The Generative AI - Gender Puzzle

Manuel Hoffmann* Sam Boysel* Frank Nagle* Kevin Xu‖

*Harvard Business School, Harvard University

‖GitHub Inc.

This version: *December 9, 2024*

Abstract: Artificial Intelligence (AI) has the potential to reshape work processes, yet its implications for gender disparities remain unclear. This study investigates the adoption and impact of GitHub Copilot, a generative AI tool, through a regression discontinuity design. While men are more likely to adopt and use Copilot, the effect of generative AI on work patterns is much more pronounced for women. Compared with men, generative AI induces women to reallocate the focus of their tasks towards core work and away from project management to a greater degree. These findings reveal a "Generative AI – Gender Puzzle": women have the potential to benefit more substantially despite lower usage. This study highlights the need to address barriers to equitable AI adoption and informs policies to ensure inclusive benefits of AI technologies.

JEL-Classification: H4, O3, J0

Keywords: Generative Artificial Intelligence, Digital Gender Gaps, Open Source Software, Knowledge Economy

---

Acknowledgement: The authors are grateful for financial and administrative support from GitHub. The findings and conclusions presented in this study are solely those of the authors.

1 Motivation

Artificial Intelligence (AI) is predicted to substantially transform the economy and work processes. The rapid adoption of AI technologies is expected to affect productivity, skill requirements, and organizational structures across industries (Brynjolfsson, Rock, and Syverson, 2021). However, it is unclear how evenly distributed any potential returns to use of this new technology may be. On the one hand, generative AI reduces the cost of access to expertise, and it may therefore increase access to previously difficult-to-obtain just in-time assistance for disadvantaged groups which can be viewed as democratizing access to mentorship. Multiple studies have shown the potential for generative AI tools to narrow skill gaps by elevating the lesser skilled workers (see Brynjolfsson, Li, and Raymond 2023; Dell'Acqua et al. 2023; Hoffmann et al. 2024). Similarly, Agrawal, Gans, and Goldfarb (2019) suggest that AI technologies can provide scalable support and reduce access barriers to expertise. On the other hand, generative AI has been shown to produce biased output which – if used at scale – may result in replicating pre-existing gaps in the workplace or even an increase in gaps between disadvantaged groups and the majority. Studies such as the one from Bender et al. (2021) emphasize that large language models and other generative AI systems often encode and propagate systemic biases present in their training data, which could exacerbate existing inequities in the labor market and the workplace. As such, it is crucial to better understand the consequences of generative AI implementation on historically disadvantaged groups.

In science, technology, engineering, and mathematics (STEM) fields, women are typically underrepresented: while women have an even representation at around 49.3% of the employed population for non-STEM jobs, they make up only 29.2% of the workforce in STEM jobs globally (World Economic Forum, 2023). Gender disparities are even more stark within the open source software ecosystem, where only 5% of top projects have female core contributors (Trinkenreich et al., 2022). Even beyond the level of representation, a broad literature documents a plethora of gender gaps in the workplace. Women are less likely to be full-time employed (Goldin, 2014), earn lower wages (Blau and Kahn, 2017), and are underrepresented in positions of leadership (Kunze and Miller, 2017) and entrepreneurship (Guzman and Kacperczyk, 2019). The OECD,

has identified closing gender gaps in the workforce as key to sustainable growth and innovation (OECD, 2020, 2024b) and with AI as one of the technologies with the most transformation potential (OECD, 2024a), it is an important question whether the effect of AI will be to reduce or increase pre-existing societal gaps across gender.

Answering this question is crucial to better understand how policies could be established around work and generative AI to broadly distribute its benefits across a broad swath of our society. The most recent literature suggests that this technology follows a trend that has been observed for past technologies, where substantial gaps in adoption and usage have been observed across men and women (Acilar and Sæbø, 2023). As such, a meta study by Otis et al. (2024) shows that women are less likely than men to adopt and use generative AI. In another study, Humlum and Vestergaard (2024) document this pattern in the context of Denmark for ChatGPT where they link survey data on usage with administrative data. While these studies have made initial progress on the reasons why women have lower rates of adoption and usage, we particularly think that interlinking the adoption as well as usage with the consequences, and, as such, any anticipated benefits and costs, can substantially enhance our understanding about technology adoption and usage, more broadly construed.

Building on Hoffmann et al. (2024), we study the use of generative AI in the context of GitHub's Copilot – a software development focused generative AI tool based on OpenAI's Generative Pre-trained Transformer (GPT) models, which also underlie ChatGPT. Copilot was developed jointly by OpenAI, an AI research organization, and GitHub, a subsidiary of Microsoft that offers a decentralized software development platform which represents the world's largest host of open source software (OSS) projects. Copilot was trained on vast amounts of publicly available software code, including that hosted as OSS on GitHub itself, and assists software development with code completion and interactive chat features. Originally released under a limited access technical preview starting in June 2021, it became more broadly available as a full product in June 2022. As part of this roll out process, GitHub offered complimentary access to subsets of users, including developers of top OSS projects hosted on GitHub. The characteristics of this roll-out process assist

us in identifying the causal effects of access to Copilot. By exploiting the introduction of Copilot during the general access period through a dataset of top developers, we compare those that were just eligible to receive access to GitHub Copilot for free to those that are not using a regression discontinuity design. We further identified women and men using the full names of those developers and we verified that the share of women we identified is nearly identical to the share of women when developers select their pronouns using a new feature on GitHub (16%).

In the setting of the generative coding AI GitHub Copilot, we find at the threshold of AI eligibility that males are as likely to adopt generative AI than females but once they have adopted it, males are using generative AI with a higher intensity. As such, while women use generative AI slightly less than males, we can see a different pattern emerge on the consequences that generative AI has after receiving it quasi-exogenously. We find that women increase their share of core work (coding) by more than males and, similarly, they reduce project management by more than males. Aside from showing important heterogeneity of the findings from Hoffmann et al. (2024) which depict the average effect, this heterogeneity is important in conjunction with usage. While the usage and adoption differential are a puzzle in itself, taken together with the stronger behavioral response of women post usage is what we call the Generative AI – Gender Puzzle: why do women not use generative AI as much as men but when they do, they are the ones who respond most strongly when a generative AI tool reduces their cost of core work?

Our study makes three key contributions to the literature on artificial intelligence and its impact on work. First, we document a novel Generative AI – Gender Puzzle, in which men are more likely to adopt and use generative coding AI tools like GitHub Copilot, but women exhibit stronger responses when they adopt the technology. Specifically, we find that women experience a larger increase in the share of time spent on the core work of coding and a more pronounced decrease in the non-core work of project management tasks compared to men. As an addition of part of the puzzle we show that the share of women as top developers is around 15%. While low, this is encouraging given that the literature points to substantially lower estimates. Second, by leveraging quasi-exogenous variation in access to generative AI, we provide causal evidence on the

gender-differentiated effects of AI tools on task allocation, shedding light on how generative AI reshapes work patterns along gender lines. Finally, our findings suggest that the usage and impact of generative AI are not uniform across demographic groups, raising important questions about the factors driving adoption behavior and the downstream consequences of AI tools in workplace settings. These insights are crucial for policymakers and organizations aiming to promote equitable AI adoption and utilization while maximizing productivity gains.

The remainder of this paper proceeds as follows. Section 2 introduces the empirical setting which includes the background and the method. In Section 3 we characterize our dataset and discuss the construction of our sample. We hone into the set of developers that obtain Copilot eligibility for free via an internal ranking. We then present our results by directly starting with the estimation across gender using the regression discontinuity design (Section 4). Section 5 concludes.

2 Empirical Setting

We leverage a setting from GitHub, the largest open source software platform in the world.[1] GitHub decided to provide top developers of the open source software ecosystem with a free generative coding AI named Copilot. GitHub Copilot is a code completetion tool based on GPT-3 transformer series and was created jointly by Microsoft and Open AI. It was launched in the beta phase in the summer of 2021 and entered the general access phase in the summer of 2022. During that phase anyone was able to purchase a monthly ($10) or yearly subscription ($100 per year) to gain access to the coding generative AI. Aside from paid access, GitHub provided numerous other channels to obtain free access to the coding AI. Among them, they allowed top developers to obtain free access to the generative AI. They ranked top developer eligible projects (repositories) and they then created a threshold to decide who is eligible for GitHub Copilot. Individuals below the threshold could obtain free access through the top developer channel and the ones above

[1]GitHub Inc. is owned by Microsoft Corp. since 2018.

would not obtain access through this particular channel. In line with this empirical setting, we use a regression discontinuity design and employ the following model:

$$Yit = \alpha_0 + \alpha_1 \mathbb{1} Eligible_{it} + \alpha_2 Ranking_{it} + \alpha_3 Eligible_{it} \times Ranking_{it} + \epsilon_{it}$$

where Yit is the outcome including GitHub Copilot adoption, usage, coding and project management. $Ranking_{it}$ is the cumulative minimum Ranking across each eligible repository of a developer. $Eligible_{it}$ is defined as $\mathbb{1}(Ranking < 0)$ which is a parameter of interest from which we identify a change of the outcome at the normalized ranking threshold while considering a bandwidth of $h \in [-100, 100]$ (see Hoffmann et al. (2024) for further details, including results for the average developer at the threshold and robustness checks).

3 Data

For the time period from July 2022 to July 2023, we use granular data on work processes, a proprietary ranking from GitHub and their threshold information, as well as name-based identification to better understand differences across gender. We create a measure of coding based on the coding tab on the project website of GitHub - which includes activities such as creating a project (repository) and uploading (pushing) code to the platform. Similarly, we use the issues tab on a project website of GitHub to define project management. On the issues tab developers coordinate with each other and the community about bug fixes and feature request and, crucially, assign each other to engage in work on a project (see Hoffmann et al. (2024) for further details).

For this project, it is crucial to identify gender. This is a daunting task. While a feature has been introduced during our intervention period in March 2023 by GitHub for developers to report their preferred pronoun (e.g. he/him/his or she/her/hers), its coverage is relatively low. As such, we rely on predicting gender based on first and last name and conduct a validity check using the pronouns.

— Table 1 about here —

Table 1 provides descriptive statistics for the sample of top developers within our bandwidth who provided a first and last name. Based on our name-based prediction measure of gender, We find that sixteen percent of those developers are women.[2] This is consistent with the sample top developers that report pronouns consistent with identifying as female (16.6%).

The Table further shows descriptive statistics across the outcome variables of interest from coding, project management as well as adoption and usage. Across gender those measures appear to be relatively similar. The only pronounced difference appears in the cumulative Copilot days. Purely for descriptive purposes, we display skill-based measures that have been used in Hoffmann et al. (2024). The number of days (tenure) that a developer is on the platform is higher for males than females and consistent with this idea, male developers have a higher number of achievements on the platforms (batches provided by GitHub). Interstingly, women seem to have more followers than men, but they are on average not as central to projects when measured by core work as coding contributions to projects.

4 Results

We inspect heterogeneous determinants and consequences of generative AI across gender.

— Figure 1 about here —

Figure 1 shows treatment effects at the threshold for males and females and the difference of the two groups for a range of outcomes from weekly and cumulative days of Copilot usage, the intensity of usage – the number of Copilot active days over the number of GitHub active days – as well as the extensive margin of adoption in terms of ever usage. It becomes evident across all usage and adoption measures that males (light green bar) are more active than females (dark green bar) which results in a negative female-male difference in usage and adoption (blue bar). While the

[2]This share is not changing for top developers independent on the bandwidth.

difference for men relative to women goes into the same direction, it is not statistically significant for the adoption measure of ever use of Copilot throughout our sample period. We conclude that women do not differently adopt GitHub Copilot from men but once they adopt, they do not use it as much as men. The literature documents substantially lower rates of AI usage & adoption for females relative to males (Otis et al. 2024; Humlum and Vestergaard 2024) and as such our findings are only partially consistent with the prior literature. While providing the tool for free has a similar effect across gender through a change in adoption there may be some inherent differences in the way the tool is used during work across gender which may explain the differential intensity of usage.

— Figure 2 about here —

Figure 2 shows treatment effects at the threshold for males and females and the difference of the two groups on the core work of coding and the non-core work of project management.[3] We find that women (dark green bar) respond substantially stronger than males (light green bars) when they have access to Copilot and that they exhibit a larger relative increase in the share of coding and a substantially larger drop in the share of project management (blue bars). While the usage and adoption differential are a puzzle in itself, taken together with the stronger behavioral response of women post adoption and usage creates the Generative AI – Gender Puzzle: why do women not use generative AI as much as men but when they do, they are the ones for a responds most strongly when a generative AI tool reduces their cost of core work?

5 Conclusion

We document gender differences of generative AI determinants and its consequences through a natural experiment that was established by GitHub where top developers received access to a free coding AI called Copilot at a threshold and others did not receive access through this channel. By

[3]We define coding and project management for each project and developer on the GitHub platform as activities visible on the coding and issues tab, respectively. We follow the definitions established in Hoffmann et al. (2024).

identifying gender by first and last name, we establish that the share of women among top developers is around 16% (and there are around 84% men). We then used the regression discontinuity design and for those developers, we find that women are as likely as men to adopt the coding AI but they are using it less frequently than men. In contrast, once they have generative AI available, women are increasing their core work of coding and reducing their project management share by more than men. We call this pattern the Generative AI - Gender Puzzle since one may conclude that women - while they are using the technology less - are actually the ones who are potentially benefiting more from it.

In this paper we not only establish a new Generative AI - Gender Puzzle but we also put the well known technology gap of usage across gender in a new light. The findings indicate that the usage patterns of generative AI differ significantly between genders. Potential reasons that ought to be explored is that women may intentionally use AI as a mentor when barriers present themselves but otherwise they may not lean on it. Alternatively, the tasks that women are engaging in may be more favorable for input from the generative AI or women may simply be making more effective use of generative AI than men.

References

Acilar, Ali and Øystein Sæbø. 2023. "Towards understanding the gender digital divide: A systematic literature review." *Global knowledge, memory and communication* 72 (3):233–249.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. *The economics of artificial intelligence: an agenda*. University of Chicago Press.

Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the dangers of stochastic parrots: Can language models be too big?" In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.

Blau, Francine D and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of economic literature* 55 (3):789–865.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond. 2023. "Generative AI at work." Tech. rep., National Bureau of Economic Research.

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

Dell'Acqua, Fabrizio, Edward McFowland, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. 2023. "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality." *Harvard Business School Technology & Operations Mgt. Unit Working Paper* 24 (013).

Goldin, Claudia. 2014. "A grand gender convergence: Its last chapter." *American economic review* 104 (4):1091–1119.

Guzman, Jorge and Aleksandra Olenka Kacperczyk. 2019. "Gender gap in entrepreneurship." *Research Policy* 48 (7):1666–1680.

Hoffmann, Manuel, Sam Boysel, Frank Nagle, Sida Peng, and Kevin Xu. 2024. "Generative AI and the Nature of Work." *Harvard Business School Strategy Unit Working Paper* (25-021):25–021.

Humlum, Anders and Emilie Vestergaard. 2024. "The Adoption of ChatGPT." *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2024-50).

Kunze, Astrid and Amalia R Miller. 2017. "Women helping women? Evidence from private sector data on workplace hierarchies." *Review of Economics and Statistics* 99 (5):769–775.

OECD. 2020. *Advancing Gender Balance in the Workforce: A Collective Responsibility*. OECD Publishing.

———. 2024a. "Anticipating and Managing Emerging Global Transformations." URL https://www.oecd.org/en/about/programmes/strategic-foresight/anticipating-and-managing-emerging-global-transformations.html. Accessed: 12/1/2024.

———. 2024b. "Women in Inclusive Entrepreneurship." URL https://www.oecd.org/en/topics/sub-issues/inclusive-entrepreneurship/women-in-inclusive-entrepreneurship.html. Accessed: 12/1/2024.

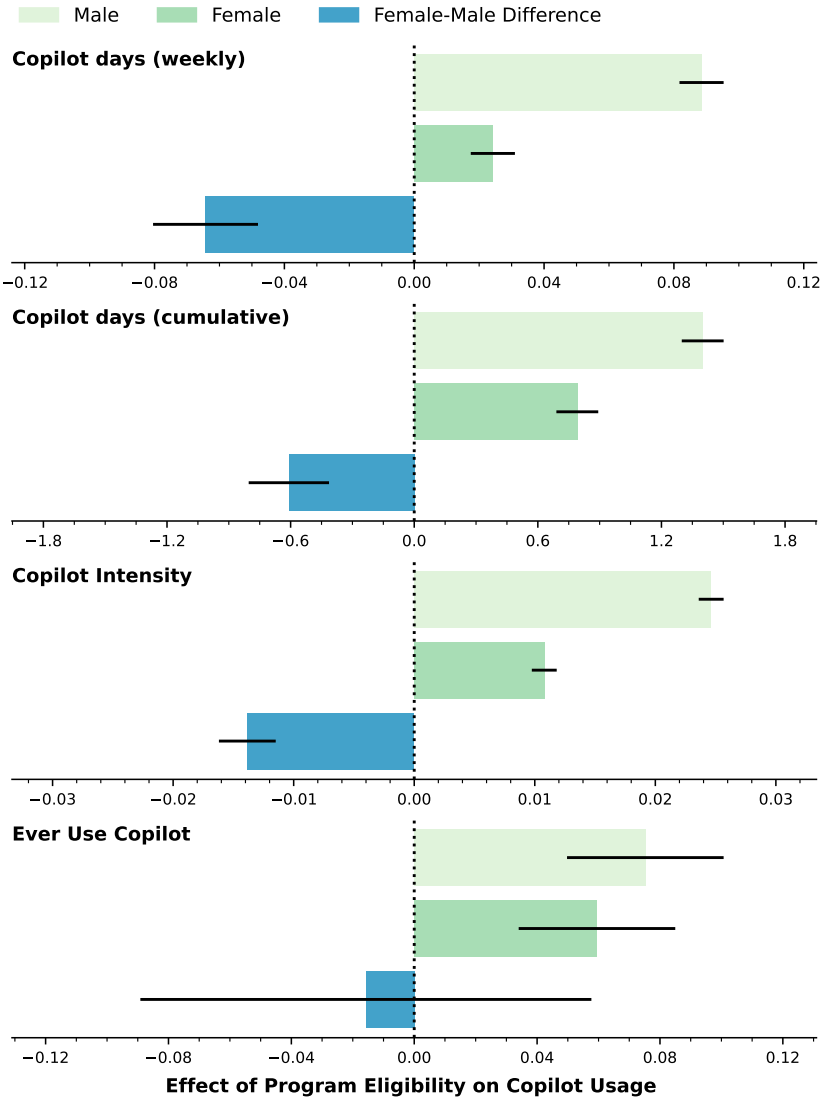
Otis, Nicholas G, Berkeley Haas, Katelyn Cranney, Solene Delecourt, and Rembrand Koning. 2024. "Global Evidence on Gender Gaps and Generative AI." Tech. rep., Center for Open Science.

Trinkenreich, Bianca, Igor Wiese, Anita Sarma, Marco Gerosa, and Igor Steinmacher. 2022. "Women's participation in open source software: A survey of the literature." *ACM Transactions on Software Engineering and Methodology (TOSEM)* 31 (4):1–37.

World Economic Forum. 2023. "Global Gender Gap Report." Tech. rep., World Economic Forum.

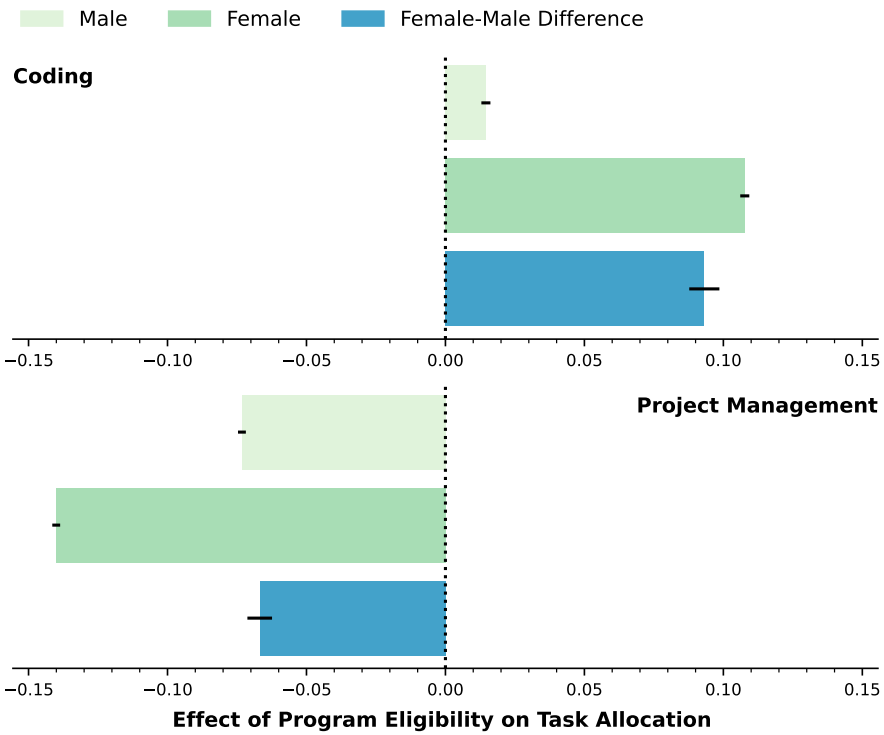
Figures and Tables

Figure 1: ADOPTION OF GENERATIVE AI ACROSS GENDER



Note: The figure shows intent-to-treat effects at the eligibility threshold across measures of GitHub Copilot generative AI usage and adoption.

Figure 2: CODING AND PROJECT MANAGEMENT GENDER-DIFFERENCES OF GENERATIVE AI



Note: The figure shows intent-to-treat treatment effects at the eligibility threshold across measures of core work of coding and non-core work of project management.

Table 1: Descriptive Statistics

	Pooled	**Male**	**Female**
Gender Measure			
Female	0.16	0	1
Outcome Measures			
Coding (Share)	0.45	0.45	0.46
Project Management (Share)	0.36	0.35	0.38
Copilot Days (Weekly)	0.11	0.11	0.10
Copilot Days (Cumulative)	4.57	4.67	3.99
Ever Copilot	0.12	0.12	0.11
Copilot Intensity	0.03	0.03	0.02
Skill Measures			
Tenure (days)	967	990	820
Number of Achievements	9.01	9.12	8.32
Number of Followers	96	96	100
Project Centrality	0.36	0.37	0.27
N	4,342	3,732	610

Note: The table shows mean statistics for a range of variables for the overall sample and split by males and females.