Estimating the Impact of ChatGPT on Developer Productivity Using Italy's ChatGPT Ban as Natural Experiment

RICHARD SCHULZ*, ETH Zurich, Switzerland

SIMON SCHUHMACHER*, University of Zurich, Switzerland

Motivation & Related Work

There has been substantial anecdotal evidence suggesting that AI-assisted coding tools such as ChatGPT and GitHub Copilot have significantly increased developer productivity. However, recent empirical research has shown mixed results [1, 3, 4]. While some studies find notable efficiency gains from integrating generative AI into software development workflows—for instance, faster code completion and reduced debugging time—other work points to limited or highly context-dependent effects, with improvements varying by task complexity, developer experience, and integration depth.

Despite this growing literature, much of the existing evidence is correlational or based on controlled experiments detached from real-world settings. We still know little about how sudden changes in access to such tools affect developers in practice. In particular, the short-lived ChatGPT ban in Italy in early 2023 offers a rare quasi-experimental opportunity to observe developers' behavioral response to an abrupt disruption in access to a key AI tool.

Building on Kreitmeir and Raschky [2], who documented a 50% decline in Italian project releases during the ban, this research moves one step closer to identifying how this productivity loss materialized. However, the effect on the granular, day-to-day development process remains unquantified. It is unclear if developers wrote less code, or if the quality of their code suffered, leading to downstream effects. This research fills that gap by investigating how the sudden removal of a key AI tool impacts the fundamental activities of coding: the volume and quality of contributions.

Methodology & Data

The analysis will use historical GitHub activity data from users in Italy, Austria, and France for the weeks surrounding the late March/early April 2023 ban. To this end, we will leverage the GH Archive dataset,[1] which provides a comprehensive record of public GitHub events, including *PushEvents*, *ReleaseEvents*, and *IssueEvents*. Because GH Archive data does not include information about developers' locations, we will infer user countries based on their public profile information, following the methods from prior research [2]. This information will be retrieved through the GitHub GraphQL API.

As *PushEvents* data do not contain actual code changes, we will also use the GitHub REST API to fetch corresponding commit details, enabling us to compute lines of code added and deleted. Finally, we will integrate data from all three sources to construct a unified dataset containing relevant activity metrics at the developer-day level.

We will employ a **Difference-in-Differences (DiD)** quasi-experimental design, replicating the framework from the original paper to ensure causal inference. The analysis will compare the change in developer metrics in Italy (treatment group) against those in Austria and France (control group) before and after the ban's implementation.

---

*Both authors contributed equally to this research.
[1]https://www.gharchive.org/

Authors' Contact Information: Richard Schulz, rschul@ethz.ch, ETH Zurich, Zurich, Switzerland; Simon Schuhmacher, simon.schuhmacher@uzh.ch, University of Zurich, Zurich, Switzerland.

In contrast to the original study—which measured aggregate productivity through project releases—our analysis zooms in on code-level and issue-level metrics that capture the day-to-day development process. The main dependent variables are:

- **Code Volume**: The number of lines of code (LOC) added and deleted per developer per day. This provides a granular measure of active coding behavior.
- **Bug-Related Issue Creation**: As an indirect quality proxy, we analyze whether the number of bug-related issues (identified through issue titles and labels containing keywords such as "bug," "fix," or "error") increased in the weeks following the ban within affected repositories.

All variables will be standardized at the developer-day level to account for individual differences in coding activity. Control variables will include developer experience (tenure on GitHub), project size, and temporal trends (day-of-week effects).

Research Hypotheses

- **H1 (Volume)**: The ChatGPT ban caused a statistically significant reduction in the daily volume of code contributed by developers in the treatment group, measured through lines of code added and deleted per user per day.
- **H2 (Defect Incidence)**: Projects affected by the ban experienced an increase in bug-related issue creation in the subsequent weeks compared to the control group, indicating that the temporary loss of ChatGPT led to a higher defect rate or more post-release debugging effort.

Project Outline

(1) **Replication & Validation**: First, we reproduce the main result from Kreitmeir and Raschky [2] – the temporary 50% drop in Release events for Italian developers.
(2) **Extension & Analysis**: Second, we apply the validated model to our new dependent variables (Volume and Defect Incidence) to test hypotheses H1 and H2. Together, these analyses will offer a richer understanding of how generative AI tools influence developer productivity — not only in terms of total output but also in terms of the quality of the produced code.

References

[1] Zheyuan (Kevin) Cui, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. 2025. The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers. (June 2025). https://www.microsoft.com/en-us/research/publication/the-effects-of-generative-ai-on-high-skilled-work-evidence-from-three-field-experiments-with-software-developers/

[2] David H. Kreitmeir and Paul A. Raschky. 2023. The Unintended Consequences of Censoring Digital Technology – Evidence from Italy's ChatGPT Ban. arXiv:2304.09339 [econ.GN] https://arxiv.org/abs/2304.09339

[3] Elise Paradis, Kate Grey, Quinn Madison, Daye Nam, Andrew Macvean, Vahid Meimand, Nan Zhang, Ben Ferrari-Church, and Satish Chandra. 2024. How much does AI impact development speed? An enterprise-based randomized controlled trial. arXiv:2410.12944 [cs.SE] https://arxiv.org/abs/2410.12944

[4] Sida Peng, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. The Impact of AI on Developer Productivity: Evidence from GitHub Copilot. arXiv:2302.06590 [cs.SE] https://arxiv.org/abs/2302.06590