The Heterogeneous Productivity Effects of Generative AI: Evidence from Italy's ChatGPT Ban*

David Kreitmeir†

Paul A. Raschky‡

March 1, 2024

Abstract

We analyse the individual productivity effects of Italy's ban on ChatGPT, a generative pretrained transformer chatbot. We compile data on the daily coding output quantity and quality of over 36,000 GitHub users in Italy and other European countries and combine these data with the sudden announcement of the ban in a difference-in-differences framework. Among the affected users in Italy, we find a short-term increase in output quantity and quality for less experienced users and a decrease in productivity on more routine tasks for experienced users.

*We would like to thank Yashdeep Dahiya for outstanding research assistance and Samantha Eyler-Driscoll for manuscript editing and proofreading. This paper supersedes a previous version titled "The Unintended Consequences of Censoring Digital Technology – Evidence from Italy's ChatGPT Ban", April 2023.

†Department of Economics and SoDa Labs, Monash University (davidkreitmeir.github.io and david.kreitmeir1@monash.edu)

‡Department of Economics and SoDa Labs, Monash University (praschky.github.io and paul.raschky@monash.edu)

1 Introduction

The public release of OpenAI's ChatGPT provided near universal[1] access to generative artificial intelligence (AI) tools at no or very low cost. Its subsequent quick adoption[2] broadened the discussion about the impact of generative AI on society and its potential to boost worker productivity by performing relatively complex tasks and producing (seemingly) novel output, all while requiring only minimal technological knowledge on the part of users. However, ChatGPT also has the tendency to produce wrong or faulty outputs (e.g., "hallucinations") that, in the absence of expert knowledge, are difficult to detect and costly to rectify and might ultimately undermine the productivity of some workers (Dell'Acqua et al., 2023).

One of the focal points in the discussion on generative AI's societal impact is its ability to create and generate new content and knowledge. Similarly to prior advances in AI, generative AI can enhance productivity by replacing more routine tasks (Brynjolfsson et al., 2023; Kanazawa et al., 2022; Noy and Zhang, 2023; Peng et al., 2023) or improving users' decision accuracy (Kleinberg et al., 2017; Almog et al., 2024; Cho, 2023). The feature that sets this generation of AI apart from previous ones is that, with its access to the universe of online knowledge, generative AI combines domain-specific information with rules, lending it the ability to create new content and ultimately opening the possibility of extending the production possibility frontier beyond an individual's current level of training or expertise.

However, the accuracy of current generative AI models' performance in some tasks such as text summarization or generation, combined with its clarity and confidence of delivery, might create the *illusion* that it enhances productivity in other domains. Inaccurate, faulty or "hallucinated" output may not be immediately detected and could be used as an input in a knowledge worker's subsequent production flow. For instance, Kabir et al. (2024) analysed ChatGPT's response to 517 programming questions and

[1]Countries where ChatGPT is not accessible include China, Eritrea, Iran, North Korea, Russia and Saudi Arabia, among others.

[2]According to OpenAI, by November 2023, ChatGPT was recording approximately 100 million weekly users. https://techcrunch.com/2023/11/06/openais-chatgpt-now-has-100-million-weekly-active-users/

found that 52% of its answers were incorrect and that the users presented with these answers overlooked these errors 39% of the time. Nevertheless, users still tend to prefer to use ChatGPT because of its comprehensive responses (Kabir et al., 2024) and the confident language of the responses (Li et al., 2023).

For some tasks (e.g., content writing), the process of detecting faulty output or rectifying generative AI–driven errors might be quick, while for others (e.g., software development), the same process can be tedious and time consuming.[3] There are also wide differences in the accuracy of generative AI output, driven not only by the complexity of the underlying task but also the size and quality of the underlying training data. The tools can leverage a very large online text corpus to predict the next word in tasks such as creative writing and chatting, but the training data for software development and code creation are limited to a relatively small number of online forums (e.g., Stack Overflow), where the ground truth can be noisy.[4]

In cases where the underlying task is more complex and the output requires accuracy to be ultimately useful (e.g. software development), relying on generative AI might prolong task completion and decrease workers' output quality (Dell'Acqua et al., 2023). Such problems might be more acute among less experienced workers, who may have less domain knowledge and require more time to detect and correct errors. Less experienced workers might also be more prone to continue using the tool because the alternatives (e.g., acquiring the knowledge and skills themselves) appear to be even costlier.

In this paper, we use observational data to analyse the heterogeneous effects of ChatGPT on the output quantity and quality of experienced and less experienced software

[3]While the motivation for the relatively early public release of ChatGPT and other large-language models (LLMs) was to improve their performance with the human-generated data collected from user interactions with the tools, their output quality remains noisy, and expert knowledge is often required to accurately judge this quality. Moreover, Chen et al. (2023) show that ChatGPT's performance on a number of tasks, including generating code, actually declined from version 3.5 to version 4.0, calling into question whether its performance and accuracy will continuously improve over time. In addition, del Rio-Chanona et al. (2023) show that the widespread use of ChatGPT has led to a decline in usage of online help forums such as Stack Overflow, which in return will decrease the human-generated ground truth data that can be used to improve AI models.

[4]For example, for less routine, more complex and more niche questions, the answers provided on Stack Overflow are not necessarily correct or are just initial solution suggestions instead of working solutions. While these suggestions might have received upvotes, signaling to the LLM their "usefulness" for its training, in reality, the content of the answers and ChatGPT subsequent output might be only an untested and ultimately not functioning code routines.

developers. In particular, we exploit Italy's sudden announcement of a ChatGPT ban as a natural experiment to examine the ban's short-run effects on GitHub users' productivity. We find that the ban had no systematic effect on the overall output of more experienced developers and only some small negative effects on their output for more routine tasks (resolving issues and debugging). However, among less experienced users, the short-term lack of access to ChatGPT increased both the amount of output and its quality. For this group of users, the likelihood that we observe any output-related activity on GitHub is approximately 10% higher for the two business days following the ban. This effect size shrinks for the subsequent days. In the same vein, we find some tentative evidence that Internet users in Italy adapted fairly quickly to the legislation by increasing their use of virtual private networks (VPNs) and encrypted routing to circumvent the ban. A placebo test shows that our results are unlikely to be driven by unobserved, seasonal factors. The results are also robust to our using alternative outcome variables and analysing the effect at the user–repository–day level.

Our results present some first nonexperimental empirical evidence on the effects of restricting access to generative AI on workers' performance in more complex tasks. Importantly, we show that the effects of generative AI are heterogeneous by worker skill type.

Our study complements the existing, largely experimental, literature on the effects of generative AI on worker productivity in less complex tasks (e.g., content writing, customer support), where generative AI output is less error-prone, by examining a setting with more complex tasks, where AI-generated output can be less accurate or more faulty. Existing work by Brynjolfsson et al. (2023) and Noy and Zhang (2023) has found mainly positive productivity effects of generative AI and stronger effects for less experienced workers in the contexts of customer support and content writing tasks. In contrast, our results suggest that, for more complex tasks (e.g., code development),[5] generative AI does not

[5]Related work by Peng et al. (2023) and Chatterjee et al. (2024) has found positive effects of GitHub Copilot on both the productivity and job satisfaction of software developers. While GitHub Copilot is also an AI-based tool developed by GitHub and OpenAI, it is specifically a code completion tool. In contrast, ChatGPT is a chat-based (rather than auto-complete) tool and can be used to produce entirely new code segments/programs based on a human language prompt; in such cases, the accuracy of generative AI is highly variable (Kabir et al., 2024).

necessarily boost the productivity of less skilled workers and can even decrease their output quantity and quality. Our results, therefore, confirm the findings of Dell'Acqua et al. (2023) in a controlled experiment environment that, for tasks beyond ChatGPT's current capabilities, using ChatGPT increases the time a worker spends on a task. We complement their results by highlighting the differential effects by knowledge worker skill level. Our finding of heterogeneous effects for more complex tasks also empirically complements the larger discussion in economics on technological change and inequality in the labour market (e.g., Acemoglu, 2002; Autor et al., 2003; Goldfarb and Tucker, 2019; Acemoglu and Restrepo, 2020), in particular the productivity and labour market effects of AI (e.g., Brynjolfsson et al., 2017; Agrawal et al., 2019; Acemoglu, 2021; Eloundou et al., 2023).

The paper is organised as follows: Section 2 provides some background on ChatGPT and the Italian ban on the technology in 2023. Section 3 describes the data. Section 4 presents empirical results on the ban's effect on worker productivity, and Section 5 concludes.

2 ChatGPT and the Italian Ban

ChatGPT, an LLM created by US startup OpenAI, has been used by millions of people since it launched in November 2022. Trained on a vast corpus of text data from the Internet as it was in 2021, this large-scale AI language model uses a transformer-based neural network to process natural language. During the training process, the model learned to identify patterns and relationships between words, phrases, and sentences, enabling it to generate text.[6]

ChatGPT is accessible via a public website (chatgpt.openai.com) or an application programming interface (API), and almost anyone[7] can sign up for a free account. The interface is designed like a chat environment where the user writes "prompts" and ChatGPT answers. Interactions can range from casual chats and search-like queries to more

[6]Such as this very paragraph.

[7]Before Italy, countries including China, Russia and North Korea had already banned ChatGPT.

complex exchanges such as creative writing of a text or creation of recipes based on prompts. ChatGPT can also write code in multiple programming languages on the basis of a simple prompt.

On April 1, 2023, the Italian data protection authority (Garante per la protezione dei dati personali) blocked use of the ChatGPT chatbot, citing privacy concerns, and announced an investigation into OpenAI's compliance with the European Union's General Data Protection Regulation (GDPR). In particular, the authority stated that there was no legal basis for the mass collection and storage of personal data to train the algorithms underlying the platform's operation.[8] The ban was lifted in late April after OpenAI responded to the data protection authority's privacy concerns.[9]

3 Data

GitHub Data We access individual-level, real-time activity data for GitHub users in Italy (treatment) and Austria, France, and Spain (control) in the week prior to and that immediately after the ChatGPT ban in Italy (March 27–April 11, 2023).

GitHub is the world's largest online code hosting platform, used for storage of and joint work on coding projects (so-called repositories).[10] All modifications to a GitHub repository are automatically timestamped and stored, and GitHub permits tracking of any iterations of specific files and lines of code. Every action taken by a team member is automatically recorded, with details about the kind and substance of the modification, the files and code lines affected, and the date the changes were performed. Anyone with access to a repository can examine and download the history of iterations and actions, and given GitHub's history of developing open-source software, a significant portion of its repositories are not access restricted, meaning that the project activity information is available to everyone. Thus, public GitHub repositories provide a direct,

[8] Shiona McCallum, "ChatGPT banned in Italy over privacy concerns", BBC 01/04/2023, https://www.bbc.com/news/technology-65139406

[9] Shiona McCallum, "ChatGPT accessible again in Italy", BBC 28/04/2023, https://www.bbc.com/news/technology-65431914

[10] The programming languages most commonly represented in GitHub repositories are `Python` (17.38%), `Java` (11.77%), `Go` (10%), `JavaScript` (9.95%), and `C++` (9.66%). In comparison, `R`-related repositories account for only 0.074% of all pull requests on GitHub. https://madnight.github.io/githut

real-time measure of labour activity for millions of software and code developers worldwide (McDermott and Hansen, 2021).[11]

We access data on individual GitHub users' activity from the GitHub archive, which is updated daily and contains all public event data. The GitHub archive data are hosted on Google's BigQuery warehouse system and can be accessed with a query on Google's cloud infrastructure. GitHub user information such as the year of GitHub user account creation was downloaded with the GitHub GraphQL API.[12] The two datasets are merged via the unique GitHub user login.

We use the individual-level action data to construct three sets of outcome variables: The first group is labelled *Output quantity* and includes aggregate *Output* limited to "productive" actions, aggregate *Output* as defined by Shen (2023), *Commits*, and *Issues closed*. The second group, *Output quality*, contains the *Bug fix ratio*, *Pull requests (PR) merged* and the *PR merge ratio*. The third group, *Task complexity*, consists of *Files edited per PR* and *Lines added per PR*. A detailed description of the construction and definition of each variable is provided in Appendix Table A.1.

On a daily basis, these actions are relatively rare events at the user level. Hence, we transform each into a binary indicator that takes 1 if one of the actions in a category is recorded for the user on a given day and takes 0 otherwise.[13] Descriptive statistics at the user–day and user level are presented in Appendix Table B.1.

Package Repositories We compile a list of packages hosted on GitHub for ten analytical programming languages: `C`, `C++`, `Go`, `Java`, `JavaScript`, `Julia`, `Perl`, `Python`, `R`, and `Rust`. We rely in the first instance on the community–curated "Awesome Lists" to locate GitHub repositories for "popular" packages in each language. In a second step,

[11]GitHub data have been used in empirical research on software developers' productivity during the onset of COVID-19 (Forsgren, 2021), the impact of COVID-19 on daily and weekly patterns of individual labour allocation (McDermott and Hansen, 2021), the effects of working from home on individual productivity (Shen, 2023), the effect of air pollution on individual output (Holub and Thies, 2023), and the relationship between social links and the likelihood of joining professional software development teams (Casalnuovo et al., 2015).

[12]The Python scripts written to access the GitHub user information for the respective countries are available at `https://GitHub.com/sodalabsio/GitHub_scrape`.

[13]The distribution of day–user-level counts of the main event variables during the sample period is presented in Figure B.1 in the appendix.

we scrape the information on all packages hosted on the official software repositories for `Python` (`pypi`), `R` (`CRAN`) and `Julia` (`JuliaRegistries`) to retrieve information on each package's GitHub repository. We make use of the standardized GitHub URL structure to identify the *owner* and *name* of a package repository.[14] To identify the GitHub user accounts other than the owner that contribute to a package repository, we use information on each individual GitHub user's activity from January 2011 until March 2023. We restrict the list of *GitHub event types* to "productive" events to select primarily accounts that made at least one substantial contribution to a package repository.[15] Moreover, we winsorise the sample at the 1st and 99th percentiles to safeguard against outliers.[16] Our final list of package contributors and owners comprises 483,855 unique GitHub user accounts, of which 5,920 are part of our baseline sample.

4 Effect of the ChatGPT Ban on GitHub Output

To analyse the effect of the Italian ChatGPT ban on GitHub users' output, we estimate variants of the following difference-in-difference (DID) event-study model:

$$Y_{it} = \alpha_i + \lambda_t + \gamma_{dow} + \sum_{\tau=-4}^{-2} \beta_\tau D_{it}^\tau + \sum_{\tau=0}^{3} \beta_\tau D_{it}^\tau + \epsilon_{it}, \tag{1}$$

where Y_{it} denotes the outcome variable, which, in the baseline analysis, is one of the nine output and task variables. D is a dummy variable equalling one for observations in the treatment group at event–day τ and zero otherwise, with $\tau = -1$ as the reference period; α_i is a vector of user-specific fixed effects; λ_t is day (date) fixed effects; γ_{dow} is day-of-the-week fixed effects; and ϵ_{it} is the error term. The parameters of interest are β_τ. We cluster standard errors at the user level.

[14] The stylized URL for a package repository is `https://GitHub.com/[account name hosting the repository]/[repository name]` (for instance, `https://github.com/numpy/numpy`).

[15] Our set of "productive" event types comprises `PullRequestEvent`, `PullRequestReviewEvent`, `PullRequestReviewCommentEvent`, `PushEvent`, and `ReleaseEvent`.

[16] Note that we exclude `bot` accounts from the list of contributors prior to winsorising.

4.1 Baseline Results

Table 1 presents the main results. The upper panel shows the estimated coefficients for D for each of the nine output variables, four different samples (overall, less experienced, experienced and package contributors) and a time period covering the last two working days (March 27–28, 2023) before and the first two working days after the ban took effect (April 3–4, 2023). We find post-ban increases for the Italian users on most of our output quantity, output quality and task complexity measures. For example, the aggregate output measure increases by 0.012. At a mean of 0.23, this corresponds to an increase in the average likelihood of an output-relevant GitHub event of approximately 5%.[17] These effects are, however, driven only by the less experienced users, whose aggregate output increases by 0.024, which translates to a 10% increase in the average likelihood of our observing any output action on GitHub. In contrast, the estimated coefficients for experienced users and package contributors are close to zero and not statistically significant, with the notable exception of the *Bug fix ratio* for package contributors. Here, we find a statistically significant, negative effect of the ban. Taken together, the results suggest that the ban actually increased the productivity of less experienced coders. After the ban took effect, both the amount and quality of their output increased, and they tended to edit more files and write more lines of code. This supports the idea that, prior to the ban, ChatGPT might have had a disruptive effect on the output quantity and quality of less experienced coders. Code produced by ChatGPT might not have compiled, might have required a lot of time to debug or might have distracted less experienced coders in other ways. On the other hand, the ban did not systematically affect experienced coders' output.

In the lower panel, we reestimate the specifications from the upper panel with a longer pre- and post-treatment period. Overall, the pattern of the results remains qualitatively stable. Quantitatively, the effect size for some of the output measures decreases, indicating that users might have found ways to adapt to the new situation or circumvent the ban. In addition, we now find a systematic decrease in experienced users' and package

[17] Refer to Table B.1 for summary statistics and Tables B.2 and B.3 for an extended version of Table 1 containing, among others, the dependent variable mean for each regression.

contributors' likelihood of closing issues, consistent with proficient users using ChatGPT mainly for debugging purposes. Figure 1 presents the estimated event-study coefficients β_τ from specification 1 for the outcomes of interest for each user type. Importantly, a joint F-test of whether all coefficients on the preevent relative time indicators are jointly zero cannot be rejected at conventional levels for any outcome, alleviating concerns about preexisting trends. Moreover, the event-study results reveal that the treatment effect for less experienced users did indeed peak two days after the ban.

4.2 Robustness Checks

We conduct a number of robustness checks and present the results in the appendix. First, we consider a number of alternative outcomes and show that our results are robust under alternative definitions of the output measure and hold for other types of GitHub actions, as well (Table B.4).[18] Second, the overall changes in output quantity and quality could be driven by unobserved factors occurring at the same time as the introduction of the ban.[19] In Section B.2, we present the results of a placebo test for a set of GitHub actions—creating a wiki page for a repository and making a repository public—for which ChatGPT is not used. Third, in Table B.5, we replace the binary outcome variables with continuous variables and show that our results are also robust on the intensive margin. Fourth, while we believe that the unexpected implementation of the ban in only one European country makes the assumption of parallel trends credible in our empirical setting, we nevertheless conduct a sensitivity analysis of possible post-treatment violations of parallel trends relative to pre-treatment violations. Appendix Figure B.2 presents relative magnitudes bounds, $\Delta^{RM}\left(\bar{M}\right)$, for the average treatment effect over the first two post-treatment days and main outcomes under different values of $\bar{M}$, as suggested by Rambachan and Roth (2023).[20] Finally, for a subset of users working on multiple repositories, we construct a new panel dataset at the user–repository–day level (Appendix

[18]Table B.4 also presents estimates on `Releases`, confirming the results presented in a previous version of the paper.

[19]For example, Easter Monday, an important public holiday in Italy, fell on April 10.

[20]For instance, $\bar{M} = 1$ restricts the post-treatment violations of parallel trends over the first two post-treatment days to be no larger than the maximal pre-treatment violation of parallel trends over two consecutive days.

Section C). This allows us to include repository (project) and user fixed effects. Despite a reduction in the number of users in both the control and treatment groups, the results largely confirm the patterns from our baseline, user-level, analysis. Interestingly, we also find a small negative effect of the ban on experienced users' output quantity relative to that of their European peers working on the same repository.

4.3 Discussion

Data from public GitHub repositories include code and software projects from a variety of organisations and individuals. Some of these are open-source development projects (e.g., APIs) from private-sector companies, some are general open-source projects developed by a community of volunteers (and therefore are closer in character to public goods), and others are owned by research organisations or individual developers. Given data limitations, it is not possible to distinguish the type of project.

It is possible that some Italian users immediately used tools (e.g., VPNs) to circumvent the ban. Using data on Google searches for VPN services and usage data for TOR[21] (see Appendix D), we show a sudden jump in circumvention activity among Italian Internet users in the days after the ban. Despite the easy access to circumvention technology, many corporations and organisations actually prohibit the use of VPN and TOR tools on their devices and networks, implying that their use may be limited to mainly private devices and home networks. More importantly, we still find systematic effects on output despite this circumvention activity, and one can interpret our results as a lower bound. Another concern is that our finding of heterogeneity between less experienced and experienced users could be driven by the latter's greater skill in circumventing the ban. However, the systematic effects of the ban on tasks related to fixing bugs and closing issues suggests that this is not the case.

There are a number of follow-up questions that we cannot empirical analyse because of limitations in the data. First, while generative AI might disrupt the production flow of

[21]The TOR (The Onion Router) network is an open-source overlay network of thousands of network relays that conceals a user's IP address. Unfortunately, we cannot access actual VPN usage data at daily level.

less experienced workers by providing faulty results, another possibility is that ChatGPT is simply a distraction. While it is not clear how this would explain the effect heterogeneity, more detailed data on the actual use of ChatGPT could help inform the design of workplace policies around generative AI.[22] Second, more detailed data would also shed light on the question of why, after the initial increase in output and quality, we observe a decrease in the effect size for less experienced users in subsequent days. One explanation, in line with the conclusions of Kabir et al. (2024) and Li et al. (2023), could be that less experienced users still prefer to use ChatGPT as a support tool because it generates accessible and easy-to-use responses and because the costs of pursuing alternative solutions (e.g., acquiring the necessary coding skills) is relative high. Finally, our study provides evidence on the productivity effects of (the ban on) generative AI in only the very short run because the ban was short-lived and circumventing it was relatively easy.

5 Conclusion

We present novel evidence of the short-term effects of generative AI (ChatGPT) on the productivity of knowledge workers using high-frequency, observational data from over 36,000 software developers in Italy and other European countries. We use the sudden ban on ChatGPT in Italy as a natural experiment and show that the access restriction distorted output quantity and quality. Our results not only present some first empirical evidence of the widespread adoption of ChatGPT in software and code development but also show that the productivity effects of ChatGPT (and restrictions on it) differ by skill level. Our findings have the following policy implications: For some, more complex tasks, generative AI can produce faulty and erroneous output that is difficult to detect, in particular for less experienced individuals. This calls for a more targeted use of the tool in both education and work. AI-based tools that harness the power of LLMs in a more controlled form, that generate a clearly defined output and that are not based on simple text prompts (e.g., GitHub Copilot) offer guard rails to ensure more domain-specific

[22]For example, existing generative AI tools such as GitHub Copilot are in general productivity enhancing because they are designed for specific tasks; Copilot, for example, only completes code and is not an open-ended chatbot.

use. Our findings also indicate that even well-intended government-mandated blocking of digital technology (to protect privacy) can lead to short-term output disruptions and costs for society. Sudden bans can be easily circumvented with VPN tools, but these adjustment activities simultaneously distort production processes and negatively impact productivity in professions that rely on the banned technology. Thus, our research also implies that policymakers should consider the potential economic cost of digital technology bans before imposing them.

References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1):7–72.

Acemoglu, D. (2021). Harms of AI. Working Paper 29247, National Bureau of Economic Research.

Acemoglu, D. and Restrepo, P. (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy*, 128(6):2188–2244.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2019). Artificial Intelligence: The ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, 33(2):31–50.

Almog, D., Gauriot, R., Page, L., and Martin, D. (2024). AI oversight and human mistakes: Evidence from Centre Court. arXiv 2401.16754.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent rechnological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4):1279–1333.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative AI at work. Working Paper 31161, National Bureau of Economic Research.

Brynjolfsson, E., Rock, D., and Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. Working Paper 24001, National Bureau of Economic Research.

Casalnuovo, C., Vasilescu, B., Devanbu, P., and Filkov, V. (2015). Developer onboarding in GitHub: The role of prior social links and language experience. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, ESEC/FSE 2015, page 817–828, New York, NY, USA. Association for Computing Machinery.

Chatterjee, S., Liu, C. L., Rowland, G., and Hogarth, T. (2024). The impact of AI tools on engineering at ANZ Bank: An empirical study on GitHub Copilot within corporate environment. arXiv 2402.05636.

Chen, L., Zaharia, M., and Zou, J. (2023). How is ChatGPT's behavior changing over time? arXiv 2307.09009.

Cho, S. (2023). The effect of robot assistance on skills. mimeo, UCLA Economics.

del Rio-Chanona, M., Laurentsyeva, N., and Wachs, J. (2023). Are Large Language Models a threat to digital public goods? Evidence from activity on Stack Overflow. arXiv 2307.07367.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., and Lakhani, K. R. (2023). The diffusion of disruptive technologies. Working Paper 24-013, Harvard University.

Eloundou, T., Manning, S., Mishkin, P., and Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of Large Language Models. 2303.10130, arXiv.

Forsgren, N. (2021). An analysis of developer productivity, work cadence, and collaboration in the early days of COVID-19. Technical report, Octoverse Spotlight.

Goldfarb, A. and Tucker, C. (2019). Digital Economics. *Journal of Economic Literature*, 57(1):3–43.

Holub, F. and Thies, B. (2023). Air quality, high-skilled worker productivity and adaptation: Evidence from GitHub. CRC TR 224 Discussion Paper Series crctr224_2023_402, University of Bonn and University of Mannheim, Germany.

Kabir, S., Udo-Imeh, D. N., Kou, B., and Zhang, T. (2024). Is Stack Overflow obsolete? An empirical Sstudy of the characteristics of ChatGPT answers to Stack Overflow questions. arXiv 2308.02312.

Kanazawa, K., Kawaguchi, D., Shigeoka, H., and Watanabe, Y. (2022). AI, skill, and productivity: The case of taxi drivers. Working Paper 30612, National Bureau of Economic Research.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237–293.

Li, J., Mynatt, E., Mishra, V., and Bell, J. (2023). "Always nice and confident, sometimes wrong": Developer's experiences engaging Generative AI Chatbots versus human-powered QA platforms. arXiv 2309.13684.

McDermott, G. R. and Hansen, B. (2021). Labor reallocation and remote work during COVID-19: Real-time evidence from GitHub. Working Paper 29598, National Bureau of Economic Research.

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

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv 2302.06590.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):2555–2591.

Shen, L. (2023). Does working from home work? A natural experiment from lockdowns. *European Economic Review*, 151:104323.

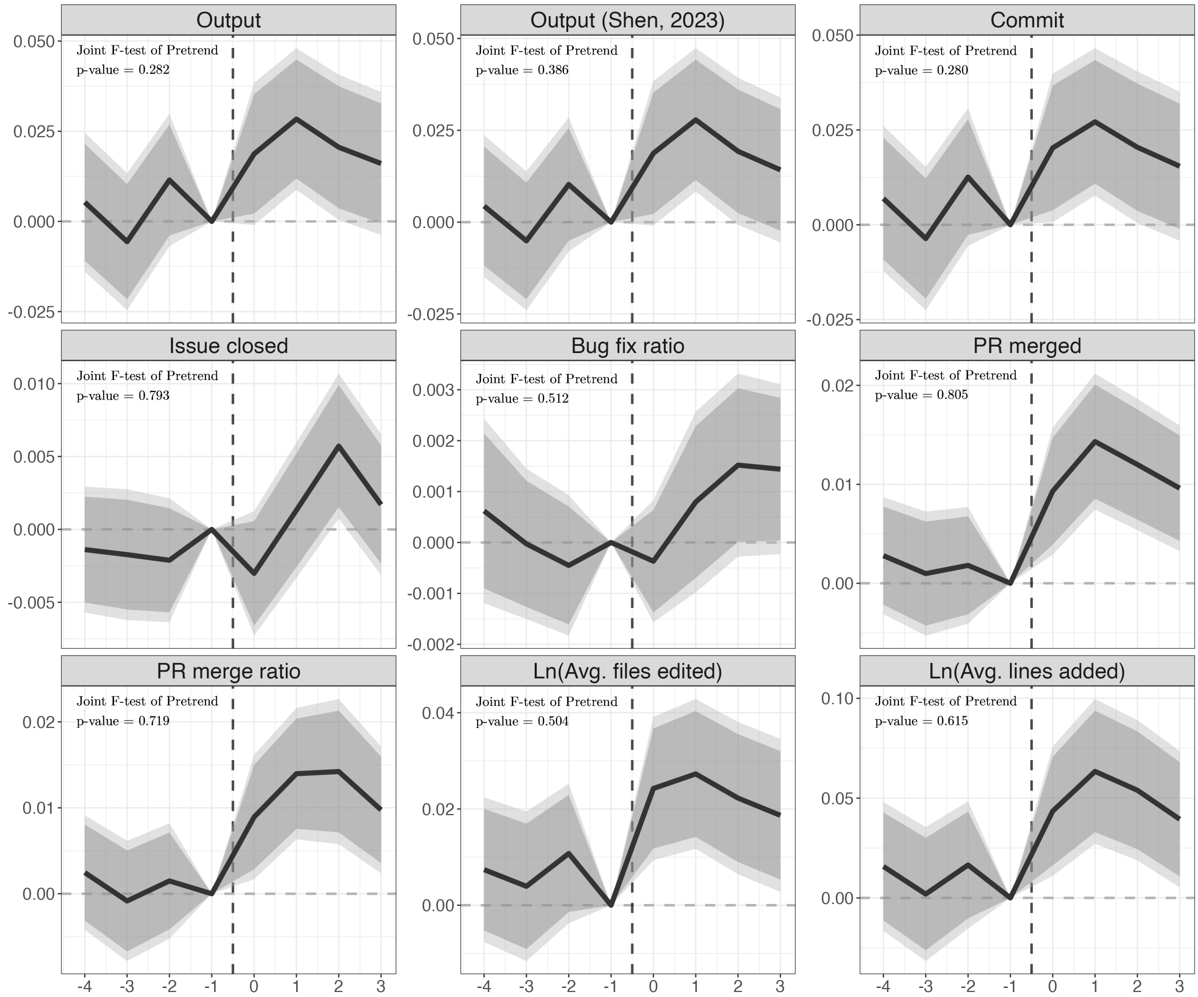
Table 1: Effect of ChatGPT Ban on GitHub Output

		Output Quantity				Output Quality			Task Complexity	
		Output	Output (Shen, 2023)	Commit	Issue closed	Bug fix ratio	PR merged	PR merge ratio	Files edited per PR	Lines added per PR
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre Mar 27–28 – Post Apr 3–4										
Overall (N = 145,496)	Treated × Post	0.0118**	0.0134**	0.0111**	-0.0004	0.0002	0.0064***	0.0069***	0.0108**	0.0238**
		(0.0055)	(0.0055)	(0.0055)	(0.0015)	(0.0006)	(0.0021)	(0.0023)	(0.0047)	(0.0104)
Less experienced (N = 74,864)	Treated × Post	0.0237***	0.0237***	0.0221***	0.0007	-0.0001	0.0099***	0.0106***	0.0201***	0.0446***
		(0.0078)	(0.0078)	(0.0078)	(0.0016)	(0.0006)	(0.0024)	(0.0025)	(0.0058)	(0.0126)
Experienced (N = 70,632)	Treated × Post	-0.0010	0.0025	-0.0006	-0.0017	0.0005	0.0025	0.0026	0.0004	0.0009
		(0.0078)	(0.0077)	(0.0076)	(0.0026)	(0.0012)	(0.0036)	(0.0041)	(0.0076)	(0.0170)
Pkg. contributor (N = 23,680)	Treated × Post	0.0147	0.0158	0.0174	-0.0076	-0.0068***	0.0006	0.0006	0.0004	-0.0130
		(0.0145)	(0.0144)	(0.0143)	(0.0054)	(0.0025)	(0.0074)	(0.0083)	(0.0153)	(0.0327)
Pre Mar 27–30 – Post Apr 3–6										
Overall (N = 290,992)	Treated × Post	0.0110***	0.0128***	0.0111***	-0.0005	0.0005	0.0059***	0.0065***	0.0103***	0.0251***
		(0.0042)	(0.0042)	(0.0042)	(0.0011)	(0.0004)	(0.0015)	(0.0017)	(0.0034)	(0.0075)
Less experienced (N = 149,728)	Treated × Post	0.0181***	0.0176***	0.0168***	0.0027**	0.0008*	0.0099***	0.0109***	0.0176***	0.0415***
		(0.0061)	(0.0061)	(0.0060)	(0.0013)	(0.0004)	(0.0018)	(0.0020)	(0.0044)	(0.0095)
Experienced (N = 141,264)	Treated × Post	0.0036	0.0079	0.0053	-0.0043**	0.0001	0.0013	0.0015	0.0022	0.0070
		(0.0059)	(0.0059)	(0.0058)	(0.0020)	(0.0008)	(0.0026)	(0.0029)	(0.0053)	(0.0118)
Pkg. contributor (N = 47,360)	Treated × Post	0.0147	0.0151	0.0188*	-0.0095**	-0.0036**	0.0002	0.0011	0.0008	-0.0030
		(0.0109)	(0.0108)	(0.0106)	(0.0040)	(0.0017)	(0.0050)	(0.0058)	(0.0106)	(0.0233)

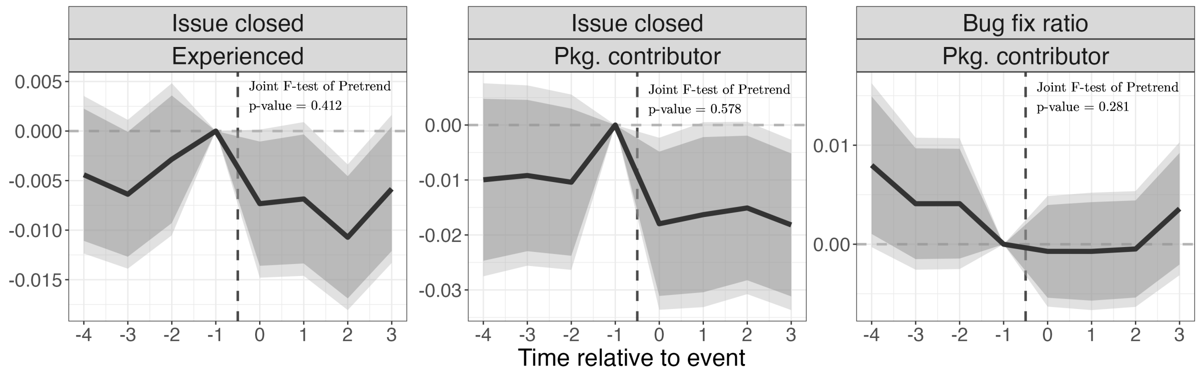
Notes: All specifications include user fixed effects. The “Less experienced” sample includes all GitHub user accounts created after or in 2017 (median), while the “Experienced” sample comprises all GitHub user accounts created before 2017. The “Pkg. contributor” sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is depicted in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Event-Study Estimates

(a) Effect of ChatGPT Ban on "Less experienced" User



(b) Effect of ChatGPT Ban on "Proficient" User



Notes: Panel A displays event-study estimates across outcomes for "less experienced" GitHub user accounts (created after or in 2017). Panel B presents event-study estimates for *(i)* "experienced" GitHub user accounts (created before 2017) and *(ii)* and accounts that are the owner and/or contributor to a(n) (analytical) programming package repository ("Pkg. contributor"). The sample period spans March 27–30 (*Pre*) and April 3–6 (*Post*). All specifications include user, time, and day-of-the-week fixed effects; 95% (90%) confidence intervals for robust standard errors clustered at the user level are depicted in light (dark) grey.

Supplementary Online Appendix

Table of Contents

A Data Appendix

Table A.1: Variable Definitions

Variable	Description
A – Output Quantity	
Events	Sum of all 15 GitHub event types, i.e., `CommitCommentEvent`, `CreateEvent`, `DeleteEvent`, `ForkEvent`, `GollumEvent`, `IssuesEvent`, `IssueCommentEvent`, `MemberEvent`, `PublicEvent`, `PullRequestEvent`, `PullRequestReviewEvent`, `PullRequestReviewCommentEvent`, `PushEvent`, `ReleaseEvent`, `WatchEvent`
Output	# Commits [`PushEvent$.size`] + # Issues closed [`IssuesEvent$.action == 'closed'`]+ # Pull requests closed [`PullRequestEvent$.action == 'closed'`] + # Releases [`ReleaseEvent`]
Output (Shen, 2023)	# Commits [`PushEvent$.size`] + # Pull requests [`PullRequestEvent`]
Output (Holub and Thies, 2023)	# Commits [`PushEvent$.size`] + # Comments on issues [`IssueCommentEvent`]+ # Comments on pull requests [`PullRequestReviewCommentEvent`] + # Comments on commits [`CommitCommentEvent`] + # Pull requests [`PullRequestEvent`] + # Issues [`IssuesEvent`]
Issue closed	# Issues closed [`IssuesEvent$.action == 'closed'`]
Commit	# Commits [`PushEvent$.size`]
Releases	# Releases [`ReleaseEvent`]
Pull Request (PR)	# Pull requests [`PullRequestEvent`]
B – Output Quality	
Bug fix ratio	(# Issues closed with "error" or "bug" label `[IssuesEvent$.action == 'closed' & REGEXP_CONTAINS(IssuesEvent$.label, 'error|bug'}]) / (# Issues closed [IssuesEvent$.action == 'closed'])`
PR merged	# Closed pull requests that were merged [`PullRequest$.pull_request.merged == 'true' & PullRequestEvent$.action == 'closed'`]

Table A.1: Variable Definitions *(continued)*

Variable	Description
PR merge ratio	(# Closed pull requests that were merged [`PullRequest$.pull_request.merged == 'true' & PullRequestEvent$.action == 'closed'`]) / (# Pull Requests closed [`PullRequestEvent$.action == 'closed'`])
C – Task Choice	
Files edited per PR	Average # files edited per closed pull request [`AVG(PullRequestEvent$.pull_request.changed_files) IF PullRequestEvent$.action == 'closed'`]
Lines added per PR	Average # lines added per closed pull request [`AVG(PullRequestEvent$.pull_request.additions) IF PullRequestEvent$.action == 'closed'`]
Interactive Activity	# Comments on issues [`IssueCommentEvent`]+ # Comments on pull requests [`PullRequestReviewCommentEvent`] + # Comments on commits [`CommitCommentEvent`]
Easy issue ratio	(# Easy issues opened + # Easy issues closed)/(# Issues opened + # Issues closed)
D – Placebo	
Wiki pages	# Wiki pages created [`GollumEvent`]
Public repos	# Repositories made public [`PublicEvent`]

Notes: The first column presents the variable name, and the second column provides a detailed description of how each variable is defined. The SQL Google BigQuery code to retrieve the required data is presented in brackets. The keywords to define an issue as "easy" are `good first issues, good first bug, good-first, documentation, polish, cleanup, simple, easy, small, trivial, minor help, wanted, junior job, newcomer, starter, beginner, newbie, novice, low hanging, low-hanging` (cf. Holub and Thies, 2023).

B GitHub User-Level Data

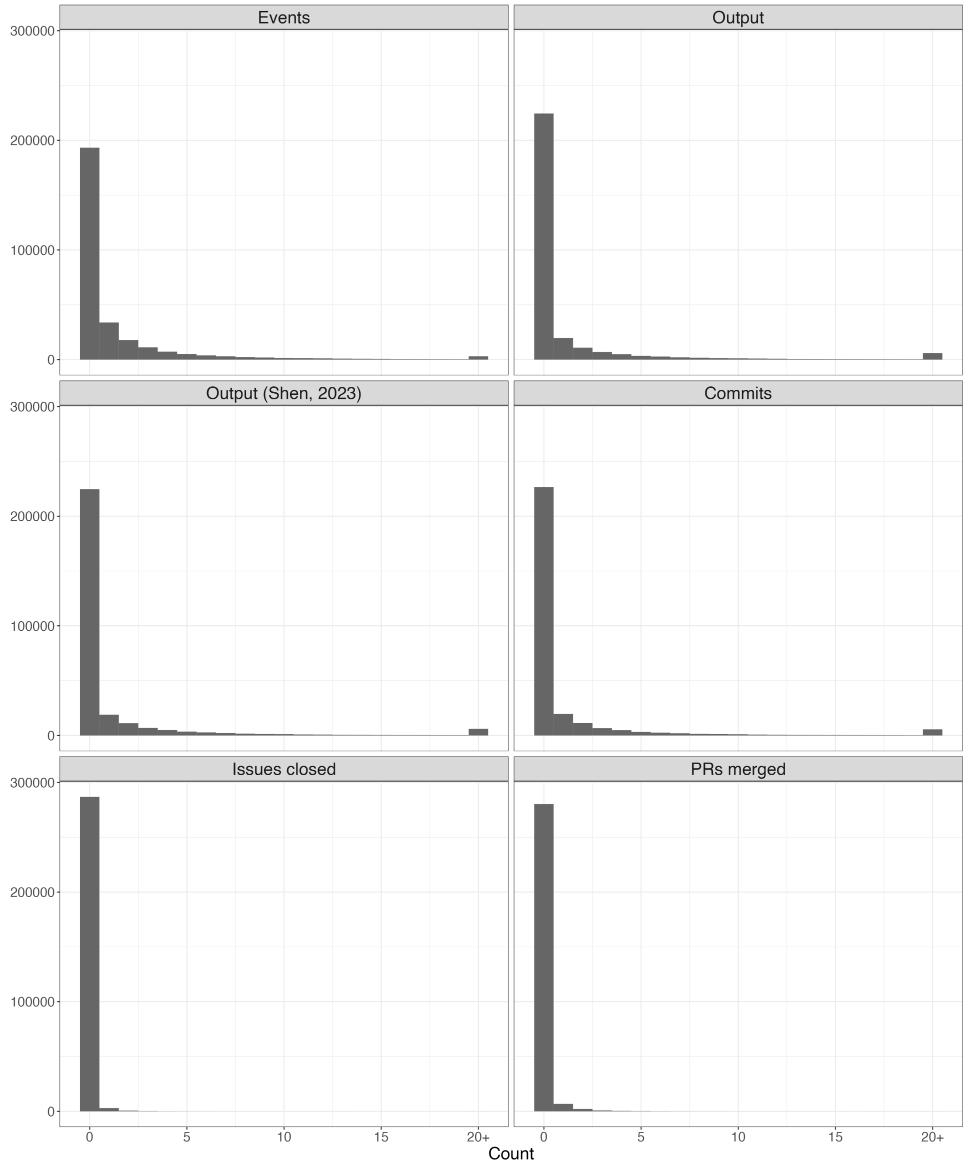
B.1 Descriptive Statistics

Table B.1: Descriptive Statistics

	Overall (36,374)		Less experienced (18,716)		Experienced (17,658)		Pkg. contributor (5,920)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A – User–Day Level (N = 290,992)								
Output	0.2286	0.4199	0.2319	0.4220	0.2251	0.4177	0.2644	0.4410
Output (Shen, 2023)	0.2283	0.4198	0.2316	0.4219	0.2249	0.4175	0.2645	0.4411
Commit	0.2215	0.4153	0.2273	0.4191	0.2155	0.4111	0.2522	0.4343
Issue closed	0.0142	0.1183	0.0089	0.0938	0.0198	0.1394	0.0267	0.1612
Bug fix ratio	0.0026	0.0484	0.0013	0.0350	0.0039	0.0595	0.0052	0.0689
PR merged	0.0371	0.1890	0.0265	0.1607	0.0483	0.2144	0.0635	0.2439
PR merge ratio	0.0396	0.2144	0.0281	0.1779	0.0518	0.2467	0.0684	0.2852
Files edited per PR	0.0698	0.4085	0.0549	0.3777	0.0857	0.4381	0.1087	0.4780
Lines added per PR	0.1593	0.9069	0.1259	0.8346	0.1947	0.9765	0.2480	1.0805
B – User Level (N = 36,374)								
	Mean	SD	Min	Median	Max			
User creation year	2016.55	3.77	2009	2017	2023			
Experienced	0.49	0.50	0	0	1			
Pkg. contributions	49.87	447.13	0	0	19638			
Pkg. owner	0.05	0.22	0	0	1			
Followers	29.49	203.18	0	6	17421			
Following	19.67	185.37	0	5	28300			
Repositories	29.91	54.54	0	17	3900			
Total events	11.93	18.87	1	5	140			

Notes: Panel A presents descriptive statistics for the baseline sample period March 27–30 (Pre) and April 3–6 (Post). The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of unique GitHub user accounts for the entire baseline sample ("Overall") and each of the sub-samples is presented in parentheses below. Panel B provides information on the individual characteristics of all GitHub user accounts in the baseline sample.

Figure B.1: Distribution of Output Quantities



Notes: Daily counts of each action type at the user level for the sample period of March 27–30 (Pre) – April 3–6 (Post) are presented. Counts above 20 are binned and labelled 20+.

B.2 Additional Results

Table B.2: Effect of ChatGPT Ban on GitHub Output – March 27–28 (Pre) – April 3–4 (Post)

		Output Quantity				Output Quality			Task Complexity	
		Output	Output (Shen, 2023)	Commit	Issue closed	Bug fix ratio	PR merged	PR merge ratio	Files edited per PR	Lines added per PR
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overall (N = 145,496)	Treated × Post	0.0118**	0.0134**	0.0111**	-0.0004	0.0002	0.0064***	0.0069***	0.0108**	0.0238**
		(0.0055)	(0.0055)	(0.0055)	(0.0015)	(0.0006)	(0.0021)	(0.0023)	(0.0047)	(0.0104)
	Post	-0.0101***	-0.0105***	-0.0093***	-0.0007	-0.0001	-0.0025**	-0.0030**	-0.0052**	-0.0098*
		(0.0024)	(0.0024)	(0.0024)	(0.0007)	(0.0003)	(0.0011)	(0.0012)	(0.0023)	(0.0052)
	Dep. var. mean	0.2325	0.2320	0.2251	0.0144	0.0026	0.0382	0.0406	0.0708	0.1621
Less experienced (N = 74,864)	Treated × Post	0.0237***	0.0237***	0.0221***	0.0007	-0.0001	0.0099***	0.0106***	0.0201***	0.0446***
		(0.0078)	(0.0078)	(0.0078)	(0.0016)	(0.0006)	(0.0024)	(0.0025)	(0.0058)	(0.0126)
	Post	-0.0185***	-0.0184***	-0.0181***	-0.0009	-0.0003	-0.0035***	-0.0037**	-0.0092***	-0.0195***
		(0.0035)	(0.0035)	(0.0035)	(0.0008)	(0.0003)	(0.0013)	(0.0014)	(0.0031)	(0.0068)
	Dep. var. mean	0.2355	0.2351	0.2305	0.0090	0.0015	0.0271	0.0286	0.0550	0.1264
Experienced (N = 70,632)	Treated × Post	-0.0010	0.0025	-0.0006	-0.0017	0.0005	0.0025	0.0026	0.0004	0.0009
		(0.0078)	(0.0077)	(0.0076)	(0.0026)	(0.0012)	(0.0036)	(0.0041)	(0.0076)	(0.0170)
	Post	-0.0015	-0.0022	-0.0002	-0.0005	0.0000	-0.0015	-0.0023	-0.0011	0.0002
		(0.0033)	(0.0033)	(0.0032)	(0.0011)	(0.0005)	(0.0017)	(0.0019)	(0.0035)	(0.0079)
	Dep. var. mean	0.2292	0.2288	0.2194	0.0201	0.0039	0.0499	0.0534	0.0876	0.1999
Pkg. contributor (N = 23,680)	Treated × Post	0.0147	0.0158	0.0174	-0.0076	-0.0068***	0.0006	0.0006	0.0004	-0.0130
		(0.0145)	(0.0144)	(0.0143)	(0.0054)	(0.0025)	(0.0074)	(0.0083)	(0.0153)	(0.0327)
	Post	-0.0101*	-0.0102*	-0.0097*	-0.0026	0.0006	0.0004	0.0007	0.0007	0.0060
		(0.0060)	(0.0060)	(0.0059)	(0.0023)	(0.0010)	(0.0032)	(0.0037)	(0.0063)	(0.0143)
	Dep. var. mean	0.2711	0.2710	0.2588	0.0269	0.0053	0.0677	0.0721	0.1131	0.2565

Notes: All specifications include user fixed effects. The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.3: Effect of ChatGPT Ban on GitHub Output — March 27–30 (Pre) – April 3–6 (Post)

		Output Quantity				Output Quality			Task Complexity	
		Output	Output (Shen, 2023)	Commit	Issue closed	Bug fix ratio	PR merged	PR merge ratio	Files edited per PR	Lines added per PR
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overall (N = 290,992)	Treated × Post	0.0110***	0.0128***	0.0111***	-0.0005	0.0005	0.0059***	0.0065***	0.0103***	0.0251***
		(0.0042)	(0.0042)	(0.0042)	(0.0011)	(0.0004)	(0.0015)	(0.0017)	(0.0034)	(0.0075)
	Post	-0.0149***	-0.0153***	-0.0140***	-0.0003	0.0000	-0.0030***	-0.0031***	-0.0056***	-0.0119***
		(0.0019)	(0.0019)	(0.0018)	(0.0005)	(0.0002)	(0.0008)	(0.0009)	(0.0017)	(0.0039)
	Dep. var. mean	0.2286	0.2283	0.2215	0.0142	0.0026	0.0371	0.0396	0.0698	0.1593
Less experienced (N = 149,728)	Treated × Post	0.0181***	0.0176***	0.0168***	0.0027**	0.0008*	0.0099***	0.0109***	0.0176***	0.0415***
		(0.0061)	(0.0061)	(0.0060)	(0.0013)	(0.0004)	(0.0018)	(0.0020)	(0.0044)	(0.0095)
	Post	-0.0218***	-0.0219***	-0.0212***	-0.0004	-0.0003*	-0.0042***	-0.0042***	-0.0086***	-0.0201***
		(0.0027)	(0.0027)	(0.0027)	(0.0005)	(0.0002)	(0.0010)	(0.0011)	(0.0023)	(0.0052)
	Dep. var. mean	0.2319	0.2316	0.2273	0.0089	0.0013	0.0265	0.0281	0.0549	0.1259
Experienced (N = 141,264)	Treated × Post	0.0036	0.0079	0.0053	-0.0043**	0.0001	0.0013	0.0015	0.0022	0.0070
		(0.0059)	(0.0059)	(0.0058)	(0.0020)	(0.0008)	(0.0026)	(0.0029)	(0.0053)	(0.0118)
	Post	-0.0078***	-0.0083***	-0.0066***	-0.0001	0.0003	-0.0018	-0.0021	-0.0024	-0.0035
		(0.0025)	(0.0025)	(0.0024)	(0.0008)	(0.0003)	(0.0012)	(0.0014)	(0.0026)	(0.0057)
	Dep. var. mean	0.2251	0.2249	0.2155	0.0198	0.0039	0.0483	0.0518	0.0857	0.1947
Pkg. contributor (N = 47,360)	Treated × Post	0.0147	0.0151	0.0188*	-0.0095**	-0.0036**	0.0002	0.0011	0.0008	-0.0030
		(0.0109)	(0.0108)	(0.0106)	(0.0040)	(0.0017)	(0.0050)	(0.0058)	(0.0106)	(0.0233)
	Post	-0.0152***	-0.0149***	-0.0145***	-0.0022	0.0004	-0.0023	-0.0027	-0.0040	-0.0059
		(0.0046)	(0.0046)	(0.0045)	(0.0016)	(0.0007)	(0.0024)	(0.0029)	(0.0047)	(0.0107)
	Dep. var. mean	0.2644	0.2645	0.2522	0.0267	0.0052	0.0635	0.0684	0.1087	0.2480

Notes: All specifications include user fixed effects. The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Alternative Outcomes

		Output Quantity				Task Choice		Placebo	
		Event	Output (Holub and Thies, 2023)	Pull request (PR)	Release	Interactive activity	Easy issue ratio	Wiki page	Public repo
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre Mar 27–28 – Post Apr 3–4									
Overall (N = 145,496)	Treated × Post	0.0224***	0.0189***	0.0066**	-0.0017	0.0023	0.0000	-0.0001	-0.0004
		(0.0065)	(0.0058)	(0.0028)	(0.0011)	(0.0029)	(0.0003)	(0.0004)	(0.0010)
Less experienced (N = 74,864)	Treated × Post	0.0313***	0.0257***	0.0119***	-0.0003	0.0025	-0.0002	-0.0006	0.0001
		(0.0091)	(0.0081)	(0.0032)	(0.0013)	(0.0032)	(0.0003)	(0.0006)	(0.0016)
Experienced (N = 70,632)	Treated × Post	0.0134	0.0119	0.0009	-0.0032*	0.0019	0.0002	0.0006	-0.0011
		(0.0094)	(0.0083)	(0.0046)	(0.0017)	(0.0051)	(0.0006)	(0.0006)	(0.0010)
Pkg. contributor (N = 23,680)	Treated × Post	0.0363**	0.0274*	-0.0003	-0.0037	-0.0025	-0.0018*	-0.0004	0.0005
		(0.0166)	(0.0154)	(0.0094)	(0.0033)	(0.0105)	(0.0009)	(0.0013)	(0.0020)
Pre Mar 27–30 – Post Apr 3–6									
Overall (N = 290,992)	Treated × Post	0.0190***	0.0147***	0.0061***	-0.0005	0.0011	0.0002	0.0001	0.0002
		(0.0050)	(0.0045)	(0.0020)	(0.0008)	(0.0022)	(0.0002)	(0.0003)	(0.0007)
Less experienced (N = 149,728)	Treated × Post	0.0244***	0.0186***	0.0107***	-0.0005	0.0016	0.0001	-0.0004	0.0004
		(0.0069)	(0.0062)	(0.0024)	(0.0011)	(0.0024)	(0.0002)	(0.0004)	(0.0011)
Experienced (N = 141,264)	Treated × Post	0.0136*	0.0109*	0.0011	-0.0006	0.0004	0.0003	0.0006	-0.0001
		(0.0072)	(0.0064)	(0.0033)	(0.0013)	(0.0039)	(0.0004)	(0.0005)	(0.0007)
Pkg. contributor (N = 47,360)	Treated × Post	0.0337***	0.0195*	-0.0034	0.0009	0.0005	-0.0009*	-0.0007	0.0001
		(0.0125)	(0.0115)	(0.0066)	(0.0024)	(0.0081)	(0.0005)	(0.0011)	(0.0011)

Notes: All specifications include user fixed effects. The “Less experienced” sample includes all GitHub user accounts created after or in 2017 (median), while the “Experienced” sample comprises all GitHub user accounts created before 2017. The “Pkg. contributor” sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

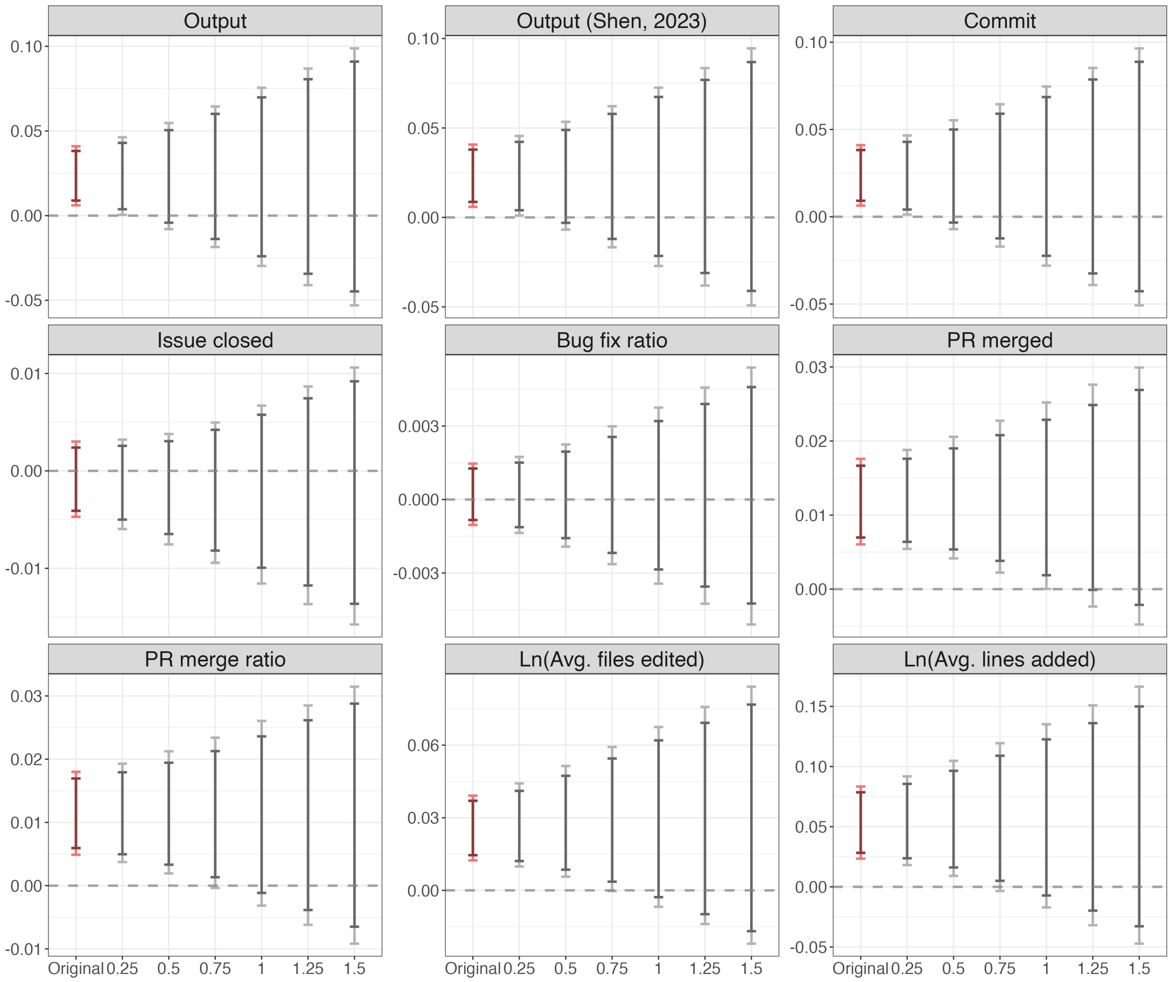
Table B.5: Event Counts

		Output Quantity						Output Quality
		Ln(1 +Events)	Ln(1 + Output)	Ln(1 + Output (Holub and Thies, 2023))	Ln(1 + Output (Shen, 2023))	Ln(1 + Commits)	Ln(1 + Issues closed)	Ln(1 + PR merged)
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pre Mar 27–28 – Post Apr 3–4								
Overall (N = 145,496)	Treated × Post	0.0247**	0.0178*	0.0203*	0.0244**	0.0166	0.0001	0.0069***
		(0.0097)	(0.0104)	(0.0104)	(0.0107)	(0.0102)	(0.0013)	(0.0021)
Less experienced (N = 74,864)	Treated × Post	0.0412***	0.0368***	0.0389***	0.0420***	0.0337**	0.0010	0.0105***
		(0.0134)	(0.0139)	(0.0140)	(0.0142)	(0.0137)	(0.0014)	(0.0024)
Experienced (N = 70,632)	Treated × Post	0.0070	-0.0033	-0.0001	0.0050	-0.0022	-0.0008	0.0027
		(0.0141)	(0.0156)	(0.0156)	(0.0161)	(0.0153)	(0.0023)	(0.0035)
Pkg. contributor (N = 23,680)	Treated × Post	0.0350	0.0138	0.0149	0.0247	0.0159	-0.0044	0.0037
		(0.0268)	(0.0313)	(0.0313)	(0.0322)	(0.0307)	(0.0045)	(0.0074)
Pre Mar 27–30 – Post Apr 3–6								
Overall (N = 290,992)	Treated × Post	0.0264***	0.0212***	0.0237***	0.0271***	0.0206***	-0.0001	0.0064***
		(0.0076)	(0.0080)	(0.0081)	(0.0083)	(0.0079)	(0.0010)	(0.0016)
Less experienced (N = 149,728)	Treated × Post	0.0366***	0.0345***	0.0353***	0.0386***	0.0316***	0.0024**	0.0102***
		(0.0105)	(0.0109)	(0.0110)	(0.0112)	(0.0108)	(0.0012)	(0.0018)
Experienced (N = 141,264)	Treated × Post	0.0157	0.0067	0.0110	0.0146	0.0087	-0.0031*	0.0021
		(0.0110)	(0.0118)	(0.0118)	(0.0123)	(0.0116)	(0.0018)	(0.0026)
Pkg. contributor (N = 47,360)	Treated × Post	0.0482**	0.0205	0.0214	0.0335	0.0225	-0.0053	0.0024
		(0.0207)	(0.0234)	(0.0234)	(0.0244)	(0.0230)	(0.0035)	(0.0049)

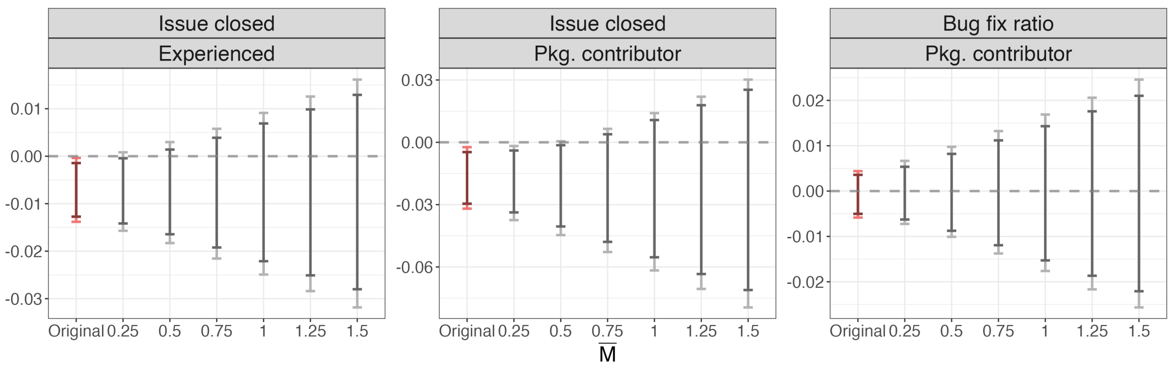
Notes: All specifications include user fixed effects. The “Less experienced” sample includes all GitHub user accounts created after or in 2017 (median), while the “Experienced” sample comprises all GitHub user accounts created before 2017. The “Pkg. contributor” sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure B.2: HonestDiD – Relative Magnitude Bounds

(a) Effect of ChatGPT Ban on "Less experienced" User



(b) Effect of ChatGPT Ban on "Proficient" User



Notes: This figure displays `HonestDiD` robust confidence interval sets of the average treatment effect across the first two post-treatment periods for different relative magnitude bounds $\Delta^{RM}\left(\bar{M}\right)$, as suggested by Rambachan and Roth (2023). Panel A displays confidence interval sets for "less experienced" GitHub user accounts and Panel B for *(i)* "experienced" accounts and *(ii)* accounts that are the owner and/or contributor to a programming package repository ("Pkg. contributor"); 95% (90%) robust confidence intervals are depicted in light (dark) grey for relative magnitude bounds and in light (dark) red for original treatment effect estimates.

C GitHub Repository–User-Level Data

C.1 Descriptive Statistics

Table C.1: Descriptive Statistics

	Overall (3,546 × 10,467)		Less experienced (2,572 × 4,003)		Experienced (3,244 × 6,464)		Pkg. contributor (2,223 × 2,548)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
A – User–Day Level (N = 186,496)								
Output	0.0122	0.1099	0.0072	0.0847	0.0152	0.1223	0.0189	0.1361
Output (Shen, 2023)	0.0152	0.1224	0.0101	0.1002	0.0182	0.1338	0.0237	0.1521
Commit	0.0095	0.0968	0.0053	0.0726	0.0119	0.1086	0.0144	0.1192
Issue closed	0.0035	0.0594	0.0021	0.0455	0.0044	0.0663	0.0052	0.0722
Bug fix ratio	0.0010	0.0305	0.0006	0.0250	0.0012	0.0333	0.0015	0.0378
PR merged	0.0048	0.0690	0.0023	0.0476	0.0063	0.0789	0.0078	0.0878
PR merge ratio	0.4620	6.7223	0.2235	4.7031	0.6038	7.6715	0.7471	8.5227
Files edited per PR	0.0095	0.1534	0.0055	0.1239	0.0119	0.1685	0.0149	0.1822
Lines added per PR	0.0211	0.3244	0.0124	0.2588	0.0263	0.3577	0.0331	0.3938

B – User Level (N = 10,467)	Mean	SD	Min	Median	Max
User creation year	2015.39	3.51	2009	2015	2023
Experienced	0.62	0.49	0	1	1
Pkg. contributions	90.21	672.04	0	0	19514
Pkg. owner	0.07	0.26	0	0	1
Followers	53.83	343.70	0	11	17421
Following	37.16	325.79	0	11	28300
Repositories	40.66	78.77	0	23	3900
Total events	3.73	10.59	0	1	138

Notes: Panel A presents descriptive statistics for the baseline sample period March 27–30 (Pre) and April 3–6 (Post). The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of unique repository × GitHub user accounts for the entire repository–user-level sample ("Overall") and for each of the subsamples is presented in parentheses below. Panel B provides information on the individual characteristics of all GitHub user accounts in the repository–user-level sample.

C.2 Additional Results

Table C.2: Effect of ChatGPT Ban on GitHub Output — Repo-Level Analysis

		Output Quantity				Output Quality			Task Complexity	
		Output	Output (Shen, 2023)	Commit	Issue closed	Bug fix ratio	PR merged	PR merge ratio	Files edited per PR	Lines added per PR
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Pre Mar 27–28 – Post Apr 3–4										
Overall (N = 93,248)	Treated × Post	-0.0011	-0.0012	-0.0018	-0.0006	0.0001	-0.0008	-0.0667	-0.0007	-0.0033
		(0.0015)	(0.0018)	(0.0014)	(0.0009)	(0.0004)	(0.0010)	(0.0957)	(0.0022)	(0.0047)
Less experienced (N = 34,768)	Treated × Post	0.0006	0.0012	-0.0006	0.0002	0.0003	0.0011	0.1042	0.0050*	0.0096
		(0.0020)	(0.0025)	(0.0018)	(0.0010)	(0.0006)	(0.0012)	(0.1128)	(0.0029)	(0.0065)
Experienced (N = 58,480)	Treated × Post	-0.0020	-0.0026	-0.0023	-0.0011	0.0000	-0.0019	-0.1649	-0.0044	-0.0114*
		(0.0022)	(0.0024)	(0.0019)	(0.0013)	(0.0006)	(0.0015)	(0.1416)	(0.0031)	(0.0065)
Pkg. contributor (N = 23,668)	Treated × Post	-0.0039	-0.0053	-0.0037	-0.0017	-0.0026**	-0.0003	-0.0078	-0.0011	-0.0025
		(0.0039)	(0.0045)	(0.0035)	(0.0023)	(0.0012)	(0.0025)	(0.2399)	(0.0050)	(0.0103)
Pre Mar 27–30 – Post Apr 3–6										
Overall (N = 186,496)	Treated × Post	-0.0014	-0.0010	-0.0013	-0.0007	0.0001	-0.0002	-0.0190	-0.0007	-0.0022
		(0.0012)	(0.0014)	(0.0011)	(0.0007)	(0.0003)	(0.0007)	(0.0704)	(0.0017)	(0.0035)
Less experienced (N = 69,536)	Treated × Post	0.0019	0.0026	0.0008	0.0016*	0.0007	0.0014*	0.1319	0.0038	0.0085*
		(0.0016)	(0.0019)	(0.0014)	(0.0008)	(0.0004)	(0.0008)	(0.0828)	(0.0025)	(0.0049)
Experienced (N = 116,960)	Treated × Post	-0.0035**	-0.0032*	-0.0025	-0.0022**	-0.0002	-0.0012	-0.1098	-0.0034	-0.0089*
		(0.0017)	(0.0019)	(0.0016)	(0.0010)	(0.0004)	(0.0011)	(0.1042)	(0.0023)	(0.0047)
Pkg. contributor (N = 47,336)	Treated × Post	-0.0051*	-0.0031	-0.0031	-0.0023	-0.0011	0.0002	0.0267	-0.0031	-0.0052
		(0.0031)	(0.0035)	(0.0029)	(0.0017)	(0.0008)	(0.0018)	(0.1735)	(0.0038)	(0.0079)

Notes: All specifications include repository × user fixed effects. The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the repository × user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Repository-Level Analysis — March 27–28 (Pre) – April 3–4 (Post)

		Output Quantity				Output Quality			Task Complexity	
		Output	Output (Shen, 2023)	Commit	Issue closed	Bug fix ratio	PR merged	PR merge ratio	Files edited per PR	Lines added per PR
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overall (N = 93,248)	Treated × Post	-0.0011	-0.0012	-0.0018	-0.0006	0.0001	-0.0008	-0.0667	-0.0007	-0.0033
		(0.0015)	(0.0018)	(0.0014)	(0.0009)	(0.0004)	(0.0010)	(0.0957)	(0.0022)	(0.0047)
	Post	0.0008	0.0011	0.0009	0.0001	0.0001	0.0007	0.0631	0.0021*	0.0041*
		(0.0007)	(0.0008)	(0.0006)	(0.0004)	(0.0002)	(0.0005)	(0.0447)	(0.0012)	(0.0024)
	Dep. var. mean	0.0124	0.0155	0.0095	0.0036	0.0010	0.0048	0.4618	0.0097	0.0212
Less experienced (N = 34,768)	Treated × Post	0.0006	0.0012	-0.0006	0.0002	0.0003	0.0011	0.1042	0.0050*	0.0096
		(0.0020)	(0.0025)	(0.0018)	(0.0010)	(0.0006)	(0.0012)	(0.1128)	(0.0029)	(0.0065)
	Post	-0.0012	-0.0012	-0.0010	0.0000	-0.0003	-0.0009*	-0.0864*	-0.0010	-0.0034
		(0.0008)	(0.0010)	(0.0007)	(0.0005)	(0.0003)	(0.0005)	(0.0515)	(0.0012)	(0.0026)
	Dep. var. mean	0.0075	0.0106	0.0056	0.0020	0.0007	0.0024	0.2373	0.0053	0.0121
Experienced (N = 58,480)	Treated × Post	-0.0020	-0.0026	-0.0023	-0.0011	0.0000	-0.0019	-0.1649	-0.0044	-0.0114*
		(0.0022)	(0.0024)	(0.0019)	(0.0013)	(0.0006)	(0.0015)	(0.1416)	(0.0031)	(0.0065)
	Post	0.0020**	0.0023**	0.0019**	0.0002	0.0004	0.0016**	0.1457**	0.0037**	0.0082**
		(0.0010)	(0.0011)	(0.0008)	(0.0006)	(0.0003)	(0.0007)	(0.0633)	(0.0017)	(0.0034)
	Dep. var. mean	0.0153	0.0184	0.0118	0.0046	0.0011	0.0062	0.5953	0.0123	0.0267
Pkg. contributor (N = 23,668)	Treated × Post	-0.0039	-0.0053	-0.0037	-0.0017	-0.0026**	-0.0003	-0.0078	-0.0011	-0.0025
		(0.0039)	(0.0045)	(0.0035)	(0.0023)	(0.0012)	(0.0025)	(0.2399)	(0.0050)	(0.0103)
	Post	0.0039**	0.0024	0.0024*	0.0003	0.0006	0.0016	0.1394	0.0058**	0.0118**
		(0.0016)	(0.0020)	(0.0014)	(0.0010)	(0.0005)	(0.0011)	(0.1053)	(0.0025)	(0.0052)
	Dep. var. mean	0.0192	0.0243	0.0142	0.0055	0.0014	0.0077	0.7437	0.0147	0.0326

Notes: All specifications include repository × user fixed effects. The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the repository × user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

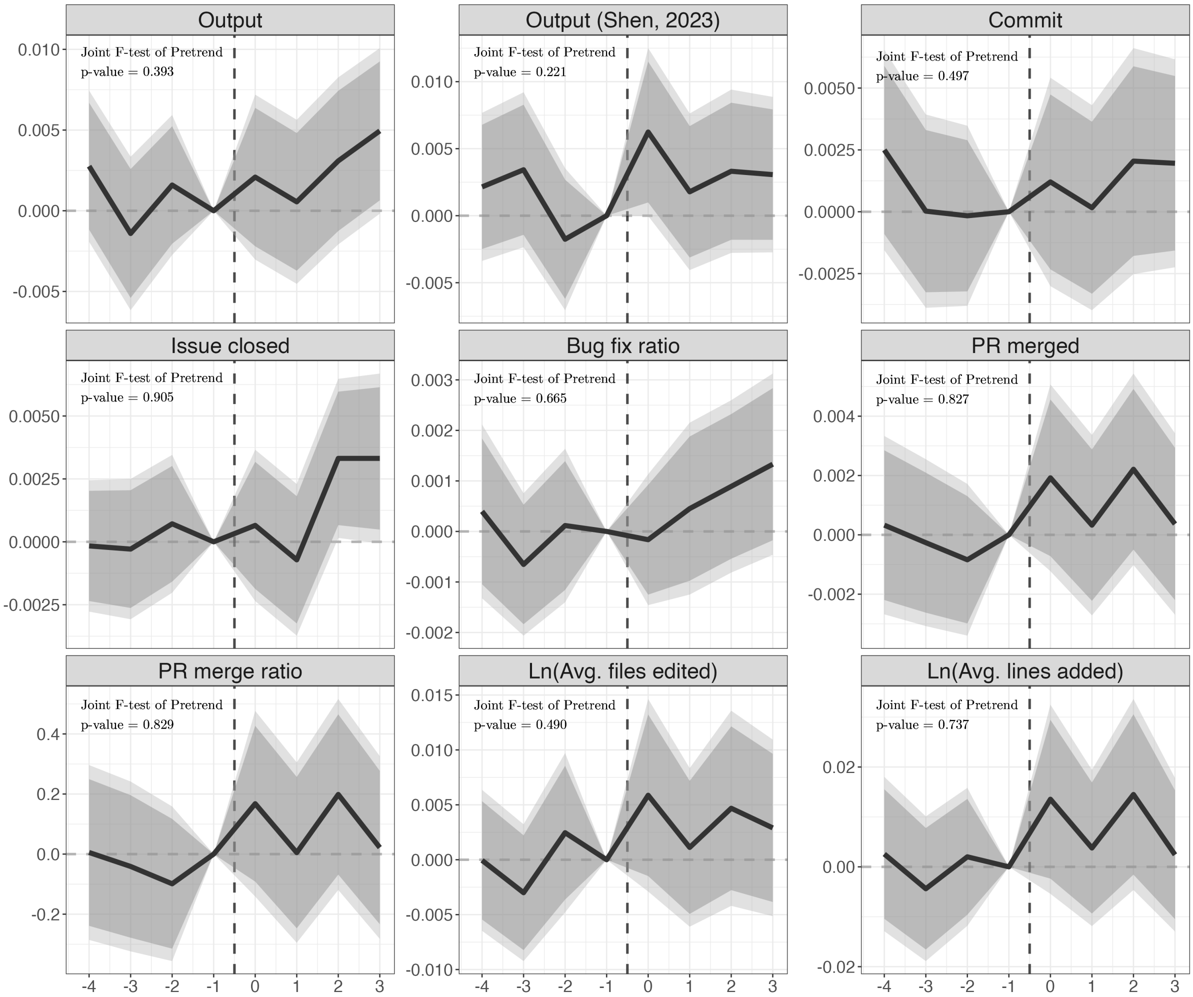
Table C.4: Repository-Level Analysis — March 27–30 (Pre) – April 3–6 (Post)

		Output Quantity				Output Quality			Task Complexity	
		Output	Output (Shen, 2023)	Commit	Issue closed	Bug fix ratio	PR merged	PR merge ratio	Files edited per PR	Lines added per PR
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Overall (N = 186,496)	Treated × Post	-0.0014	-0.0010	-0.0013	-0.0007	0.0001	-0.0002	-0.0190	-0.0007	-0.0022
		(0.0012)	(0.0014)	(0.0011)	(0.0007)	(0.0003)	(0.0007)	(0.0704)	(0.0017)	(0.0035)
	Post	0.0004	0.0004	0.0003	0.0002	0.0002	0.0003	0.0322	0.0021***	0.0036**
		(0.0005)	(0.0006)	(0.0005)	(0.0003)	(0.0002)	(0.0003)	(0.0315)	(0.0008)	(0.0016)
	Dep. var. mean	0.0122	0.0152	0.0095	0.0035	0.0010	0.0048	0.4620	0.0095	0.0211
Less experienced (N = 69,536)	Treated × Post	0.0019	0.0026	0.0008	0.0016*	0.0007	0.0014*	0.1319	0.0038	0.0085*
		(0.0016)	(0.0019)	(0.0014)	(0.0008)	(0.0004)	(0.0008)	(0.0828)	(0.0025)	(0.0049)
	Post	-0.0013**	-0.0017**	-0.0011**	-0.0004	-0.0002	-0.0008**	-0.0763**	-0.0005	-0.0026
		(0.0007)	(0.0008)	(0.0006)	(0.0004)	(0.0002)	(0.0004)	(0.0355)	(0.0009)	(0.0020)
	Dep. var. mean	0.0072	0.0101	0.0053	0.0021	0.0006	0.0023	0.2235	0.0055	0.0124
Experienced (N = 116,960)	Treated × Post	-0.0035**	-0.0032*	-0.0025	-0.0022**	-0.0002	-0.0012	-0.1098	-0.0034	-0.0089*
		(0.0017)	(0.0019)	(0.0016)	(0.0010)	(0.0004)	(0.0011)	(0.1042)	(0.0023)	(0.0047)
	Post	0.0014**	0.0016**	0.0011*	0.0005	0.0005**	0.0010**	0.0921**	0.0035***	0.0070***
		(0.0007)	(0.0008)	(0.0006)	(0.0004)	(0.0002)	(0.0005)	(0.0447)	(0.0011)	(0.0023)
	Dep. var. mean	0.0152	0.0182	0.0119	0.0044	0.0012	0.0063	0.6038	0.0119	0.0263
Pkg. contributor (N = 47,336)	Treated × Post	-0.0051*	-0.0031	-0.0031	-0.0023	-0.0011	0.0002	0.0267	-0.0031	-0.0052
		(0.0031)	(0.0035)	(0.0029)	(0.0017)	(0.0008)	(0.0018)	(0.1735)	(0.0038)	(0.0079)
	Post	0.0027**	0.0008	0.0008	0.0005	0.0004	0.0008	0.0773	0.0054***	0.0106***
		(0.0013)	(0.0015)	(0.0011)	(0.0007)	(0.0004)	(0.0009)	(0.0812)	(0.0018)	(0.0040)
	Dep. var. mean	0.0189	0.0237	0.0144	0.0052	0.0015	0.0078	0.7471	0.0149	0.0331

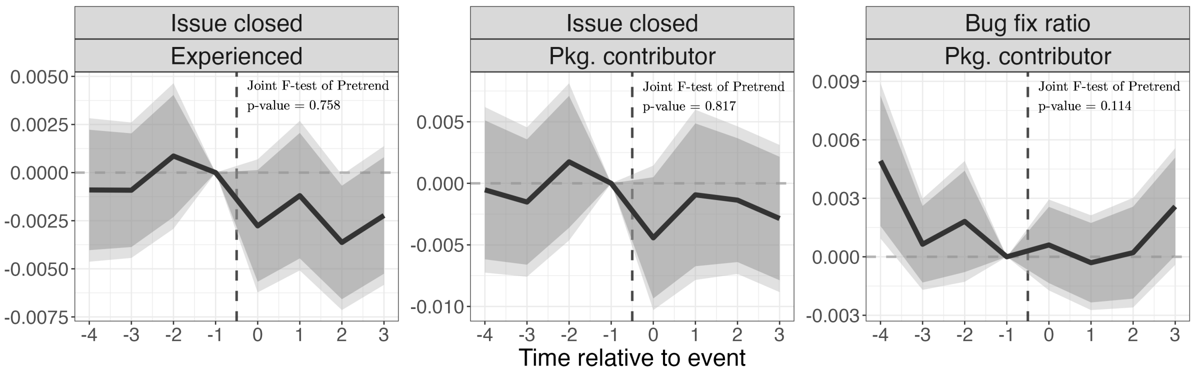
Notes: All specifications include repository × user fixed effects. The "Less experienced" sample includes all GitHub user accounts created after or in 2017 (median), while the "Experienced" sample comprises all GitHub user accounts created before 2017. The "Pkg. contributor" sample comprises all GitHub user accounts that are the owner and/or contributor to a(n) (analytical) programming package repository. The number of observations is in parentheses after each sample definition. Robust standard errors in parentheses are clustered at the repository × user level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure C.1: Repository-Level Analysis

(a) Effect of ChatGPT Ban on "Less experienced" User



(b) Effect of ChatGPT Ban on "Proficient" User



Notes: Panel A displays event-study estimates across outcomes for “less experienced” GitHub user accounts (created after or in 2017). Panel B presents event-study estimates for *(i)* “experienced” GitHub user accounts (created before 2017) and *(ii)* and accounts that are the owner and/or contributor to a(n) (analytical) programming package repository (“Pkg. contributor”). The sample period spans March 27–30 (*Pre*) and April 3–6 (*Post*). All specifications include user × repository, time, and day-of-the-week fixed effects; 95% (90%) confidence intervals for robust standard errors clustered at the user × repository level are depicted in light (dark) grey.

D User Adaption to the ChatGPT Ban

Considering that there appears to be mean reversion in the estimated effects toward the end of our sample period, we now turn our attention to adaptation behaviour. The simplest way to circumvent the ChatGPT ban was to use VPN tools or encrypted routing through, for instance, the TOR network.

D.1 Data

We collect daily data on the number of *Google searches* on the topic of "Virtual Private Networks" from *Google Trends* and on the number of users of *TOR*, an open-source software for enabling anonymous communication, from *TOR Metrics* for all 25 countries in the European Union.[23] We retrieve information on both the number of users of "standard" *TOR relays* and of *TOR bridge* relays to examine whether there were changes in the use of, in particular, *TOR bridge* relays, which are not listed publicly and therefore are more difficult for firewalls to identify.[24] We apply a log transformation to both user numbers. The sample period under consideration covers March 13, 2023, the day after the release of ChatGPT-4, until April 7, 2023, the end of the workweek post-ban. Observations on weekends are dropped from the sample since we are interested in the effect of the ban on output. Figure D.1 provides a graphic illustration of the final panel structure.

D.2 Results

To estimate the average treatment effect of the ChatGPT ban on users from Italy, we apply the generalized synthetic control method proposed by Xu (2017). The treatment effect on the treated unit (ATT) is the difference between the actual outcome and its estimated counterfactual. To obtain the counterfactual, a (cross-validated) interactive fixed effects (IFE) model is estimated for the control group data.[25] All IFE models incorporate additive unit and time fixed effects.[26] To draw inference, we rely on the parametric bootstrap procedure suggested by Xu (2017) for settings with a small number of treated units.

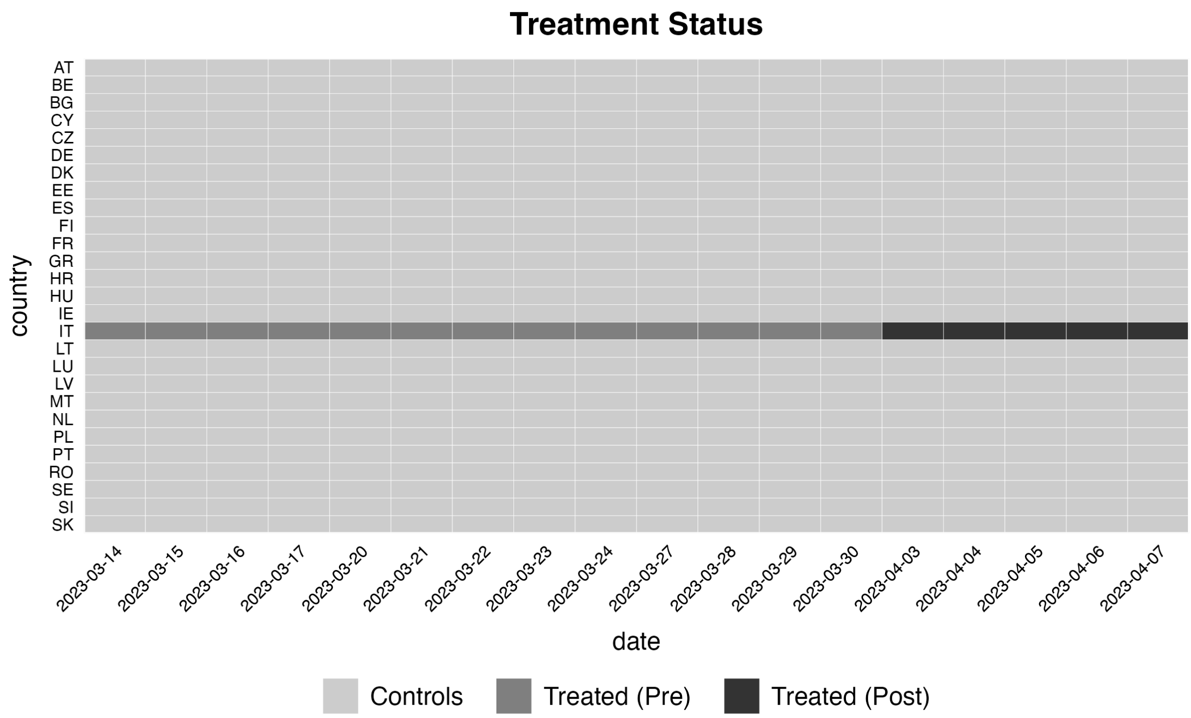
[23] Google trends data have been widely used in economic research as a predictor of human behavioural economic phenomena (Choi and Varian, 2012). For example, Böhme et al. (2020) used Google trends data on migration-related Google search terms to predict international migration, while Ginsberg et al. (2009) used trends data to predict influenza outbreaks.

[24] Note that *TOR bridge* relays can, however, slow down the connection. For more information on bridges vs. "standard" relays, please refer to the official *TOR* documentation at `https://tb-manual.torproject.org/bridges/`.

[25] Specifically, we apply the EM algorithm proposed by Gobillon and Magnac (2016) and implemented in the **R** package `gsynth` (Xu and Liu, 2022), which additionally uses treatment group information for the pre-treatment period, leading to (slightly) more precisely estimated coefficients.

[26] Note that the *Google trends* data are already standardized by country for the selected time period such that we include only time fixed effects in this case when estimating the IFE model.

Figure D.1: Panel Structure of Google Trends and TOR Data

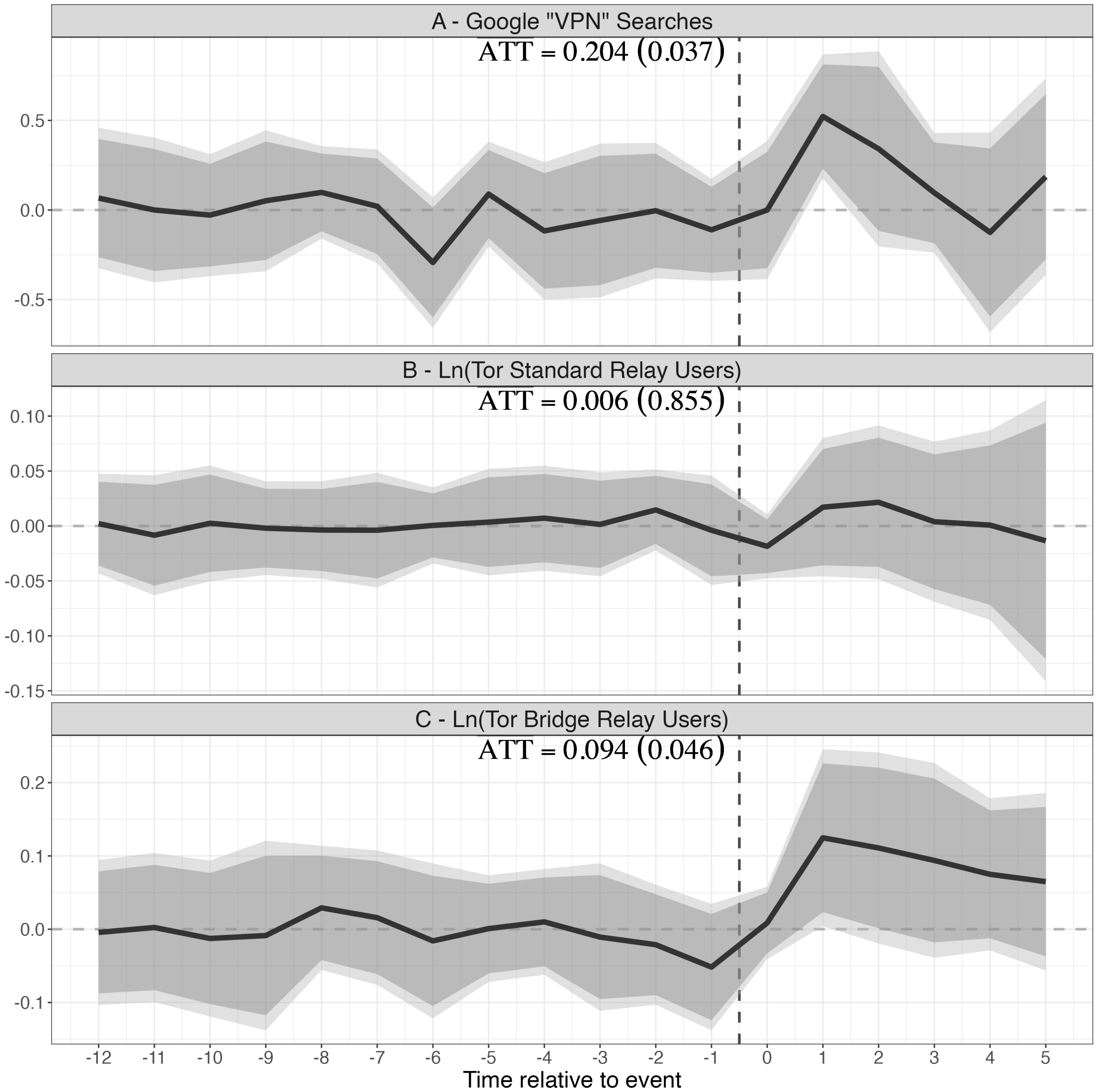


Notes: The panel structure of the datasets used in Section D—i.e., the *(i) Google trends* and the *(ii) TOR* (*"standard"* and *bridge relay*) user datasets—are displayed. Workday dates during the sample period are presented on the x-axis, and *ISO 3166-1 alpha-2* country codes are presented on the y-axis.

The top panel in Figure D.2 presents the effect of the ChatGPT ban on the number of *Google queries* on the topic of VPNs. We observe a significant increase in the share of web searches on this relative to other topics in Italy on the first working day after the ban that slowly vanishes over the next three days. The estimated effect on April 3 is sizeable: the share of searches on VPNs increases by 52.2 percentage points. On average, the share of queries on VPNs was 20.6 percentage points higher in Italy over the workweek. The observed pattern is consistent with Italian users looking for ways to access ChatGPT even after the ban and succeeding after some initial search costs. Our estimates might, however, present only *stated* preferences.

To investigate whether the ban actually led to behavioral changes among Italian users, we look at an alternative outcome: the log number of *TOR* users. The results for *TOR relay* and *TOR bridge* users are presented in Panels B and C of Figure D.2, respectively. While the number of *TOR relay* users shows only a minor increase in the days after the ban, the average treatment effect on the number of Italian *TOR bridge* users is positive and significant on the first workday after the ban. Usage of *TOR bridges* remained elevated for the entire workweek, with an increase in user numbers—on average—of approximately 9.4 percentage points. This pattern is in line with users

Figure D.2: Effect of ChatGPT Ban on Ban Circumvention Technology



Notes: The dynamic treatment effects estimates for the generalized synthetic control method of Xu (2017) are depicted. The top panel presents the ATT for the number of *Google queries* on the topic of VPNs. The bottom panel presents the ATT for *TOR bridge* relay users. The counterfactual for the treated unit (Italy) is estimated with an interactive fixed effects model; 95% (90%) confidence intervals from the parametric bootstrap procedure proposed by Xu (2017) are displayed in light (dark) grey. Additionally, the *mean* ATT over the workweek after the ChatGPT ban and its p-value (in parentheses) are presented.

resorting to *bridge* over "standard" relays to minimize the chance of their being denied access to ChatGPT since the former are more difficult for firewalls to identify.[27]

Overall, our findings are consistent with Italian users looking for and finding ways to

[27]For a discussion on denial of ChatGPT access, see the following OpenAI forum discussion: https://community.openai.com/t/access-denied-error-1020/38758/23.

circumvent the blocked access to ChatGPT.

Appendix References

Böhme, M. H., Gröger, A., and Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, 142:102347.

Choi, H. and Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, 88(s1):2–9.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics esing search engine query data. *Nature*, 457:1012–1014.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive Fixed Effects and Synthetic Controls. *Review of Economics and Statistics*, 98(3):535–551.

Holub, F. and Thies, B. (2023). Air quality, high-skilled worker productivity and adaptation: Evidence from GitHub. CRC TR 224 Discussion Paper Series crctr224_2023_402, University of Bonn and University of Mannheim, Germany.

Shen, L. (2023). Does working from home work? A natural experiment from lockdowns. *European Economic Review*, 151:104323.

Xu, Y. (2017). Generalized Synthetic Control Method: Causal inference with Interactive Fixed Effects Models. *Political Analysis*, 25(1):57–76.

Xu, Y. and Liu, L. (2022). *gsynth: Generalized Synthetic Control Method.* R package version 1.2.1.