Does AI technology deployment benefit the owner of the technology? Impact of gitHub copilot release on microsoft

Vroegindeweij, Martijn; Medappa, Poonacha; Tunç, Murat

Published in:
20th Symposium on Statistical Challenges in Electronic Commerce Research

Publication date:
2024

Does AI Technology Deployment Benefit the Owner of the Technology ? Impact of GitHub Copilot Release on Microsoft

Martijn Vroegindeweij[1], Poonacha K. Medappa[1], Murat M. Tunc[1]
[1]Tilburg School of Economics and Management, Tilburg University

Software technical previews, sometimes termed beta releases, are frequently employed by firms prior to the public launch of their software applications. This practice allows for organizations to test and refine the software, gathering valuable feedback from users before the official release (Cheng and Liu, 2012). During the technical preview, the software gets stress tested in the actual use environment, and information on any software bugs is passed on to the organization that develops the software (Dey et al., 2013; Mehra and Saha, 2018). This feedback loop allows the firm to identify issues with the software for free in place of costly in-house testing.

The use of technical previews have become commonplace for major technology launches. When Google launched Gmail in March 2004 as a technical preview, its use extended beyond Google's internal team. The company granted access to 1,000 opinion leaders, broadening the scope of initial user feedback and engagement (Chung, 2004). Everyone in the technical preview (Google employees and the opinion leaders) were given the opportunity to invite their friends and family to participate in the technical preview. In a more recent development, Microsoft unveiled GitHub Copilot, an AI-powered coding assistant, through a technical preview in June 2021, followed by its broader public release in June 2022 (Dakhel et al., 2023). During this technical preview the product was made available to a select group of users who voluntarily sign up for the program (Scheffler, 2023). The selection criteria for the allocation of technical preview is made by the organization that launches the product, with organization employees often becoming the first users of the product during this phase (Scheffler, 2023).

For organizations launching the product, the overall impact of the new technology during the technical preview can be different from its public release. This variation necessitates careful consideration of how long and to whom the technical preview should be made available (Cheng and Liu, 2012). These considerations become even more pertinent when the organization is not only the creator but also a major user of the product, directly benefiting from its functionalities and improvements. An example for this is Copilot which is an AI tool that can assist software development and is capable of writing complete code blocks, explaining code, and generating test cases for code. Considering it is actively utilised by Microsoft and GitHub employees for software development (Scheffler, 2023), its launch as technical preview (when a larger proportion of Microsoft developers use the product) and public release (when the tool is publicly available to all users both inside and outside Microsoft) may deferentially impact Microsoft and others organizations who use the product.

For researchers interested in examining the impact of a new technology release, the choice of treatment period —be it the technical preview or public release phase— presents a significant statistical challenge as they may deferentially impact the treatment and/or the control groups. Hence, we pose the following question:

How does the impact of a technical preview (TP) of AI technology differ from its public release (PR) for the organization owner of that technology?

To answer this research question, we study the impact of the technical preview and public release of GitHub Copilot on the development outcomes of Microsoft's Open Source Software (OSS) projects. The prolonged technical preview period of GitHub Copilot launch (from June 2021 to June 2022) offers an excellent opportunity to study the impact of the new technology release. Moreover, Microsoft, which owns GitHub, has shifted most of its software development activities to the GitHub platform. This unique position allows for an analysis of the impact of AI-driven code development on Microsoft by observing coding behavior within its GitHub repositories. Since its public release, GitHub Copilot has seen considerable adoption, with over 10,000 organisations, including industry leaders such as Coca-Cola and Airbnb, embracing its use (Scheffler, 2023).

Current research on AI assisted code development has shown the substantial impact of the technology on productivity. A study to assess the productivity benefits of using Copilot revealed that developers who use Copilot completed their programming task 55.8% faster than the control group (Peng et al., 2023). Another study by GitHub revealed that Copilot use can also improve the code quality (Rodriguez, 2023). Along these lines, we investigate the impact of Copilot technical preview and public release on Microsoft owned OSS projects, and if it provided any comparative benefits for the company.

Research Design: Our paper uses OSS project-level data collected from GitHub. Since we are interested in understanding the impact of the technical preview and the public release period for the organization that owns the technology released, we use projects owned by Microsoft in the treatment group (5,098 projects). As the control group we selected the OSS projects owned by the top three US technology companies that are active on OSS - Google (1,796 projects), IBM (2,358 projects) and Meta (118 projects; Table 1). Issue, commit, and release data was gathered from all projects of the four organizations that were selected for the study. Using this information, we identify two main categories of project -level outcomes: (a) development measures : commits, code additions, and file changes ; (b) maintenance measures : issues and issue resolution time (See Table 2 for the descriptions of the variables). The measures were aggregated to create a monthly-panel data set, spanning June 2020 to June 2023. Only the projects that had activity in our time range were included in our analysis.

Figure 1 graphically represents our research design and our two treatments (technical preview and public release). The main DiD model is provided below:

$$y_{i,t} = \beta_0 + \beta_1 TP/PR_{i,t} + \gamma_i + \delta_t + \epsilon_{i,t}$$

where $y_{i,t}$ refers to the outcome measures (development and maintenance) for project i in month t. $TP/PR_{i,t}$ is the IV and a binary indicator that turns to 1 when the period is either technical preview or public release depending on our treatment. The project-level fixed effects are represented by γ_i, and δ_t represents the monthly fixed effects. $\epsilon_{i,t}$ is the robust standard error clustered at the project level to account for the potential heteroskedasticity of the errors.

Results and Discussion: Tables 5 and 6 provide the results of our models estimating project development and maintenance outcomes respectively. From Table 5, we find that the introduction of Copilot as a technical preview resulted in an increase in development activity for Microsoft.This is expected as the technical preview would see a greater proportion of AI-assisted code development from within Microsoft than outside. However, Microsoft tends to loose out on this development advantage after the public release when we expect other organizations to start adopting Copilot. When it comes to the project maintenance measures(Table 6), we find that while the technical preview does not lead to any changes to the number of issues and the issue resolution time, post public release sees a significant increase in both the maintenance measures. This could imply that while AI-assisted code development boosts productivity, it might also result in heightened maintenance needs over time. Specifically, there tends to be an increase in the number of issues reported for the code developed, and resolving these issues may require more time and effort.

Statistical Challenges: This research highlights two key statistical challenges. First, the choice of treatment -technical preview or public release- of a new technology is important and can deferentially impact the treatment and/or control groups. Partialling out these effects (specifically the effect of public release) is a challenge that needs careful consideration for research that looks at studying new technology release. Second, the choice of the organizations for the control group (Google, IBM and Meta) offers both a challenge and an opportunity. The challenge comes from the fact that different organizations may have different requirements, use cases, and tool to handle AI integration into their development operations. The opportunity comes from the potential ability to create a synthetic control (synthetic Microsoft) from a weighted combination of these three and potentially other organizations (Abadie et al., 2010).

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Cheng, H. K., & Liu, Y. (2012). Optimal Software Free Trial Strategy: The Impact of Network Externalities and Consumer Uncertainty. *Information Systems Research*, *23*(2), 488–504.

Chung, J. (2004, July). For Some Beta Testers, It's About Buzz, Not Bugs. https://www.nytimes.com/2004/07/22/technology/for-some-beta-testers-it-s-about-buzz-not-bugs.html

Dakhel, A. M., Majdinasab, V., Nikanjam, A., Khomh, F., Desmarais, M. C., & Jiang, Z. M. (2023). GitHub Copilot AI Pair Programmer: Asset or Liability? *The Journal of Systems & Software*.

Dey, D., Lahiri, A., & Liu, D. (2013). Consumer Learning and Time-Locked Trials of Software Products. *Journal of Management Information Systems*, *30*(2), 239–267.

Mehra, A., & Saha, R. L. (2018). Utilizing Public Betas and Free Trials to Launch a Software Product. *Production and Operations Management*, *27*(11), 2025–2037.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The Impact of AI on Developer Productivity: Evidence from Github Copilot. *arXiv preprint arXiv:2302.06590*.

Rodriguez, M. (2023, October). Research: Quantifying GitHub Copilot's impact on code quality - The GitHub Blog. https://github.blog/2023-10-10-research-quantifying-github-copilots-impact-on-code-quality/

Scheffler, I. (2023). GitHub CEO says Copilot will write 80% of code "sooner than later". https://www.freethink.com/robots-ai/github-copilot

A Figures

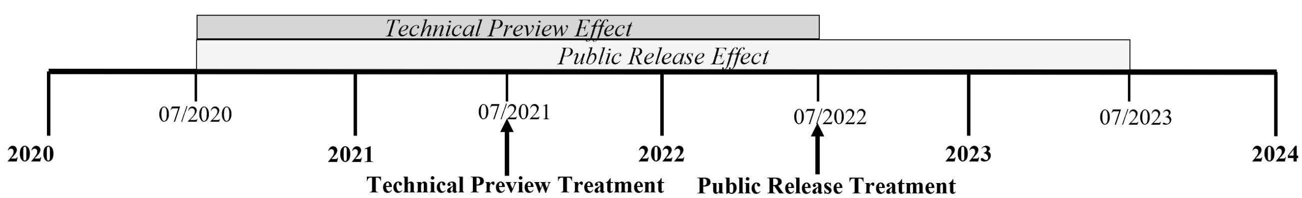

Figure 1: Technical Preview and Public Release of Copilot

B Tables

	Company	**Repositories**
Treatment	Microsoft	5,098
Control	Google	1,796
	IBM	2,358
	Meta	118
Total		**9,763**

Table 1: Summary of unique repositories for treated and control groups

Concept	Measure	Description
Dependent Variables		
	Additions	Total number of lines added to the project.
Development	ChangedFiles	Total number of files modified in the project.
	Commits	Total number of commits to the project repository.
Maintenance	Issues	Total number of issues reported.
	avgIssueResolutionTime	Average time taken to resolve an issue.
Independent Variables		
Treatment 1	Technical Preview of Copilot	1 year period of the technical preview (07/2021 - 07/2022)
Treatment 2	Public Release of Copilot	1 year period of the public release (07/2022 - 07/2023)

Table 2: Concepts and their measures

Variable	**Mean**	**Std. dev.**	**Min**	**Max**
Additions	4,210.413	128,839.4	0	3.90e+07
ChangesFiles	53.6931	934.7406	0	209,829
Commits	5.0480	37.14441	0	3,211
Issues	1.1216	21.78767	0	2,358
avgIssueResolutionTime	1,637.992	2,631.143	.0002778	26,250.08

Table 3: Descriptive statistics

	1	2	3	4	5
1. Additions	1.0000				
2. ChangedFiles	0.4431	1.0000			
3. Commits	0.1658	0.4626	1.0000		
4. Issues	0.0576	0.2177	0.3738	1.0000	
5. avgIssueResolutionTime	-0.0049	0.0017	0.0075	-0.0086	1.0000

Table 4: Pairwise Correlation

Table 5: The impact of technical preview and public release on project development measures.

DV:	Commits	Additions	FileChanges	Commits	Additions	FileChanges
CoPilot (TP)	0.037***	0.123***	0.047***	0.0363***	0.122***	0.0485***
	(0.011)	(0.031)	(0.017)	(0.011)	(0.031)	(0.017)
CoPilot (PR)				0.010	0.069**	0.002
				(0.012)	(0.030)	(0.018)
Project FE	✓	✓	✓	✓	✓	✓
Month FE	✓	✓	✓	✓	✓	✓
N	234,250	234,250	234,250	337,320	337,320	337,320
Adj. R^2	0.621	0.522	0.585	0.561	0.452	0.523

Note: All DVs are log-transformed. Robust standard errors clustered at the project level are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table 6: The impact of technical preview and public release on project maintenance measures.

DV:	Issues	Issue Resolution Time	Issues	Issue Resolution Time
CoPilot (TP)	0.007	0.002	0.005	0.003
	(0.004)	(0.015)	(0.004)	(0.016)
CoPilot (PR)			0.031***	0.033**
			(0.004)	(0.016)
Control for Issues		✓		✓
Project FE	✓	✓	✓	✓
Month FE	✓	✓	✓	✓
N	234,250	234,250	337,320	337,320
Adj. R^2	0.484	0.779	0.456	0.738

Note: All DVs are log-transformed. Robust standard errors clustered at the project level are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01