	0.133	0.031	<0.001*
Supply					
Number of people who viewed SWE job postings that are more senior than their current role, extensive margin	-0.007	0.202	-0.033	0.003	0.033
New Hires					
Number of new SWE hires who are entry-level, extensive margin	0.029	0.441	0.066	0.004	<0.001*
Number of new SWE hires, extensive margin	0.017	0.532	0.032	0.006	0.002
Number of new SWE hires who are senior level, extensive margin	0.013	0.270	0.049	0.005	0.004
Number of new SWE hires who are entry-level, Winsorized	0.047	1.187	0.039	0.019	0.012

Note: Number of firms=24,517. Number of total time periods=65. Number of groups (starting months)=18. *Statistically significant at 5% level after Benjamini-Hochberg multiple hypothesis correction.

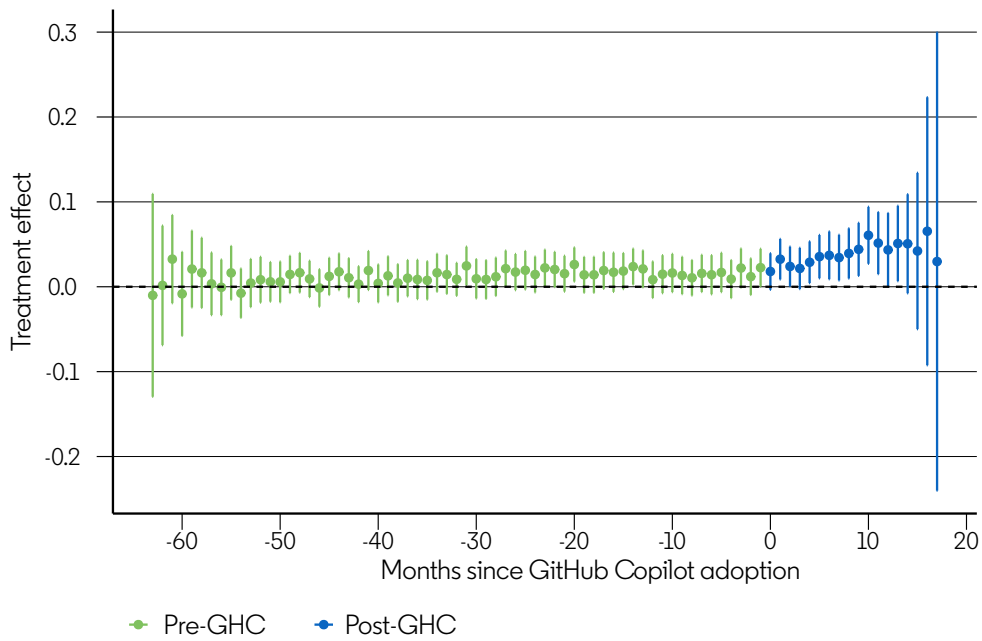
Figure 2 presents the event studies' charts for the two outcomes which remained statistically significant even after correction for multiple hypotheses. In both cases, there appears to be a very subtle increase in the treatment effect the longer that time has elapsed since first adoption of GHC. The event study charts for the other four statistically significant findings from Table 3 are presented in Appendix Figures A.6-A.9.

Figure 2: Event study charts for select outcomes

Panel 2.A: Average number of non-programming skills new hires have



Panel 2.B: Number of new SWE hires who are entry-level, extensive margin



5.2. Comparison of high and low penetration GHC firms

In the primary model from Section 5.1, the treatment group pooled together firms who had high penetration (many GHC licenses per GH license) and firms who had low penetration. However, there is a wide distribution of the percent of GH licenses having GHC licenses. Appendix Figure A.5 presents the histogram of this penetration rates, displaying this wide variation. We next split the treatment group into two subgroups: high concentration firms and low concentration firms. To do so, we calculate the median concentration rate across GHC firms, which is 0.213 (just over 1 GHC license per 5 GH licenses). We classify a GHC firm high-adoption if their maximum penetration exceeds the median, and low-adoption if it does not exceed the median.

From this, we next perform two additional analyses:

1. Treatment: high penetration firms; control: low penetration firms, never treated, and not-yet-treated firms.
2. Treatment: high penetration firms; control: low penetration firms; dropped from analysis=never-treated and not-yet-treated firms.

The first assignment thus considers low-penetration as not treated. This approach addresses the problem of including firms with low GHC license rates (maximum below 0.213 of all GH users at the firm having access to GHC) while still retaining a large comparison group. The second assignment posits that the best comparison group for those who adopt GHC and have high rates to be those firms that also adopt GHC, but have lower rates. However, this comes at the cost of much smaller control groups.

Table 4 presents the statistically significant results when we use treatment assignment 1. Here, we do not lose firms from our sample, but rather move low penetration firms from treatment to control. As a result, we should be able to better isolate effects of GHC adoption if they exist if higher penetration yields higher treatment impacts. We find that 20 of the outcomes & specifications out of the 70 tested are statistically significant at $p < 0.1$, with 15 of these being significant at $p < 0.05$, and five are still statistically significant after making correction for multiple hypotheses.

Table 4: Statistically significant findings, high adoption GHC vs. all other firms

	Coef.	Control mean	% increase	Std. error	p-value
Skills of existing SWEs					
Average number of new non-programming skills added per member who added at least one skill, filtered [†]	-0.038	0.588	-0.065	0.009	<0.001*
Average number of new soft skills added by SWEs who added at least one skill [†]	-0.010	0.067	-0.154	0.005	0.050
Number of SWEs at firm who added at least one skill, filtered [†]	1.012	3.650	0.277	0.396	0.011
Share of SWEs who added at least one skill that added at least one business skill, filtered [‡]	-0.008	0.109	-0.073	0.005	0.092
Share of SWEs who added at least one skill that added at least one soft skill, filtered [‡]	-0.013	0.082	-0.163	0.005	0.013
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills [§]	-0.017	0.486	-0.036	0.007	0.010
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills, filtered [‡]	-0.023	0.482	-0.048	0.008	0.006
Skills of new hires					
Average number of non-programming skills new hires have [†]	0.184	1.527	0.120	0.039	<0.001*
Demand					
Number of job posts in US with no degree requirement, extensive margin [†]	0.005	0.045	0.103	0.003	0.081
Number of jobs posts for SWEs, extensive margin [†]	0.011	0.155	0.069	0.006	0.052
Supply					
Number of people who viewed SWE job postings that are more senior than their current role, Winsorized [†]	-0.027	0.775	-0.034	0.016	0.090
Number of people who viewed SWE job postings that are more senior than current role, extensive margin [†]	-0.007	0.202	-0.034	0.004	0.097
Share of current SWEs from the firm who viewed at least one SWE job at a higher level	-0.001	0.016	-0.069	0.000	0.023
New hires					
Number of new SWE hires who are entry-level, Winsorized [†]	0.062	1.187	0.053	0.018	<0.001*
Number of new SWE hires who are entry-level, extensive margin [†]	0.026	0.441	0.059	0.006	<0.001*
Number of new SWE hires who are executive level, extensive margin [†]	0.002	0.002	0.923	0.001	0.003
Number of new SWE hires who are senior level, Winsorized [†]	0.042	0.530	0.080	0.012	0.001*
Number of new SWE hires who are senior level, extensive margin [†]	0.014	0.270	0.051	0.005	0.011
Number of new SWE hires, extensive margin [†]	0.021	0.532	0.039	0.007	0.003
Share of new SWE hires who are executive level [#]	0.001	0.001	1.192	0.000	0.017

Note: Number of firms: †24,517 ‡ 1,774 §1805 ||23904. #1294. Number of total time periods=65. Number of groups (starting months)=18. *Statistically significant at 5% level after Benjamini-Hochberg multiple hypothesis correction. Filtered = filtered to sample of members who had added at least one skill prior.

Although we do not see any impact of GHC on the skills demanded in job postings from our primary specification, in this alternative model, we find existing SWEs at GHC

firms add more skills overall, but add fewer non-programming, business, and soft skills. We still find that new hires to GHC firms tend to have more non-programming skills on their profile as a result of adopting GHC. Interestingly, for this sample specification, we find evidence that GHC adoption results in increased demand for SWEs (an increase in job posts for SWEs over what we would predict in absence of GHC). We also see a marginally significant increase in the number of job postings for SWEs with no degree requirement (sample limited to the US given data quality). We also still see increases in overall new hires, as well as increases in entry level and senior IC level. We now also see evidence for increases in the hiring of SWEs at the executive level.

Table 5 presents the results with the second treatment and sample specification: we compare high adoption GHC firms only to low adoption GHC firms, and do not include never-treated firms. Interestingly, the results are again largely consistent, especially in the strongest results: GHC leads to more hiring of ICs, and those new hires have more non-programming skills. The other results are also often similar, such as more skills added by existing SWEs, but fewer of those skills being non-programming, industry, soft, or business skills.

5.3. Sample sensitivity analysis

We next examine how sensitive our findings are to some of the data construction assumptions. Specifically, we test different matching criteria between LinkedIn and GitHub data, as well as minimum average LinkedIn SWE count thresholds.

For the match quality criteria, in the main specification we include any firm we matched between LinkedIn and GitHub. We alternatively test strong matches only (as defined in the Appendix), as well as retaining all firms passing the minimum SWE threshold in LinkedIn data. This means that we include non-GitHub firms with SWEs as potential control companies.

For the minimum SWE threshold, in the main specification we require at least 10 SWEs in the LinkedIn data on average each month. We alternatively test 30 SWEs minimum as well as 100 SWEs minimum.

Table 5: Statistically significant findings, high adoption GHC vs. low adoption GHC

	Coef.	Control mean	% increase	Std. error	p-value
Skills of existing SWEs					
Average number of new industry skills added by SWEs who added at least one skill, filtered [†]	-0.026	0.463	-0.056	0.015	0.078
Average number of new non-programming skills added per member who added at least one skill, filtered [†]	-0.051	0.588	-0.088	0.011	<0.001*
Average number of new soft skills added by SWEs who added at least one skill, filtered [†]	-0.012	0.067	-0.182	0.006	0.029
Number of SWEs at firm who added at least one skill, filtered [†]	0.609	3.650	0.167	0.333	0.068
Share of SWEs who added at least one skill that added at least one business skill, filtered [‡]	-0.010	0.109	-0.094	0.005	0.049
Share of SWEs who added at least one skill that added at least one soft skill, filtered [‡]	-0.015	0.082	-0.190	0.006	0.010
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills§	-0.017	0.486	-0.035	0.007	0.014
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills, filtered [‡]	-0.020	0.482	-0.042	0.008	0.014
Skills of new hires					
Average number of non-programming skills new hires have [†]	0.103	1.527	0.067	0.045	0.022
Supply					
Number of people who viewed SWE job postings from the firm [†]	0.713	5.063	0.141	0.425	0.094
Number of people who viewed SWE job postings that are more senior than their current role, Winsorized [†]	-0.032	0.775	-0.042	0.016	0.045
Number of people who viewed SWE job postings that are more senior than their current role, extensive margin [†]	-0.009	0.202	-0.044	0.004	0.045
Share of current SWEs from the firm who viewed at least one SWE job at a higher level	-0.001	0.016	-0.075	0.001	0.023
New hires					
Number of new SWE hires who are entry-level [†]	0.326	2.619	0.124	0.158	0.040
Number of new SWE hires who are entry-level, Winsorized [†]	0.064	1.187	0.054	0.020	0.001
Number of new SWE hires who are executive level, extensive margin [†]	0.001	0.002	0.726	0.001	0.020
Number of new SWE hires who are senior level, Winsorized [†]	0.045	0.530	0.085	0.012	<0.001*
Share of new SWE hires who are executive level [#]	0.001	0.001	1.171	0.000	0.022

Note: Number of firms: †8,181 ‡885 §895 ||7986. #712. Number of total time periods=65. Number of groups (starting months)=18. *Statistically significant at 5% level after Benjamini-Hochberg multiple hypothesis correction. Filtered = filtered to sample of members who had added at least one skill prior.

The results of these sensitivity checks are shown in Appendix Table A.6 for each outcome. By and large, they tell a very consistent narrative of the estimated impacts as the main specification. For the statistically significant findings from our main specification (Table 3), five of the six outcomes remain statistically significant in the same direction across each of the seven alternative samples tested. Specifically, across all eight samples we find GHC leads to a higher average number of non-programming skills new hires have, a lower number of people who viewed SWE job postings that are more senior than their current role (extensive margin), a higher probability of hiring at least one new SWE each month, and a higher number of new SWE hires who are entry-level (Winsorized and extensive margin). The sixth outcome – the probability of hiring at least one new senior-level IC SWE each month – remains statistically significant and the same sign in four of the seven alternative specifications, and not statistically significant in the other three.

There are two additional outcomes which, although not statistically significant in our main specification, are significant in several of the other samples. First, in some of these samples, we find GHC leads to more non-programming skills added by existing SWEs (when filtered to LI members with at least one skill pre-treatment). Second, we find a positive impact on the number of new SWE hires who are senior level individual contributors (Winsorized). The latter we found significant at the extensive margin in the main specification, but in every other specification we now find the count also significant and positive. Additionally, one other outcome is significant in two of the alternative specifications (but not the main specification): new hires having more programming skills (in addition to having more non-programming skills).

All of the remaining findings are not statistically significant in all or all but one (for two outcomes) of the 8 sample specifications in the sensitivity analysis. The consistency of the results across outcomes – both for those statistically significant and for those not – offers reassurance that at least our results are not artifacts of specific data construction decisions made in this paper.

6. Discussion

This paper aims to provide early evidence of the estimated causal impact that a specific generative AI tool—GitHub Copilot (GHC)—has on a specific population—software engineers (SWEs). To do so, we join data on firm adoption of GHC with LinkedIn data on labor market outcomes. Using a difference-in-differences approach with staggered treatment adoption and multiple time periods, we examine several outcomes. We find that GHC leads to 13.3% more non-programming skills among new hires over the control group mean, 3.3% fewer views of job postings by existing SWEs to outside more senior positions, 3.2 percentage point (pp) higher probability of hiring new SWEs each month, led primarily by more entry-level individual contributor (IC) SWEs (6.6 pp higher likelihood, with 3.2% higher count hired monthly), as well as 4.9% higher probability of hiring at least one senior IC each month. Interestingly, we find no evidence of new skills listed by existing SWEs in the timeframe investigated in our main specification, and no shift in skills demanded in job postings or in the number of job postings, among other hypotheses investigated. These results are very consistent across several alternative sample decisions regarding match quality and firm SWE size.

In alternative specifications, we define the treatment group by above-median maximum fraction of GH licenses with GHC licenses. We compare these against all other firms (low-penetration GHC, not-yet-treated, and never-treated) as well as just against low-penetration GHC. We find similar results as in the main specification, but additionally find evidence of GHC leading to two other groups of findings: existing SWEs add more skills due to GHC, particularly non-programming skills; and a higher demand (measured by job posts) for SWEs, including a higher probability of job posts without degree requirements.

Our findings contribute to the literature on the impacts of GAI on work and organizations, as well as the literature on the determinants and outcomes of SWE labor forces. We also shed light on the potential channels through which generative AI tools affect SWE teams, such as skill acquisition, demand and supply, and hiring. We hypothesize that GAI augments worker productivity in a way that helps them upskill in non-programming, people skills, and leads to increases in demand for these higher

productivity workers. While we are unable to causally identify these hypotheses as being the mechanisms under which the labor outcomes are realized, our findings are consistent with this framing.

While our paper reveals several interesting findings on the initial impacts of GAI on SWEs, our study has limitations and caveats that should be acknowledged. First, our identification strategy relies on the assumption that the control group firms offer a valid counterfactual for the trajectory of the outcomes in absence of treatment—the parallel trends assumption. While we find no statistically significant differences in pre-treatment period trends, this is not direct proof of the validity of the parallel trends after treatment assumption. Ultimately, our results hinge on these control firms being valid comparisons. The doubly robust approach we utilize, along with several sensitivity checks of different samples yielding similar results—also offer reassurance, although still not proof. The assumption that the adoption of GAI tools is exogenous to the outcomes of interest, conditional on the covariates and fixed effects, assumes no unobserved confounders or reverse causality. For example, it is possible that some firms may adopt GAI tools in response to changes in their SWE performance or demand, or that some firms may have unobserved characteristics that make them more likely to adopt GAI tools and affect their SWE outcomes. Future research can examine alternative approaches and evidence in support of these hypotheses.

Second, our data sources and sample selection may introduce some external validity and measurement error biases. We rely on data from LinkedIn and GitHub, which may not capture the full population of SWEs and firms in the industry. Additionally, our outcome variables and GAI tools (GHC specifically) may not capture the full spectrum of SWE performance and innovation. If SWEs are using GAI tools outside of GHC (such as ChatGPT), then this could attenuate our findings through treatment cross-over.

Third, we have to match firms across these two platforms using heuristic methods, which may result in some false positives or negatives. Additionally, we have to impose some minimum thresholds on the number of SWEs and the duration of activity for each firm, which may exclude some smaller or newer firms from our sample. Where possible,

we estimate several different versions of the sampling and outcomes, and we find consistent results.

Fourth, our study does not account for the potential spillover or general equilibrium effects of GAI tools on the SWE labor market and industry. We treat each firm as an independent unit of analysis (making the SUTVA assumption), but it is possible that the adoption of GAI tools by some firms may affect the outcomes of other firms through channels such as knowledge diffusion, competition, or network externalities. Moreover, we focus on the short-term and medium-term effects of GAI tools, but it is possible that the long-term effects may differ due to general equilibrium effects if eventually all SWEs have access and utilization of GAI, as well as changes in the supply and demand of SWE skills, wages, education, or regulation. To address these issues, future research could use dynamic models of the SWE labor market and industry, directly examine spillover through networks, and examine longer time horizons and counterfactual scenarios.

Despite these limitations, our study has important implications for both research and practice. Ultimately, we focus on a set of outcomes that are observable and quantifiable from our data sources, but there may be other aspects of SWE performance and innovation that are more qualitative or subjective, such as code readability, maintainability, security, or user satisfaction. Other research has started to find positive relationships between GAI and these productivity measures, and hiring. This study contributes to this research agenda by examining the downstream impacts on skills and labor market outcomes.

For researchers, our study suggests that GAI tools are a promising and fruitful area of inquiry, as they have the potential to transform the nature and outcomes of work in various domains, especially those that involve complex, creative, and cognitive tasks. For practitioners, our study provides evidence-based guidance on how GAI tools may enhance their SWE teams and organizations, as well as how to anticipate and mitigate the possible challenges and risks associated with these tools. Ultimately, this study offers potential encouragement that GAI tools may have positive and significant impacts on SWE teams and organizations.

References

- Acemoglu, D., & Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics* (Vol. 4, pp. 1043–1171). Elsevier.
<https://www.sciencedirect.com/science/article/pii/S0169721811024105>
- Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2), 3–30.
<https://doi.org/10.1257/jep.33.2.3>
- Autor, D. H., & Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5), 1553–1597.
<https://doi.org/10.1257/aer.103.5.1553>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Boussioux, L., N. Lane, J., Zhang, M., Jacimovic, V., & Lakhani, K. R. (2023). The Crowdless Future? How Generative AI Is Shaping the Future of Human Crowdsourcing. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4533642>
- Brynjolfsson, E., Li, D., & Raymond, L. (2023a). *Generative AI at Work* (w31161; p. w31161). National Bureau of Economic Research. <https://doi.org/10.3386/w31161>
- Brynjolfsson, E., Li, D., & Raymond, L. (2023b). *Generative AI at Work* (w31161; Issue w31161, p. w31161). National Bureau of Economic Research. <https://doi.org/10.3386/w31161>
- Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, 108, 43–47.
<https://doi.org/10.1257/pandp.20181019>
- Callaway, B., & Sant’Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
<https://doi.org/10.1016/j.jeconom.2020.12.001>
- Dell’Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Kraymer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4573321>
- Doshi, A. R., & Hauser, O. (2023). Generative Artificial Intelligence Enhances Creativity. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4535536>
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2024). GPTs are GPTs: Labor market impact potential of LLMs. *Science*, 384(6702), 1306–1308.
<https://doi.org/10.1126/science.adj0998>
- Felten, E. W., Raj, M., & Seamans, R. (2018). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings*, 108, 54–57.
<https://doi.org/10.1257/pandp.20181021>
- Furman, J., & Seamans, R. (2019). AI and the Economy. *Innovation Policy and the Economy*, 19, 161–191. <https://doi.org/10.1086/699936>

- Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. (2023). ChatGPT has Mastered the Principles of Economics: Now What? *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4356034>
- Girotra, K., Meincke, L., Terwiesch, C., & Ulrich, K. T. (2023). Ideas are Dimes a Dozen: Large Language Models for Idea Generation in Innovation. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4526071>
- Goldin, C., & Katz, L. F. (1998). The Origins of Technology-Skill Complementarity. *The Quarterly Journal of Economics*, 113(3), 693–732. <https://doi.org/10.1162/003355398555720>
- Kalliamvakou, E. (2022, September). Research: Quantifying GitHub Copilot’s impact on developer productivity and happiness. *Research: Quantifying GitHub Copilot’s Impact on Developer Productivity and Happiness*. <https://github.blog/2022-09-07-research-quantifying-github-copilots-impact-on-developer-productivity-and-happiness/>
- Katz, L. F., & Murphy, K. M. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1), 35–78.
<https://doi.org/10.2307/2118323>
- Kimbrough, K., & Carpanelli, M. (2023, August). *Preparing the Workforce for Generative AI*. LinkedIn Economic Graph Institute.
<https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/PDF/preparing-the-workforce-for-generative-ai.pdf>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health*, 2(2), e0000198.
<https://doi.org/10.1371/journal.pdig.0000198>
- Otis, N. G., Clarke, R. P., Delecourt, S., Holtz, D., & Koning, R. (2023). *The Uneven Impact of Generative AI on Entrepreneurial Performance*. <https://doi.org/10.31219/osf.io/hdjpk>
- Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot (Version 1)*. arXiv.
<https://doi.org/10.48550/ARXIV.2302.06590>
- Senz, K. (2023, April). *Is AI Coming for Your Job?* <https://hbswk.hbs.edu/item/is-ai-coming-for-your-job>
- Webb, M. (2019). The Impact of Artificial Intelligence on the Labor Market. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3482150>
- Zhao, S. (2023, February). *GitHub Copilot now has a better AI model and new capabilities*. <https://github.blog/2023-02-14-github-copilot-now-has-a-better-ai-model-and-new-capabilities/>
- Zhu, T. J., Fritzler, A., & Orlowski, J. (2020). Jobs, Skills and Migration Trends Methodology & Validation Results. *World Bank Group*.
<https://documents1.worldbank.org/curated/en/827991542143093021/pdf/World-Bank-Group-Linked-In-Data-Insights-Jobs-Skills-and-Migration-Trends-Methodology-and-Validation-Results.pdf>

APPENDIX

Matching Methodology

GitHub and LinkedIn companies are matched by company name. Before matching, GitHub companies are aggregated to their highest parent company and the company names are cleaned by removing non-alphanumeric characters and common company phrases (e.g., “corporation” or “inc” or “enterprise”) to extract a consistent name root to match on. The GitHub primary company name used to match is the most recent name used followed by the second most recent name used if there was a name change with the same GitHub account ID. Similarly, LinkedIn names are cleaned by removing the same non-alphanumeric characters and company phrases before matching. LinkedIn companies are filtered to having an average of at least 10 software engineers (SWEs) listing the company as a LinkedIn position across the time period starting in January 2018. LinkedIn companies have two company names: 1) raw name which is unique and 2) canonical name which is not unique.

First, GitHub companies with the most GitHub licenses that produce matching edge cases ($n=30$) are manually matched to their corresponding LinkedIn companies and are labeled as “strong match”. Next, GitHub company names are exact string matched to both LinkedIn raw and canonical company names separately and the LinkedIn company with the highest number of current software engineers is selected. If the GitHub company names matched to both LinkedIn raw and canonical names, then the LinkedIn company with the highest number of SWEs is selected and also labeled as “strong match”.

For the GitHub company names that match to only one or none of the LinkedIn raw or canonical names, we then try to fuzzy-match them to LinkedIn raw names. We filter to names that are at least 3 characters long and search for the GitHub name pattern inside the LinkedIn name (e.g., “gh_in_li” match where “fictionalcorp” is found in “fictionalcorptechnologies”) and select the match with the highest number of SWEs. We repeat this vice versa for searching for the LinkedIn name pattern inside the GitHub name (“li_in_gh”).

The matches from raw name, canonical name, gh_in_li, and li_in_gh matching are

pooled and the match with the highest number of SWEs is selected. These selected matches are then deduped by selecting the matches with the most GitHub licenses while the rest are unmatched. This produces a one-to-one mapping between GitHub and LinkedIn companies.

Manual and exact matches are labeled as “strong” and 90% of inferred fuzzy matches (“li_in_gh” and “gh_in_li”) are labeled as “weak”. 10% of inferred fuzzy matches that are manually validated with more SWE and fewer inferred matches meaning the string is not common are labeled are “strong”. The following tables show the breakdown of match confidence and matching method.

Table A.1 Match quality for primary sample

Match confidence	Number of firms	% of sample
Strong match	15,070	61.5%
Weak match	9,447	38.5%

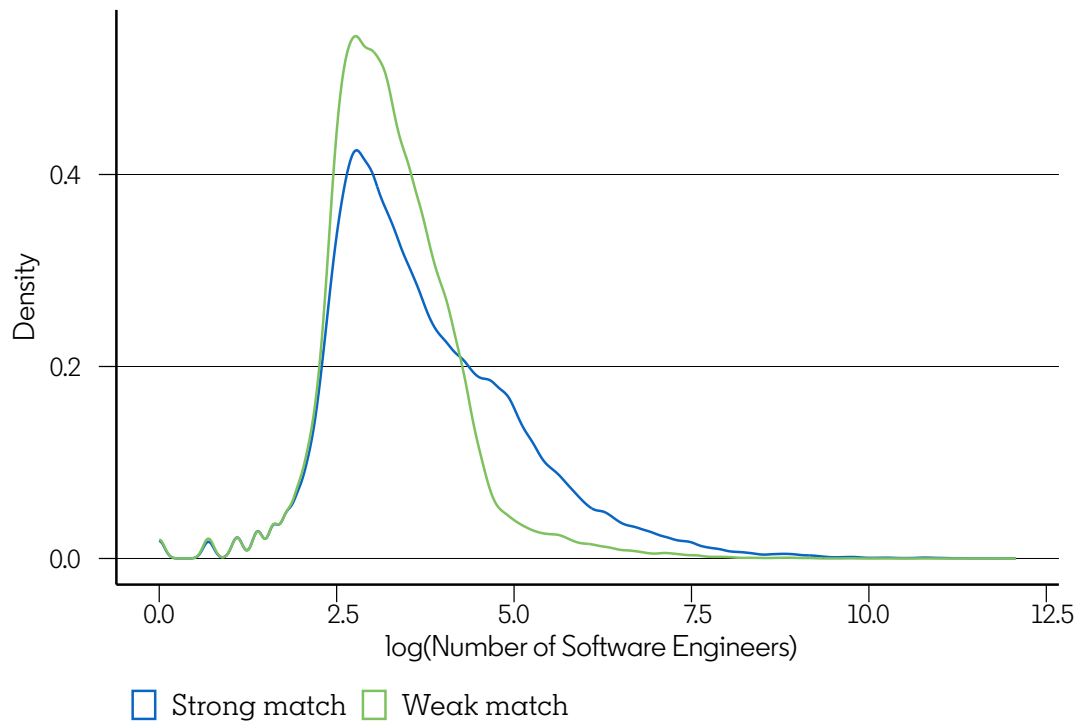
Table A.2 Match quality by type of match

Match confidence	Match method	Number of firms	% of sample
Strong	canonical_name	1,649	6.7%
Strong	gh_in_li	638	2.6%
Strong	li_in_gh	1,215	5.0%
Strong	manual	30	0.1%
Strong	raw_name	11,538	47.1%
Weak	gh_in_li	4,733	19.3%
Weak	li_in_gh	4,714	19.2%

Table A.3: Sample composition by match confidence type

Match confidence	GHC licensee	Non-GHC licensee	% GHC	% not GHC
Strong match	52,156	927,394	5.3%	94.7%
Weak match	16,573	597,482	2.7%	97.3%

Figure A.1: Distribution of SWEs by match quality



Supplementary Figures and Tables

Figure A.2: Distribution of number of months firms held GHC licenses in year after first adoption

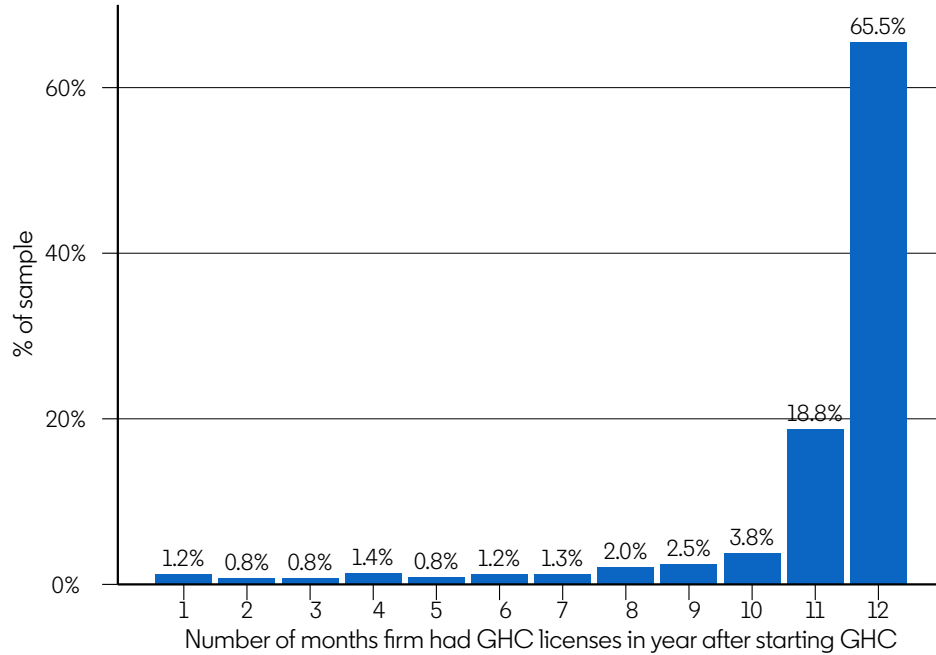


Figure A.3: Distribution of firm size (number of SWEs), by GHC adoption timing

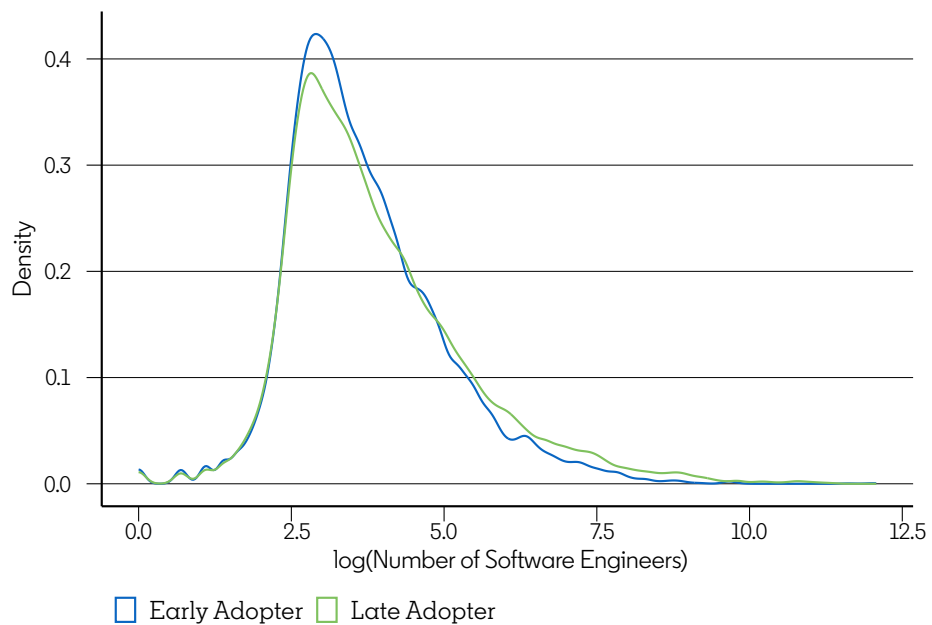


Figure A.4: Distribution of number of GH licenses, by GHC adoption timing

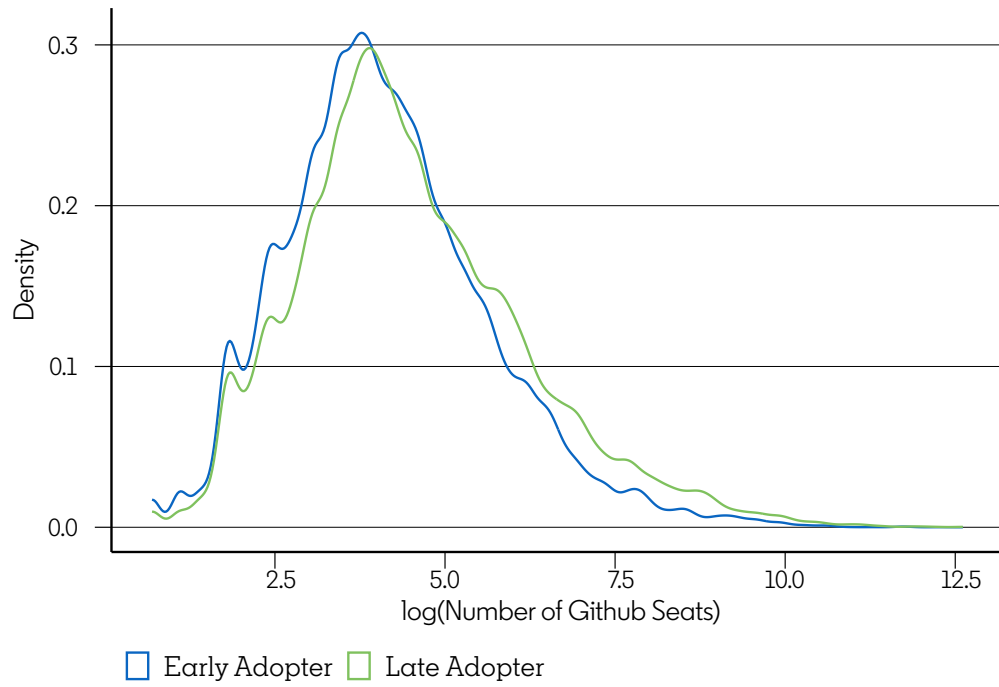


Figure A.5: Distribution of GHC licenses per GH licenses

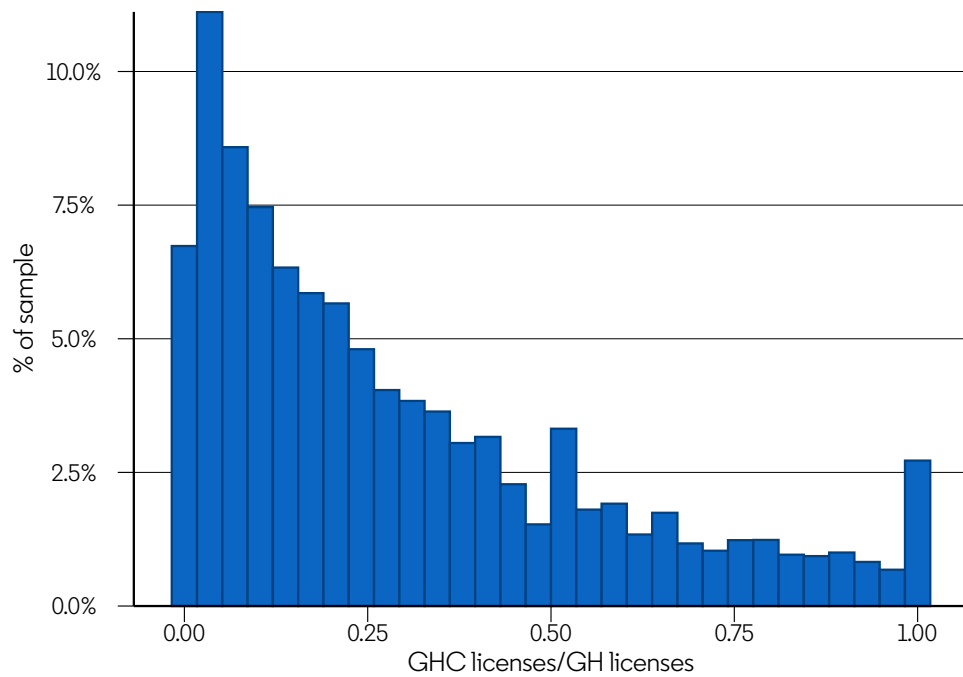


Figure A.6 Number of people who viewed SWE job postings that are more senior than their current role, extensive margin

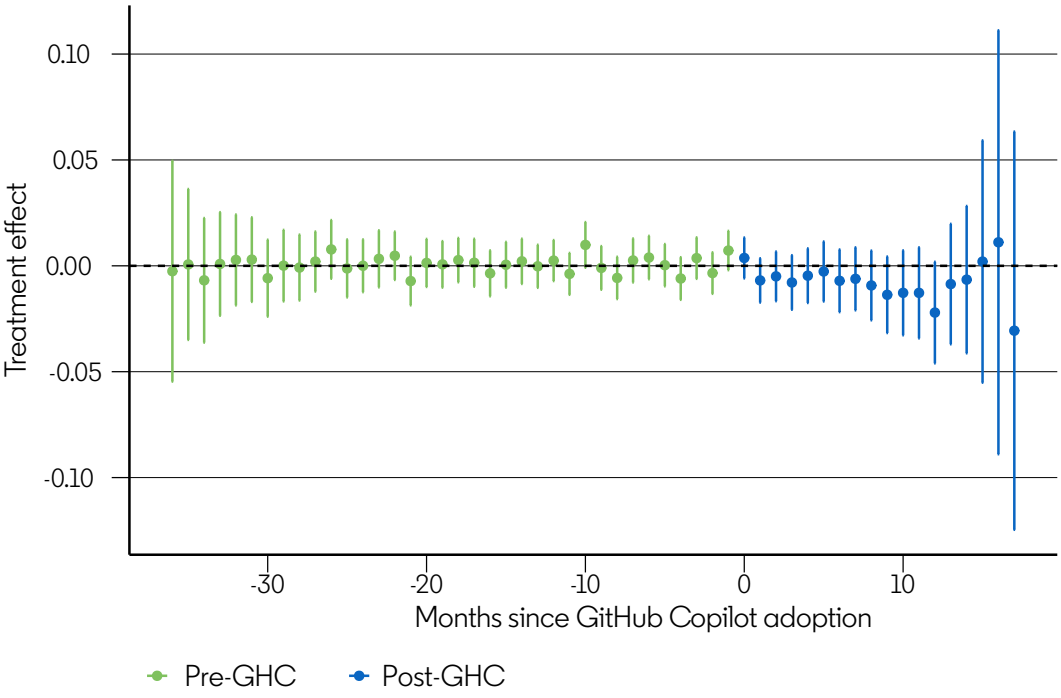


Figure A.7 Number of new SWE hires who are entry-level, Winsorized

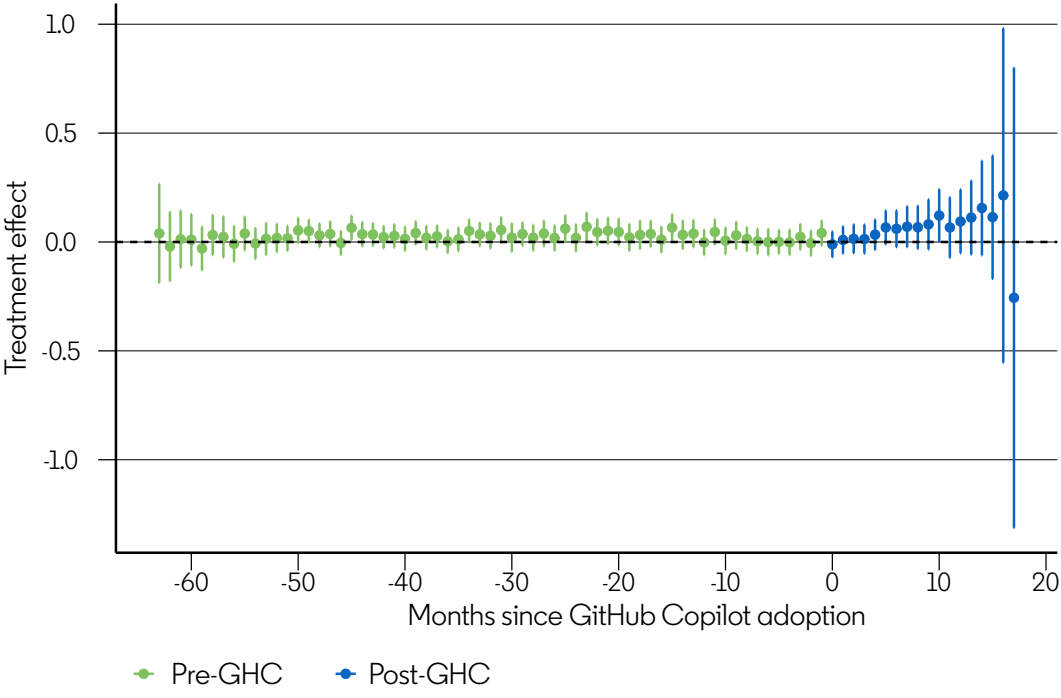


Figure A.8 Number of new SWE hires who are senior level, extensive margin

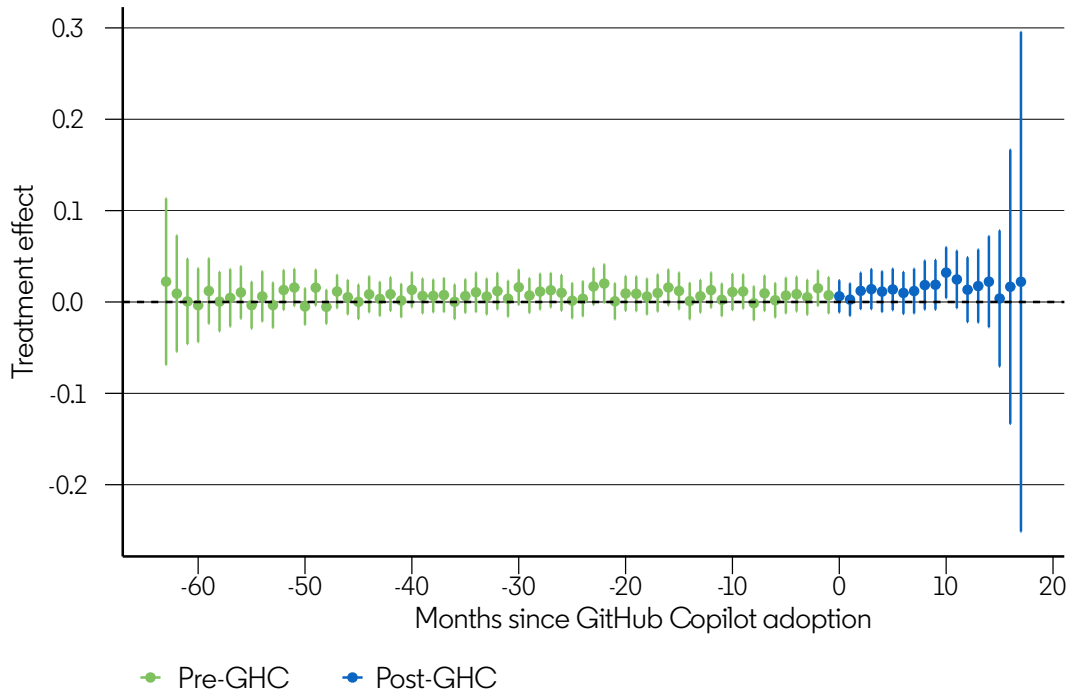


Figure A.9 Number of new SWE hires, extensive margin

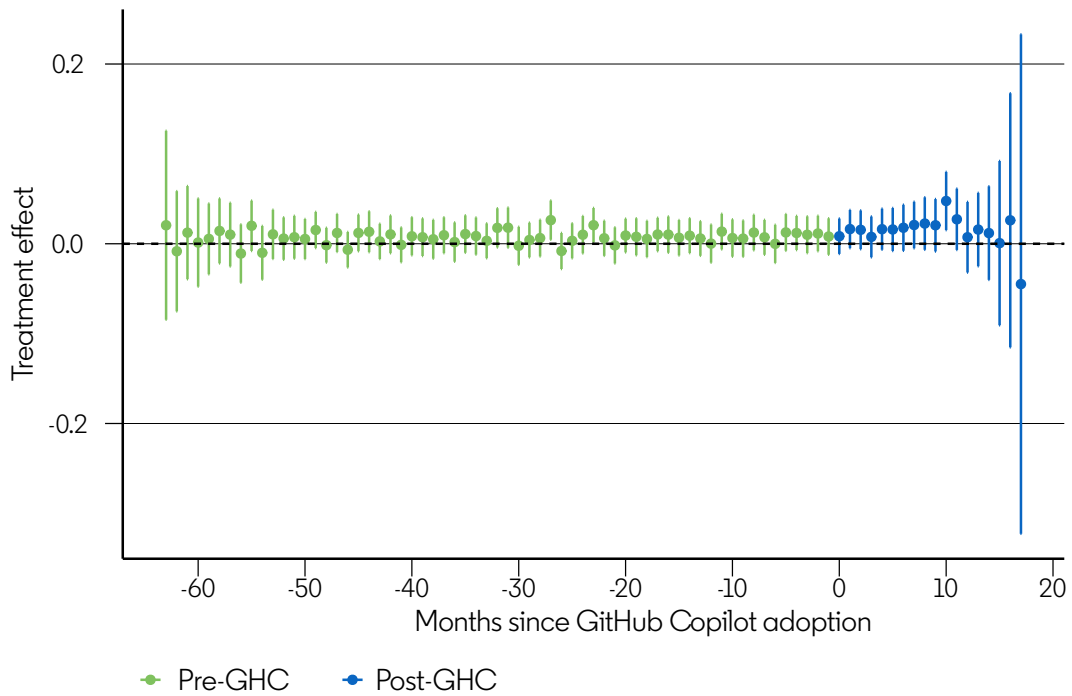


Table A.4: Descriptive statistics on outcomes

	All firms	Firm/months without GHC	Firm/months with GHC
Skills demanded in job postings			
Average number of non-programming skills associated with each job post	0.326	0.313	0.616
Average number of non-programming skills associated with each job post	(1.197)	(1.171)	(1.650)
Average number of programming skills associated with each job post	0.518	0.506	0.786
Average number of programming skills associated with each job post	(1.751)	(1.739)	(1.993)
Skills of existing SWEs			
Average number of new business skills added by SWEs who added at least one skill, filtered	0.098	0.096	0.154
Average number of new business skills added by SWEs who added at least one skill, filtered	(0.408)	(0.405)	(0.474)
Average number of new disruptive tech skills added by SWEs who added at least one skill, filtered	0.692	0.693	0.676
Average number of new disruptive tech skills added by SWEs who added at least one skill, filtered	(1.276)	(1.286)	(1.030)
Average number of new industry skills added by SWEs who added at least one skill, filtered	0.479	0.465	0.790
Average number of new industry skills added by SWEs who added at least one skill, filtered	(0.984)	(0.971)	(1.203)
Average number of new non-programming skills added per member who added at least one skill	0.642	0.628	0.955
Average number of new non-programming skills added per member who added at least one skill	(1.331)	(1.323)	(1.472)
Average number of new non-programming skills added per member who added at least one skill, filtered	0.599	0.590	0.806
Average number of new non-programming skills added per member who added at least one skill, filtered	(1.275)	(1.272)	(1.339)
Average number of new programming skills added per member who added at least one skill	1.584	1.570	1.896
Average number of new programming skills added per member who added at least one skill	(2.533)	(2.535)	(2.458)
Average number of new programming skills added per member who added at least one skill, filtered	1.472	1.471	1.493
Average number of new programming skills added per member who added at least one skill, filtered	(2.417)	(2.434)	(2.001)

Average number of new soft skills added by SWEs who added at least one skill, filtered	0.070	0.067	0.134
Average number of new soft skills added by SWEs who added at least one skill, filtered	(0.301)	(0.297)	(0.378)
Average number of new tech skills added by SWEs who added at least one skill, filtered	1.073	1.070	1.141
Average number of new tech skills added by SWEs who added at least one skill, filtered	(1.855)	(1.865)	(1.601)
Number of SWEs at firm who added at least one skill in the given month	4.684	4.166	16.174
Number of SWEs at firm who added at least one skill in the given month	(69.679)	(62.610)	(159.577)
Number of SWEs at firm who added at least one skill, filtered	4.055	3.732	11.224
Number of SWEs at firm who added at least one skill, filtered	(55.857)	(53.116)	(98.456)
Number of SWEs who added at least one non-programming skill in the given month	2.367	2.083	8.658
Number of SWEs who added at least one non-programming skill in the given month	(35.937)	(31.948)	(85.193)
Number of SWEs who added at least one non-programming skill in the given month, filtered	2.014	1.842	5.839
Number of SWEs who added at least one non-programming skill in the given month, filtered	(28.022)	(26.515)	(50.926)
Number of SWEs who added at least one programming skill in the given month	3.855	3.459	12.630
Number of SWEs who added at least one programming skill in the given month	(56.641)	(51.443)	(124.874)
Number of SWEs who added at least one programming skill in the given month, filtered	3.327	3.093	8.531
Number of SWEs who added at least one programming skill in the given month, filtered	(45.647)	(43.860)	(74.858)
Share of SWEs who added at least one skill that added at least one business skill, filtered	0.111	0.110	0.131
Share of SWEs who added at least one skill that added at least one business skill, filtered	(0.245)	(0.245)	(0.236)
Share of SWEs who added at least one skill that added at least one disruptive tech skill, filtered	0.570	0.575	0.491
Share of SWEs who added at least one skill that added at least one disruptive tech skill, filtered	(0.387)	(0.389)	(0.353)
Share of SWEs who added at least one skill that added at least one industry skill, filtered	0.459	0.454	0.547
Share of SWEs who added at least one skill that added at least one industry skill, filtered	(0.389)	(0.390)	(0.352)

Share of SWEs who added at least one skill that added at least one soft skill, filtered	0.084	0.082	0.126
Share of SWEs who added at least one skill that added at least one soft skill, filtered	(0.213)	(0.212)	(0.231)
Share of SWEs who added at least one skill that added at least one tech skill, filtered	0.708	0.710	0.659
Share of SWEs who added at least one skill that added at least one tech skill, filtered	(0.354)	(0.355)	(0.336)
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills	0.489	0.487	0.519
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills	(0.385)	(0.387)	(0.342)
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills, filtered	0.484	0.482	0.508
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills, filtered	(0.390)	(0.392)	(0.354)
Share of SWEs who added at least one skill who had at least one of those skills be programming skills	0.829	0.832	0.776
Share of SWEs who added at least one skill who had at least one of those skills be programming skills	(0.290)	(0.290)	(0.286)
Share of SWEs who added at least one skill who had at least one of those skills be programming skills, filtered	0.825	0.829	0.760
Share of SWEs who added at least one skill who had at least one of those skills be programming skills, filtered	(0.296)	(0.296)	(0.302)
Skills of new hires			
Average number of non-programming skills new hires have	1.548	1.531	1.936
Average number of non-programming skills new hires have	(2.764)	(2.754)	(2.960)
Average number of programming skills new hires have	4.030	3.994	4.824
Average number of programming skills new hires have	(6.014)	(5.993)	(6.399)
Demand			
Number of job posts in US for SWEs with no degree requirement	0.029	0.028	0.047
Number of job posts in US for SWEs with no degree requirement	(2.650)	(2.685)	(1.717)
Number of job posts in US for SWEs with no degree requirement, extensive margin	0.004	0.004	0.007
Number of job posts in US for SWEs with no degree requirement, extensive margin	(0.062)	(0.060)	(0.086)
Number of job posts in US with no degree requirement	3.296	3.115	7.304
Number of job posts in US with no degree requirement	(159.034)	(150.864)	(285.393)

Number of job posts in US with no degree requirement, Winsorized	0.048	0.045	0.098
Number of job posts in US with no degree requirement, Winsorized	(0.213)	(0.208)	(0.298)
Number of job posts in US with no degree requirement, extensive margin	0.048	0.045	0.098
Number of job posts in US with no degree requirement, extensive margin	(0.213)	(0.208)	(0.298)
Number of jobs posts for SWEs	3.500	3.367	6.463
Number of jobs posts for SWEs	(80.997)	(81.749)	(61.947)
Number of jobs posts for SWEs, Winsorized	0.545	0.517	1.175
Number of jobs posts for SWEs, Winsorized	(1.494)	(1.457)	(2.054)
Number of jobs posts for SWEs, extensive margin	0.164	0.156	0.336
Number of jobs posts for SWEs, extensive margin	(0.370)	(0.363)	(0.472)
Share of firm's job posts for SWEs that do not require degree	0.020	0.020	0.021
Share of firm's job posts for SWEs that do not require degree	(0.122)	(0.122)	(0.124)
Share of firm's job posts which are for SWEs	0.173	0.174	0.161
Share of firm's job posts which are for SWEs	(0.249)	(0.250)	(0.230)
Supply			
Number of job applicants for SWE job postings from the firm	1.602	1.350	7.186
Number of job applicants for SWE job postings from the firm	(20.049)	(17.406)	(50.659)
Number of job applicants for SWE job postings from the firm, Winsorized	0.458	0.424	1.225
Number of job applicants for SWE job postings from the firm, Winsorized	(1.222)	(1.172)	(1.872)
Number of job applicants for SWE job postings from the firm, extensive margin	0.170	0.161	0.386
Number of job applicants for SWE job postings from the firm, extensive margin	(0.376)	(0.367)	(0.487)
Number of job applicants for SWE job postings more senior than their current position	0.694	0.580	3.236
Number of job applicants for SWE job postings more senior than their current position	(10.014)	(8.765)	(24.778)
Number of job applicants for SWE job postings more senior than their current position, Winsorized	0.177	0.164	0.475
Number of job applicants for SWE job postings more senior than their current position, Winsorized	(0.517)	(0.497)	(0.784)
Number of job applicants for SWE job postings more senior than their current position, extensive margin	0.117	0.109	0.292
Number of job applicants for SWE job postings more senior than their current position, extensive margin	(0.321)	(0.311)	(0.455)

Number of people who viewed SWE job postings from the firm	5.966	5.200	22.955
Number of people who viewed SWE job postings from the firm	(77.739)	(69.697)	(179.018)
Number of people who viewed SWE job postings from the firm, Winsorized	1.654	1.546	4.036
Number of people who viewed SWE job postings from the firm, Winsorized	(3.917)	(3.782)	(5.699)
Number of people who viewed SWE job postings from the firm, extensive margin	0.255	0.244	0.496
Number of people who viewed SWE job postings from the firm, extensive margin	(0.436)	(0.429)	(0.500)
Number of people who viewed SWE job postings that are more senior than their current role	3.056	2.648	12.125
Number of people who viewed SWE job postings that are more senior than their current role	(44.511)	(39.905)	(102.579)
Number of people who viewed SWE job postings that are more senior than their current role, Winsorized	0.839	0.784	2.063
Number of people who viewed SWE job postings that are more senior than their current role, Winsorized	(2.061)	(1.990)	(3.011)
Number of people who viewed SWE job postings that are more senior than their current role, extensive margin	0.214	0.204	0.442
Number of people who viewed SWE job postings that are more senior than their current role, extensive margin	(0.410)	(0.403)	(0.497)
Share of current SWEs from the firm who applied for at least one SWE job	0.009	0.009	0.023
Share of current SWEs from the firm who applied for at least one SWE job	(0.029)	(0.028)	(0.045)
Share of current SWEs from the firm who applied for at least one SWE job at a higher level	0.004	0.004	0.010
Share of current SWEs from the firm who applied for at least one SWE job at a higher level	(0.016)	(0.015)	(0.023)
Share of current SWEs from the firm who viewed at least one SWE job	0.035	0.033	0.075
Share of current SWEs from the firm who viewed at least one SWE job	(0.082)	(0.079)	(0.115)
Share of current SWEs from the firm who viewed at least one SWE job at a higher level	0.017	0.017	0.038
Share of current SWEs from the firm who viewed at least one SWE job at a higher level	(0.046)	(0.045)	(0.064)
New hires			
Number of new SWE hires	4.231	4.161	5.784
Number of new SWE hires	(50.234)	(50.607)	(41.061)
Number of new SWE hires who are entry-level	2.662	2.631	3.342

Number of new SWE hires who are entry-level	(36.291)	(36.714)	(25.118)
Number of new SWE hires who are entry-level, Winsorized	1.191	1.189	1.230
Number of new SWE hires who are entry-level, Winsorized	(2.045)	(2.038)	(2.191)
Number of new SWE hires who are entry-level, extensive margin	0.440	0.441	0.410
Number of new SWE hires who are entry-level, extensive margin	(0.496)	(0.497)	(0.492)
Number of new SWE hires who are executive level	0.003	0.003	0.006
Number of new SWE hires who are executive level	(0.112)	(0.108)	(0.182)
Number of new SWE hires who are executive level, extensive margin	0.002	0.002	0.003
Number of new SWE hires who are executive level, extensive margin	(0.046)	(0.045)	(0.058)
Number of new SWE hires who are interns	0.309	0.300	0.509
Number of new SWE hires who are interns	(9.442)	(9.424)	(9.824)
Number of new SWE hires who are interns, Winsorized	0.063	0.063	0.082
Number of new SWE hires who are interns, Winsorized	(0.244)	(0.242)	(0.274)
Number of new SWE hires who are interns, extensive margin	0.063	0.063	0.082
Number of new SWE hires who are interns, extensive margin	(0.244)	(0.242)	(0.274)
Number of new SWE hires who are senior level	1.262	1.232	1.929
Number of new SWE hires who are senior level	(12.976)	(12.899)	(14.559)
Number of new SWE hires who are senior level, Winsorized	0.538	0.532	0.679
Number of new SWE hires who are senior level, Winsorized	(1.081)	(1.073)	(1.228)
Number of new SWE hires who are senior level, extensive margin	0.273	0.271	0.314
Number of new SWE hires who are senior level, extensive margin	(0.445)	(0.444)	(0.464)
Number of new SWE hires, Winsorized	1.865	1.855	2.093
Number of new SWE hires, Winsorized	(3.076)	(3.059)	(3.422)
Number of new SWE hires, extensive margin	0.532	0.533	0.523
Number of new SWE hires, extensive margin	(0.499)	(0.499)	(0.499)
Share of new SWE hires who are executive level	0.001	0.001	0.001
Share of new SWE hires who are executive level	(0.020)	(0.020)	(0.017)
Share of new hires who are SWEs	0.183	0.184	0.151
Share of new hires who are SWEs	(0.270)	(0.271)	(0.233)

Note: mean and standard deviation (in parentheses) reported.

Appendix Table A.5: Full results from primary sample

	Coef.	Control mean	% increase over control	Std. Err.	p-value	# firms
Skills demanded in job postings						
Average number of non-programming skills associated with each job post	0.016	0.311	0.050	0.021	0.456	24,517
Average number of programming skills associated with each job post	-0.004	0.504	-0.009	0.024	0.852	24,517
Skills of existing SWEs						
Average number of new business skills added by SWEs who added at least one skill, filtered to members who have added skills previously	-0.003	0.095	-0.031	0.005	0.566	24,517
Average number of new disruptive tech skills added by SWEs who added at least one skill, filtered to members who have added skills previously	0.006	0.692	0.008	0.013	0.669	24,517
Average number of new industry skills added by SWEs who added at least one skill, filtered to members who have added skills previously	-0.006	0.463	-0.013	0.013	0.626	24,517
Average number of new non-programming skills added per member who added at least one skill	-0.004	0.625	-0.006	0.016	0.829	24,517
Average number of new non-programming skills added per member who added at least one skill, filtered to members who have added skills previously	-0.016	0.588	-0.027	0.010	0.120	24,517
Average number of new programming skills added per member who added at least one skill	0.032	1.567	0.021	0.029	0.264	24,517
Average number of new programming skills added per member who added at least one skill, filtered to members who have added skills previously	0.033	1.470	0.023	0.024	0.169	24,517

	Coef.	Control mean	% increase over control	Std. Err.	p-value	# firms
Average number of new soft skills added by SWEs who added at least one skill, filtered to members who have added skills previously	-0.004	0.067	-0.067	0.004	0.272	24,517
Average number of new tech skills added by SWEs who added at least one skill, filtered to members who have added skills previously	0.021	1.069	0.020	0.017	0.205	24,517
Number of SWEs at firm who added at least one skill in the given month	0.273	4.031	0.068	0.667	0.682	24,517
Number of SWEs at firm who added at least one skill, filtered to members who have added skills previously	0.721	3.650	0.198	0.818	0.378	24,517
Number of SWEs who added at least one non-programming skill in the given month	0.090	2.010	0.045	0.488	0.854	24,517
Number of SWEs who added at least one non-programming skill in the given month, filtered to members who have added skills previously	-0.053	1.798	-0.030	0.376	0.887	24,517
Number of SWEs who added at least one programming skill in the given month	0.258	3.355	0.077	0.620	0.678	24,517
Number of SWEs who added at least one programming skill in the given month, filtered to members who have added skills previously	0.122	3.032	0.040	0.575	0.831	24,517
Share of SWEs who added at least one skill that added at least one business skill, filtered to members who have added skills previously	-0.002	0.109	-0.019	0.003	0.543	1,774
Share of SWEs who added at least one skill that added at least one disruptive tech skill, filtered to members who have added skills previously	-0.002	0.575	-0.004	0.006	0.678	1,774

	Coef.	Control mean	% increase over control	Std. Err.	p-value	# firms
Share of SWEs who added at least one skill that added at least one industry skill, filtered to members who have added skills previously	0.001	0.454	0.003	0.005	0.819	1,774
Share of SWEs who added at least one skill that added at least one soft skill, filtered to members who have added skills previously	0.000	0.082	-0.001	0.004	0.976	1,774
Share of SWEs who added at least one skill that added at least one tech skill, filtered to members who have added skills previously	-0.007	0.711	-0.010	0.006	0.214	1,774
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills	-0.007	0.486	-0.015	0.005	0.129	1,805
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills, filtered to members who have added skills previously	-0.009	0.482	-0.018	0.006	0.120	1,774
Share of SWEs who added at least one skill who had at least one of those skills be programming skills	0.003	0.832	0.004	0.003	0.346	1,805
Share of SWEs who added at least one skill who had at least one of those skills be programming skills, filtered to members who have added skills previously	0.003	0.829	0.004	0.004	0.483	1,774
Skills of new hires						
Average number of non-programming skills new hires have*	0.203	1.527	0.133	0.031	0.000	24,517
Average number of programming skills new hires have	0.090	3.988	0.022	0.070	0.198	24,517
Demand						
Number of job posts in US for SWEs with no degree requirement	0.026	0.028	0.940	0.020	0.176	24,517

	Coef.	Control mean	% increase over control	Std. Err.	p-value	# firms
Number of job posts in US for SWEs with no degree requirement, extensive margin	0.001	0.004	0.200	0.002	0.640	24,517
Number of job posts in US with no degree requirement	2.130	3.107	0.686	4.080	0.602	24,517
Number of job posts in US with no degree requirement, Winsorized	0.004	0.045	0.091	0.003	0.242	24,517
Number of job posts in US with no degree requirement, extensive margin	0.004	0.045	0.088	0.003	0.244	24,517
Number of jobs posts for SWEs	-1.560	3.334	-0.468	2.135	0.465	24,517
Number of jobs posts for SWEs, Winsorized	0.013	0.512	0.026	0.027	0.626	24,517
Number of jobs posts for SWEs, extensive margin	0.007	0.155	0.045	0.005	0.176	24,517
Share of firm's job posts for SWEs that do not require degree	-0.007	0.020	-0.327	0.007	0.323	321
Share of firm's job posts which are for SWEs	-0.003	0.174	-0.016	0.003	0.376	4,663
Supply						
Number of job applicants for SWE job postings from the firm	-1.300	1.300	-0.999	0.839	0.122	24,517
Number of job applicants for SWE job postings from the firm, Winsorized	0.012	0.418	0.028	0.010	0.259	24,517
Number of job applicants for SWE job postings from the firm, extensive margin	-0.002	0.159	-0.014	0.004	0.537	24,517
Number of job applicants for SWE job postings more senior than their current position	-0.431	0.559	-0.771	0.460	0.350	24,517
Number of job applicants for SWE job postings more senior than their current position, Winsorized	0.004	0.162	0.027	0.006	0.451	24,517
Number of job applicants for SWE job postings more senior than their current position, extensive margin	-0.001	0.108	-0.010	0.004	0.776	24,517
Number of people who viewed SWE job postings from the firm	-1.245	5.063	-0.246	1.469	0.397	24,517

	Coef.	Control mean	% increase over control	Std. Err.	p-value	# firms
Number of people who viewed SWE job postings from the firm, Winsorized	0.018	1.529	0.012	0.016	0.269	24,517
Number of people who viewed SWE job postings from the firm, extensive margin	0.000	0.242	0.000	0.003	0.980	24,517
Number of people who viewed SWE job postings that are more senior than their current role	0.101	2.579	0.039	1.142	0.929	24,517
Number of people who viewed SWE job postings that are more senior than their current role, Winsorized	-0.010	0.775	-0.013	0.012	0.402	24,517
Number of people who viewed SWE job postings that are more senior than their current role, extensive margin	-0.007	0.202	-0.033	0.003	0.033	24,517
Share of current SWEs from the firm who applied for at least one SWE job	0.000	0.008	0.008	0.000	0.862	23,904
Share of current SWEs from the firm who applied for at least one SWE job at a higher level	0.000	0.004	-0.025	0.000	0.688	23,904
Share of current SWEs from the firm who viewed at least one SWE job	0.000	0.033	-0.002	0.001	0.915	23,904
Share of current SWEs from the firm who viewed at least one SWE job at a higher level	0.000	0.016	-0.028	0.000	0.263	23,904
New hires						
Number of new SWE hires	-0.634	4.132	-0.153	0.495	0.200	24,517
Number of new SWE hires who are entry-level	-0.508	2.619	-0.194	1.131	0.653	24,517
Number of new SWE hires who are entry-level, Winsorized	0.047	1.187	0.039	0.019	0.012	24,517
Number of new SWE hires who are entry-level, extensive margin*	0.029	0.441	0.066	0.004	0.000	24,517
Number of new SWE hires who are executive level	0.002	0.003	0.817	0.003	0.383	24,517
Number of new SWE hires who are executive level, extensive margin	0.002	0.002	0.780	0.002	0.353	24,517

	Coef.	Control mean	% increase over control	Std. Err.	p- value	# firms
Number of new SWE hires who are interns	-0.042	0.298	-0.142	0.661	0.949	24,517
Number of new SWE hires who are interns, Winsorized	0.001	0.062	0.008	0.003	0.879	24,517
Number of new SWE hires who are interns, extensive margin	0.000	0.062	-0.002	0.004	0.976	24,517
Number of new SWE hires who are senior level	-0.497	1.217	-0.408	0.662	0.453	24,517
Number of new SWE hires who are senior level, Winsorized	0.016	0.530	0.030	0.011	0.159	24,517
Number of new SWE hires who are senior level, extensive margin	0.013	0.270	0.049	0.005	0.004	24,517
Number of new SWE hires, Winsorized	-0.022	1.850	-0.012	0.030	0.478	24,517
Number of new SWE hires, extensive margin	0.017	0.532	0.032	0.006	0.002	24,517
Share of new SWE hires who are executive level	0.000	0.001	-0.516	0.001	0.573	1,294
Share of new hires who are SWEs	0.002	0.185	0.009	0.002	0.321	10,203

Note: Number of total time periods=65. Number of groups (starting months)=18. Balanced sample requirements creates smaller samples when firms do not meet a given condition, such as having a job posting for SWEs every month of the sample. Filtered= filtered to sample of members who had added at least one skill prior.

Table A.6: Sensitivity analysis

Match requirement	Any	Any	Any	Strong	Strong	Strong	None	None
Minimum avg. SWE count	10	30	100	10	30	100	30	100
Skills demanded in job postings								
Average number of non-programming skills associated with each job post	∅	∅	∅	∅	∅	∅	∅	∅
Average number of programming skills associated with each job post	∅	∅	∅	∅	∅	∅	∅	∅
Skills of existing SWEs								
Average number of new business skills added by SWEs who added at least one skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Average number of new disruptive tech skills added by SWEs who added at least one skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Average number of new industry skills added by SWEs who added at least one skill, filtered	∅	∅	∅	∅	∅	-	∅	∅
Average number of new non-programming skills added per member who added at least one skill	∅	∅	∅	∅	∅	∅	∅	∅
Average number of new non-programming skills added per member who added at least one skill, filtered	∅	∅	--	--	--	--	∅	-
Average number of new programming skills added per member who added at least one skill	∅	∅	∅	∅	∅	∅	∅	∅
Average number of new programming skills added per member who added at least one skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Average number of new soft skills added by SWEs who added at least one skill, filtered	∅	∅	∅	∅	∅	-	∅	∅
Average number of new tech skills added by SWEs who added at least one skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Number of SWEs at firm who added at least one skill in the given month	∅	∅	∅	∅	∅	∅	∅	∅
Number of SWEs at firm who added at least one skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Number of SWEs who added at least one non-programming skill in the given month	∅	∅	∅	∅	∅	∅	∅	∅
Number of SWEs who added at least one non-programming skill in the given month, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Number of SWEs who added at least one programming skill in the given month	∅	∅	∅	∅	∅	∅	∅	∅
Number of SWEs who added at least one programming skill in the given month, filtered	∅	∅	∅	∅	∅	∅	∅	∅

Match requirement	Any	Any	Any	Strong	Strong	Strong	None	None
Minimum avg. SWE count	10	30	100	10	30	100	30	100
Share of SWEs who added at least one skill that added at least one business skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill that added at least one disruptive tech skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill that added at least one industry skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill that added at least one soft skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill that added at least one tech skill, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill who had at least one of those skills be non-programming skills, filtered	∅	∅	∅	∅	∅	∅	-	∅
Share of SWEs who added at least one skill who had at least one of those skills be programming skills	∅	∅	∅	∅	∅	∅	∅	∅
Share of SWEs who added at least one skill who had at least one of those skills be programming skills, filtered	∅	∅	∅	∅	∅	∅	∅	∅
Skills of new hires								
Average number of non-programming skills new hires have	++	++	++	++	++	++	++	++
Average number of programming skills new hires have	∅	++	∅	∅	∅	∅	++	∅
Demand								
Number of job posts in US for SWEs with no degree requirement	∅	∅	∅	∅	∅	∅	∅	∅
Number of job posts in US for SWEs with no degree requirement	∅	∅	∅	∅	∅	∅	∅	∅
Number of job posts in US with no degree requirement	∅	∅	∅	∅	∅	∅	∅	∅
Number of job posts in US with no degree requirement, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of job posts in US with no degree requirement, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅
Number of jobs posts for SWEs	∅	∅	∅	∅	∅	∅	∅	∅
Number of jobs posts for SWEs, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of jobs posts for SWEs, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅

Match requirement	Any	Any	Any	Strong	Strong	Strong	None	None
Minimum avg. SWE count	10	30	100	10	30	100	30	100
Share of firm's job posts for SWEs that do not require degree	∅	∅	∅	∅	∅	∅	∅	∅
Share of firm's job posts which are for SWEs	∅	∅	∅	∅	∅	∅	∅	∅
Supply								
Number of job applicants for SWE job postings from the firm	∅	∅	∅	∅	∅	∅	∅	∅
Number of job applicants for SWE job postings from the firm, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of job applicants for SWE job postings from the firm, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅
Number of job applicants for SWE job postings more senior than their current position	∅	∅	∅	∅	∅	∅	∅	∅
Number of job applicants for SWE job postings more senior than their current position, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of job applicants for SWE job postings more senior than their current position, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅
Number of people who viewed SWE job postings from the firm	∅	∅	∅	∅	∅	∅	∅	∅
Number of people who viewed SWE job postings from the firm, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of people who viewed SWE job postings from the firm, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅
Number of people who viewed SWE job postings that are more senior than their current role	∅	∅	∅	∅	∅	∅	∅	∅
Number of people who viewed SWE job postings that are more senior than their current role, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of people who viewed SWE job postings that are more senior than their current role, extensive margin	--	-	--	--	--	--	-	-
Share of current SWEs from the firm who applied for at least one SWE job	∅	∅	∅	∅	∅	∅	∅	∅
Share of current SWEs from the firm who applied for at least one SWE job at a higher level	∅	∅	∅	∅	∅	∅	∅	∅
Share of current SWEs from the firm who viewed at least one SWE job	∅	∅	∅	∅	∅	∅	∅	∅
Share of current SWEs from the firm who viewed at least one SWE job at a higher level	∅	∅	∅	∅	∅	∅	∅	∅
New hires								
Number of new SWE hires	∅	∅	∅	∅	∅	∅	∅	∅

Match requirement	Any	Any	Any	Strong	Strong	Strong	None	None
Minimum avg. SWE count	10	30	100	10	30	100	30	100
Number of new SWE hires who are entry-level	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are entry-level, Winsorized	++	++	++	++	++	++	++	++
Number of new SWE hires who are entry-level, extensive margin	++	++	++	++	++	++	++	++
Number of new SWE hires who are executive level	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are executive level, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are interns	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are interns, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are interns, extensive margin	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are senior level	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires who are senior level, Winsorized	∅	++	++	++	++	++	+	++
Number of new SWE hires who are senior level, extensive margin	++	++	∅	++	++	∅	++	∅
Number of new SWE hires, Winsorized	∅	∅	∅	∅	∅	∅	∅	∅
Number of new SWE hires, extensive margin	++	++	++	++	++	++	++	++
Share of new SWE hires who are executive level	∅	∅	∅	∅	∅	∅	∅	∅

Note: ∅: not statistically significant. ++ positive and statistically significant at 5% level.
+ positive and statistically significant at 10%. -- negative and statistically significant at 5% level.
- negatively and statistically significant at 10% level. Filtered: filtered to sample of members who had added at least one skill prior.