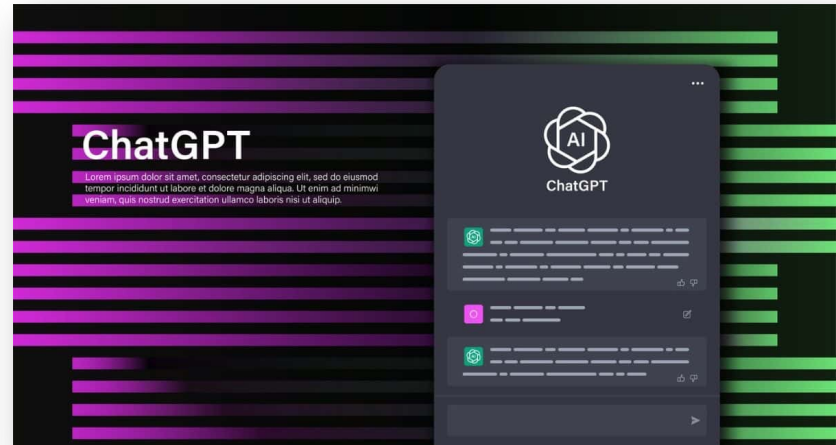
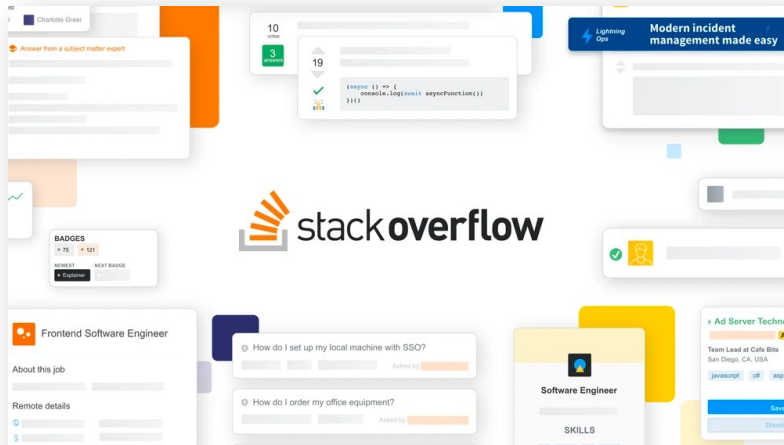


The Consequences of Generative AI for Online Knowledge Communities



Gordon Burtch
(/w Dokyun Lee & Zhichen Chen)
Boston University, Questrom School of Business



OPEN The consequences of generative AI for online knowledge communities

Gordon Burch¹, Dokyun Lee & Zhichen Chen

Generative artificial intelligence technologies, especially large language models (LLMs) like ChatGPT, are revolutionizing information acquisition and content production across a variety of domains. These technologies have a significant potential to impact participation and content production in online knowledge communities. We provide initial evidence of this, analyzing data from Stack Overflow and Reddit developer communities between October 2021 and March 2023, documenting ChatGPT's influence on user activity in the former. We observe significant declines in both website visits and question volumes at Stack Overflow, particularly around topics where ChatGPT excels. By contrast, activity in Reddit communities shows no evidence of decline, suggesting the importance of social fabric as a buffer against the community-degrading effects of LLMs. Finally, the decline in participation on Stack Overflow is found to be concentrated among newer users, indicating that more junior, less socially embedded users are particularly likely to exit.

Recent advancements in generative artificial intelligence (Gen AI) technologies, especially large language models (LLMs) such as ChatGPT, have been significant. LLMs demonstrate remarkable proficiency in tasks that involve information retrieval and content creation^{1–3}. Given these capabilities, it is important to consider their potential to drive seismic shifts in the way knowledge is developed and exchanged within online knowledge communities^{4–6}.

LLMs may drive both positive and negative impacts on participation and activity at online knowledge communities. On the positive side, LLMs can enhance knowledge sharing by providing immediate, relevant responses to user queries, potentially bolstering community engagement by helping users to efficiently address a wider range of peer questions. Viewed from this perspective, Gen AI tools may complement and enhance existing activities in a community, enabling a greater supply of information. On the negative side, LLMs may replace online knowledge communities altogether.

If the displacement effect dominates, it would give rise to several serious concerns. First, while LLMs offer innovative solutions for information retrieval and content creation and have been shown to significantly enhance individual productivity in a variety of writing and coding tasks, they have also been found to hallucinate, i.e., providing 'confidently incorrect' responses to user queries⁷, and to undermine worker performance on certain types of tasks⁸. Second, if individual participation in online communities were to decline, this would imply a decline in opportunities for all manner of interpersonal interaction, upon which many important activities depend, e.g., collaboration, mentorship, job search. Further, to the extent a similar dynamic may emerge within formal organizations and work contexts, it would raise the prospect of analogous declines in organizational attachment, peer learning, career advancement and innovation^{9–11}.

With the above in mind, we address two questions in this work. First, we examine the effects that generative artificial intelligence (AI), particularly large language models (LLMs), have on individual engagement in online knowledge communities. Specifically, we assess how LLMs influence user participation and content creation in online knowledge communities. Second, we explore factors that moderate (amplify or attenuate) the effects of LLMs on participation and content creation at online knowledge communities. By addressing these relationships, we aim to advance our understanding of the role LLMs may play in shaping the future of knowledge sharing and collaboration online. Further, we seek to provide insights into approaches and strategies that can encourage a sustainable knowledge sharing dynamic between human users and AI technologies.

We evaluate our questions in the context of ChatGPT's release, in late November of 2022. We start by examining how the release of ChatGPT impacted Stack Overflow. We show that ChatGPT's release led to a marked decline in web traffic to Stack Overflow, and a commensurate decline in question posting volumes. We then consider how declines in participation may vary across community contexts. Leveraging data on posting activity in Reddit developer communities over the same period, we highlight a notable contrast: no detectable declines in participation. We attribute this difference to social fabric; whereas Stack Overflow focuses on pure information exchange, Reddit developer communities are characterized by stronger social bonds. Further, considering heterogeneity across topic domains within Stack Overflow, we show that declines in participation varied greatly

Questrom School of Business, Boston University, Boston, MA 02215, USA. [✉]email: gburch@bu.edu

OPINION

Artificial Intelligence and Machine Learning

Generative AI Degrades Online Communities

How large language models are influencing online communities.

By Gordon Burch¹, Dokyun Lee, and Zhichen Chen

Posted Feb 16 2024



Share Download PDF Print Join the Discussion View in the ACM Digital Library

(Some) Online Knowledge Communities Are Struggling

ChatGPT Excels When It Has Training Data

ChatGPT Still Does Not Substitute for Human Social Connections

Research Agenda: Knowledge Management in the Era of Generative AI

Imagine you are at a crossroads in a complex project and you need quick answers on how to grapple with a problem. It is quite likely that you might turn to an online knowledge community for answers, one hosted by your company, or perhaps Stack Overflow, Quora, or Reddit. These communities have come to play a central role in knowledge exchange, in many corners of the economy and society, but they depend on voluntary participation from users just like you and me.

Our recent research indicates an intriguing shift is now taking place: generative AI technologies, such as OpenAI's large language model (LLM) ChatGPT, are disrupting

DOI
10.1145/3624732

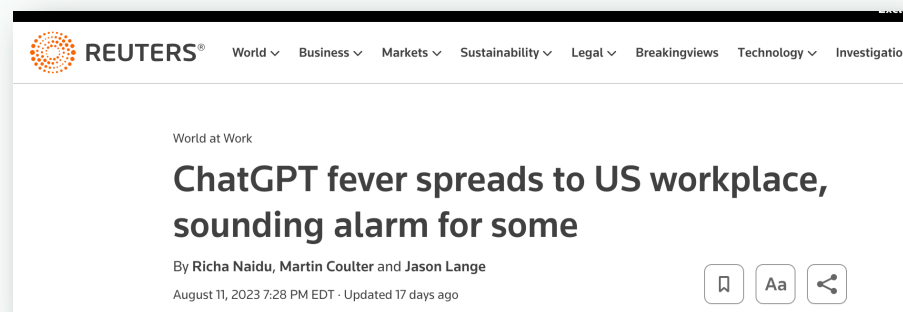
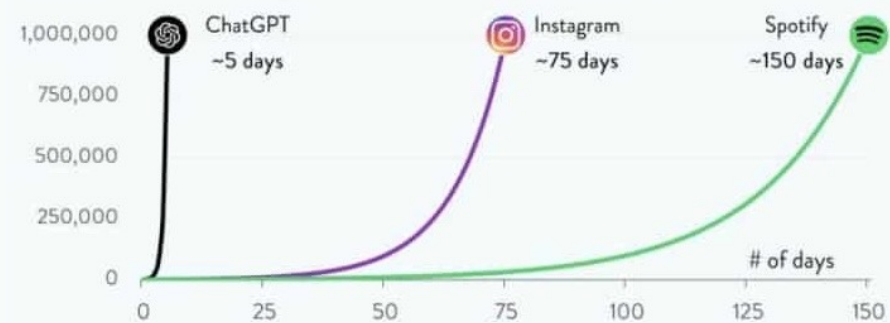
March 2024 Issue
Vol. 67 No. 3
Pages: 40–42

[Table of Contents](#) →

Widespread Adoption of LLMs

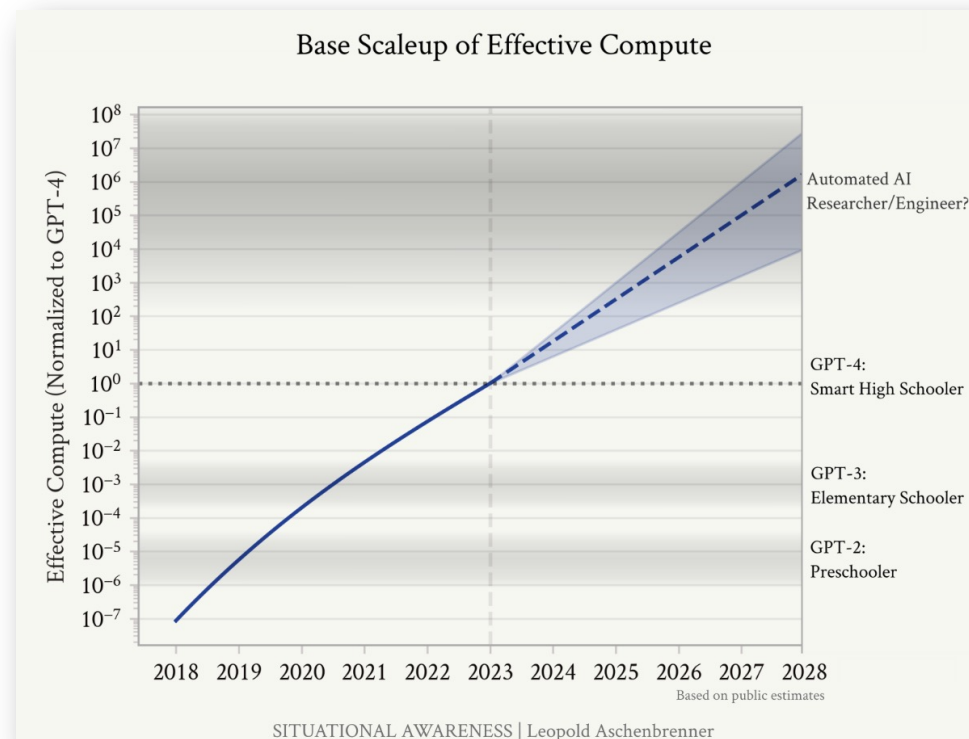
- Typical workplace use cases (Business.com survey of 2,000 American workers):

- Writing and refining emails, letters and reports.
- **Brainstorming.**
- **Debugging code.**
- **Researching a topic.**



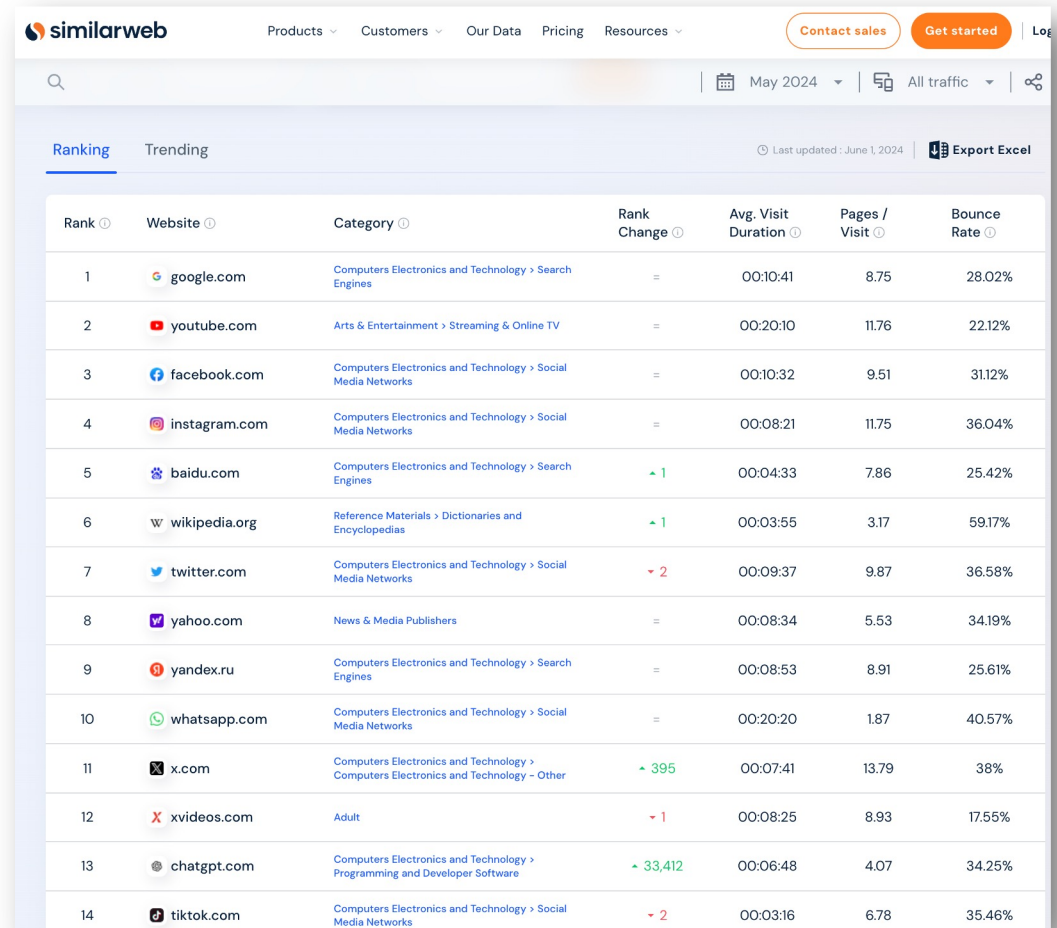
LLMs Exhibit Amazing Capabilities

- Amazing ability to produce code, performance on standardized tests, etc.
- Presently operating like a smart high-schooler. What is to come with collaborative agents, etc.?



When and to What Extent Does Gen AI Substitute for Human Communities?

- The usage of these tools now rivals usage of some of the world's largest online communities and social media services, which also focus on content production and knowledge dissemination ...
 - Reddit, X (Twitter), Facebook Groups, Wikipedia, LinkedIn, *Stack Exchange*

A screenshot of the Similarweb website showing a global ranking of websites. The page includes a navigation bar with links to Products, Customers, Our Data, Pricing, and Resources. It also features a search bar, a date selector set to May 2024, and a traffic source selector set to All traffic. The main content area displays a table of website rankings. The table has columns for Rank, Website, Category, Rank Change, Avg. Visit Duration, Pages / Visit, and Bounce Rate. The top 14 websites are listed, with Google.com at rank 1 and TikTok.com at rank 14. The table is updated as of June 1, 2024, and there is an option to export the data to Excel.

Rank	Website	Category	Rank Change	Avg. Visit Duration	Pages / Visit	Bounce Rate
1	google.com	Computers Electronics and Technology > Search Engines	=	00:10:41	8.75	28.02%
2	youtube.com	Arts & Entertainment > Streaming & Online TV	=	00:20:10	11.76	22.12%
3	facebook.com	Computers Electronics and Technology > Social Media Networks	=	00:10:32	9.51	31.12%
4	instagram.com	Computers Electronics and Technology > Social Media Networks	=	00:08:21	11.75	36.04%
5	baidu.com	Computers Electronics and Technology > Search Engines	▲ 1	00:04:33	7.86	25.42%
6	wikipedia.org	Reference Materials > Dictionaries and Encyclopedias	▲ 1	00:03:55	3.17	59.17%
7	twitter.com	Computers Electronics and Technology > Social Media Networks	▼ 2	00:09:37	9.87	36.58%
8	yahoo.com	News & Media Publishers	=	00:08:34	5.53	34.19%
9	yandex.ru	Computers Electronics and Technology > Search Engines	=	00:08:53	8.91	25.61%
10	whatsapp.com	Computers Electronics and Technology > Social Media Networks	=	00:20:20	1.87	40.57%
11	x.com	Computers Electronics and Technology > Computers Electronics and Technology - Other	▲ 395	00:07:41	13.79	38%
12	xvideos.com	Adult	▼ 1	00:08:25	8.93	17.55%
13	chatgpt.com	Computers Electronics and Technology > Programming and Developer Software	▲ 33,412	00:06:48	4.07	34.25%
14	tiktok.com	Computers Electronics and Technology > Social Media Networks	▼ 2	00:03:16	6.78	35.46%

Concerns

1st Order: LLMs ‘Hallucinate’ (Cao et al. 2021; Kabir et al. 2023; Kocoń et al. 2023) **and Non-experts Don’t Know What They Don’t Know**

- Most coding errors that ChatGPT makes are conceptual, not syntactic. LLMs may reply with incorrect information without obvious signs, e.g., code execution errors (Kabir et al. 2023).

Finding 3

ChatGPT rarely makes syntax errors for code answers. The majority of the code errors are due to applying wrong logic or implementing non-existing or wrong API, Library, or Functions.

2nd Order: A Decline in Online Community Participation Reflects a Broader Decline in Peer Connection

- Peer connection is important for learning which others in the organization have relevant expertise, particularly for junior employees (Wu and Kane 2021). Peer connection is also important for innovation.
- Peer connection is important for organizational attachment and for career advancement (Förderer and Burtch 2023; Huang and Zhang 2016).

Our Research Questions

To what extent does information-seeking (from peers) decline?

Test: changes in web traffic and question posting volumes at online knowledge communities.

When are those effects largest?

Test: heterogeneity in contributions by topic, user characteristics, and contrasting social vs. asocial communities.

Are there second-order declines in quality of knowledge sharing?

Test: changes in average net-voting outcomes for Stack Overflow answers.

Study Context

- **Stack Overflow**

- The first, and largest online community within the Stack Exchange ecosystem, created in 2008. It is presently one of the most popular online communities for Q&A related to programming.
- Stack Overflow Teams is a business product offering that serves a similar function within organizations, for knowledge sharing among employees.



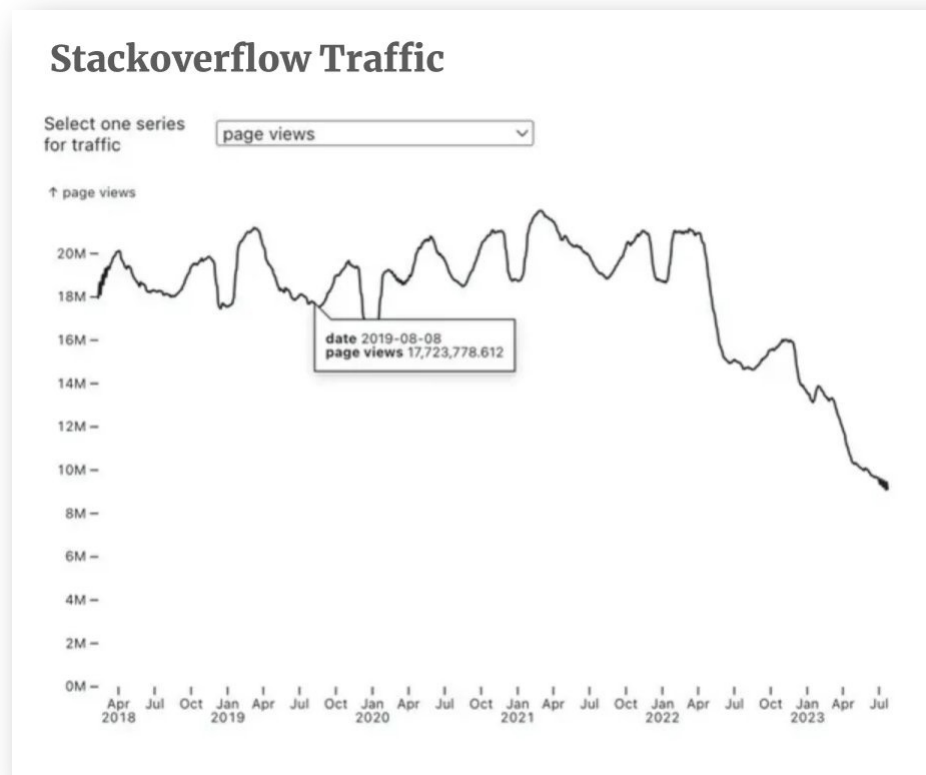
Study Context

- **Reddit**

- 10th most visited website in the world, 6th most visited in the US as of February 2023. A general-purpose online community, comprised of various sub-communities, i.e., sub-Reddits, including robust programming communities.
- We consider sub-Reddits that focus upon topics represented among our ‘popular’ Stack Overflow topic tags.



Effects on Participation: Web Traffic



Effects on Participation: Web Traffic

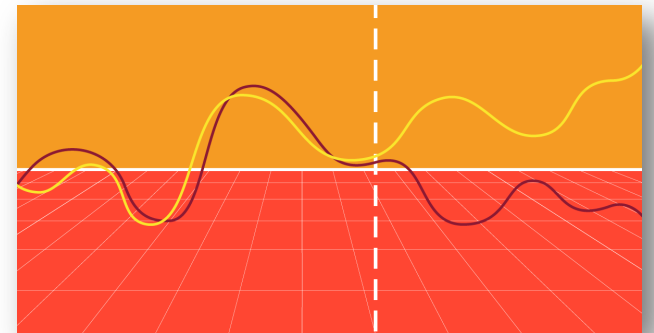
- **Data**

- Time series of daily visitor volumes across the 1,000 most popular websites on the internet from SimlarWeb.



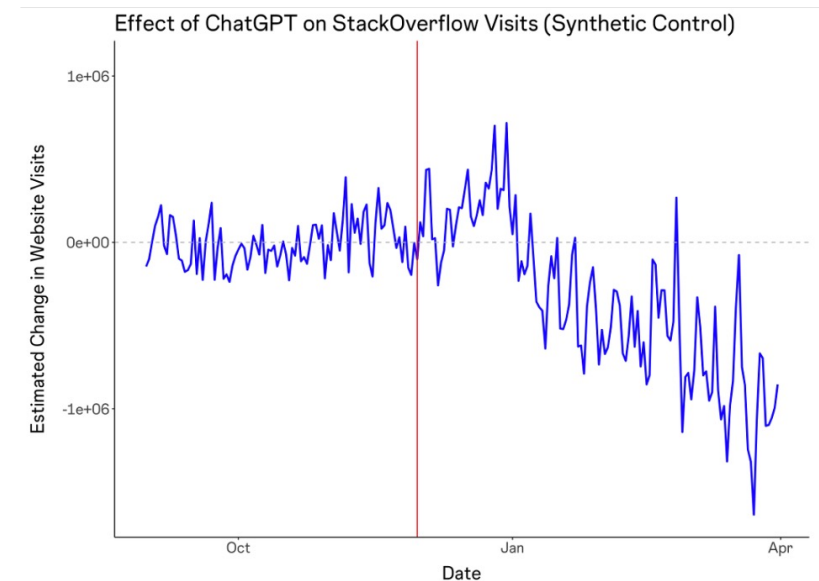
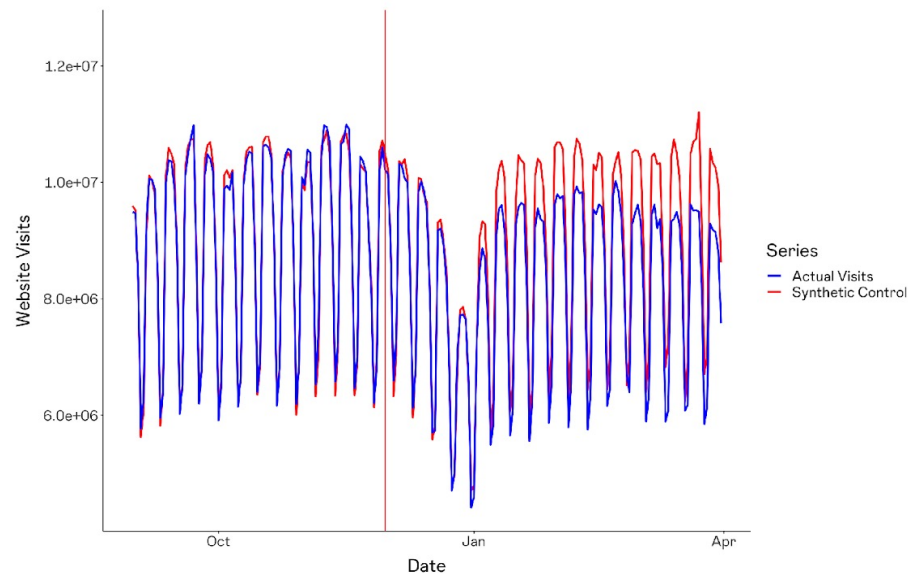
- **Method**

- Synthetic control using LASSO, or SCUL (Hollingsworth and Wing 2020; Greathouse 2022).
- The method yields a point estimate over time of the average treatment effect on the treated.



Effects on Participation: Web Traffic

We estimate that, as of March 31st, 2023, StackOverflow.com had lost ~1 million daily visits following ChatGPT's release (~15% of regular web traffic).



Effects on Participation: Question Volumes

- **Data:**

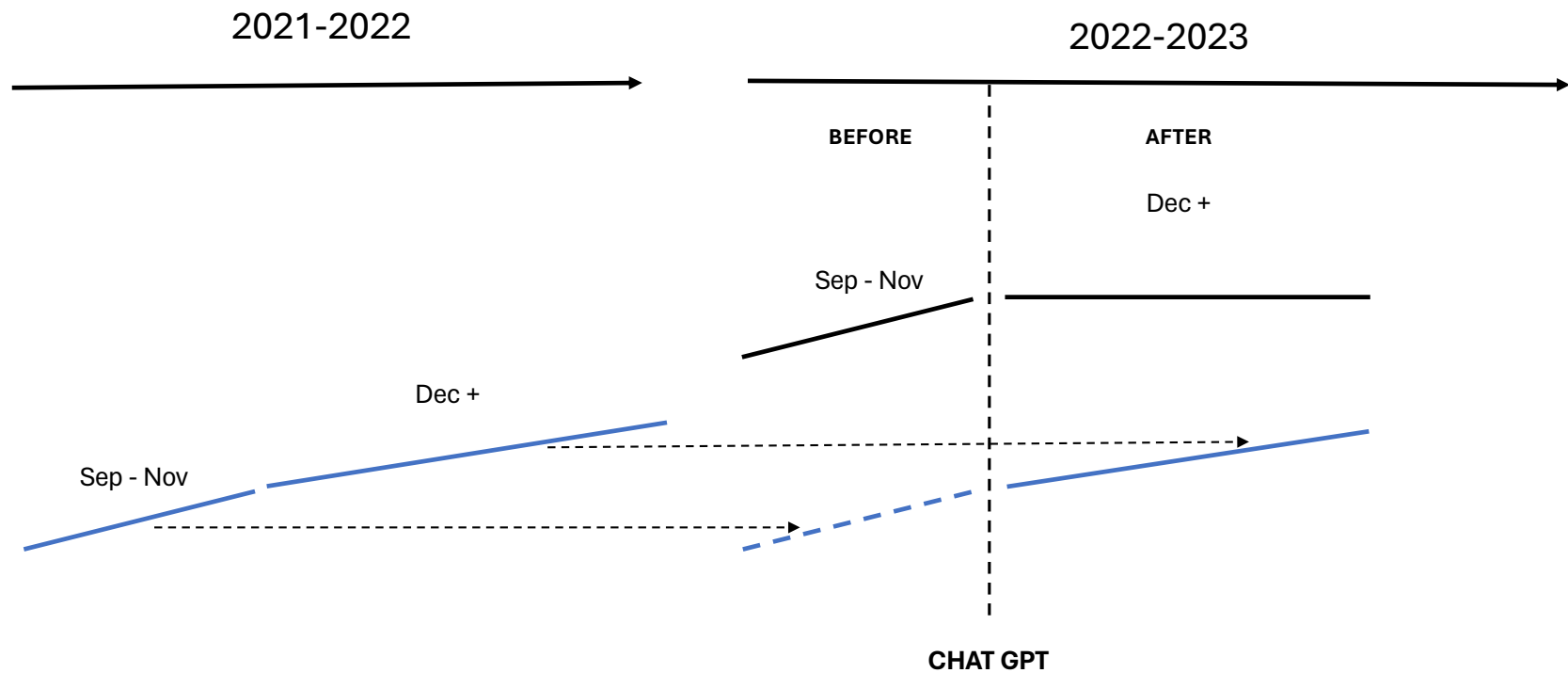
- We identify the top 50 most common topic-tags in our raw Stack Overflow question-data. We then construct counts of the number of questions posted per topic, per day, over our period of study.

- **Method:**

- We employ the approach of Goldberg et al. (2023) and Eichenbaum et al. (2023). We use measures of question volumes with a particular topic tag over an aligned set of calendar dates from one year-prior, as control.

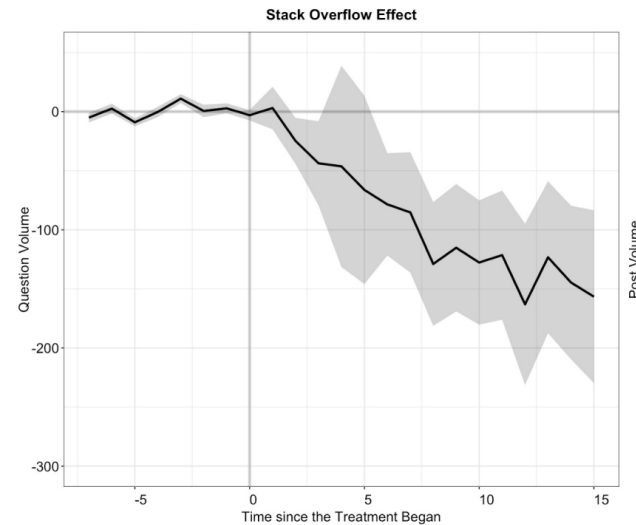
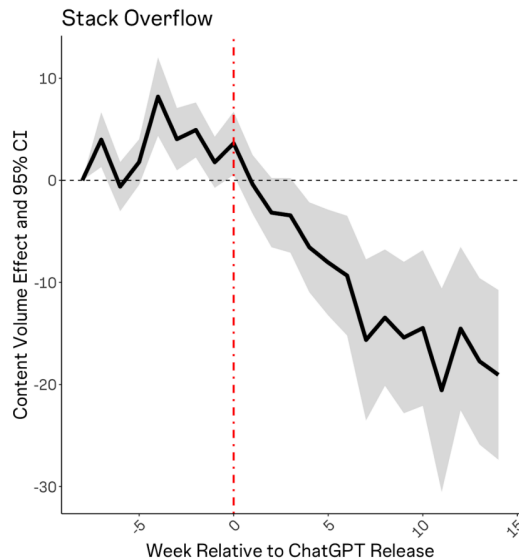
$$Volume_{i,c,t} = Treat_t + Treat_t \cdot RelWeek_c + Topic_i + W_c + e_{i,c,t}$$

Effects on Participation: Question Volumes



Effects on Participation: Question Volumes

For the top 50 topic tags on Stack Overflow, we observe a systematic decline in average question volumes over time. We see a ~15% decline in daily average question volumes; ~150 fewer questions per topic-week, on average.



Heterogeneous Effects: Which Topics?

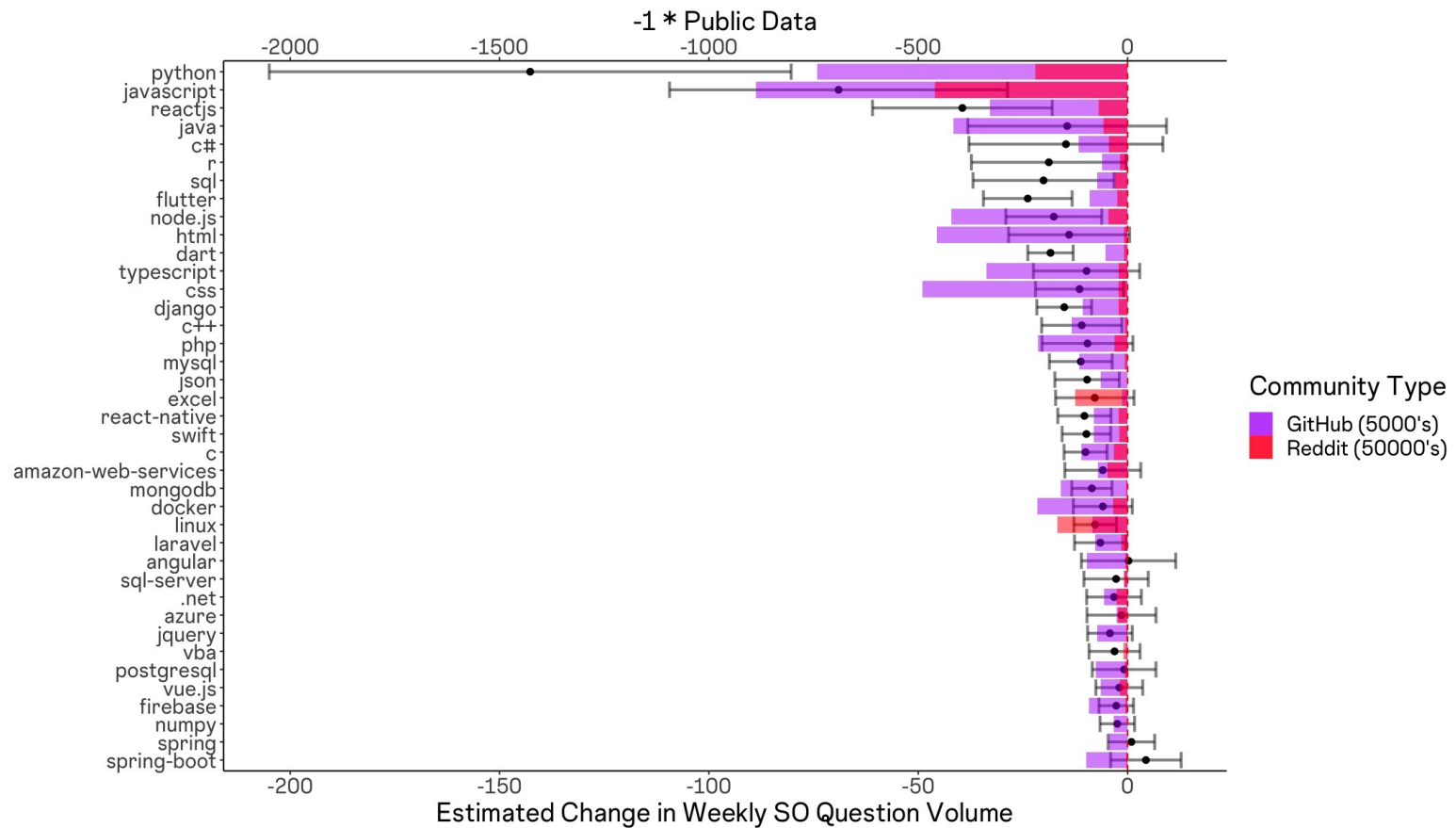
- **Data:**

- We focus on programming languages and software development frameworks.
- Same data as noted previously for this subset of areas.
- We collect additional data on the volume of GitHub repositories indicating usage of the relevant language / framework, and the total count of subscribers to relevant sub-Reddits.

- **Method:**

- We examine how our estimates of question declines vary across topics. We repeat our estimate of Equation (1), per topic.
- We correlate the magnitude of effects to the volume of publicly available data on that topic in GitHub repositories and Reddit activity.

Heterogeneous Effects: Which Topics?



Heterogeneous Effects: Which Communities?

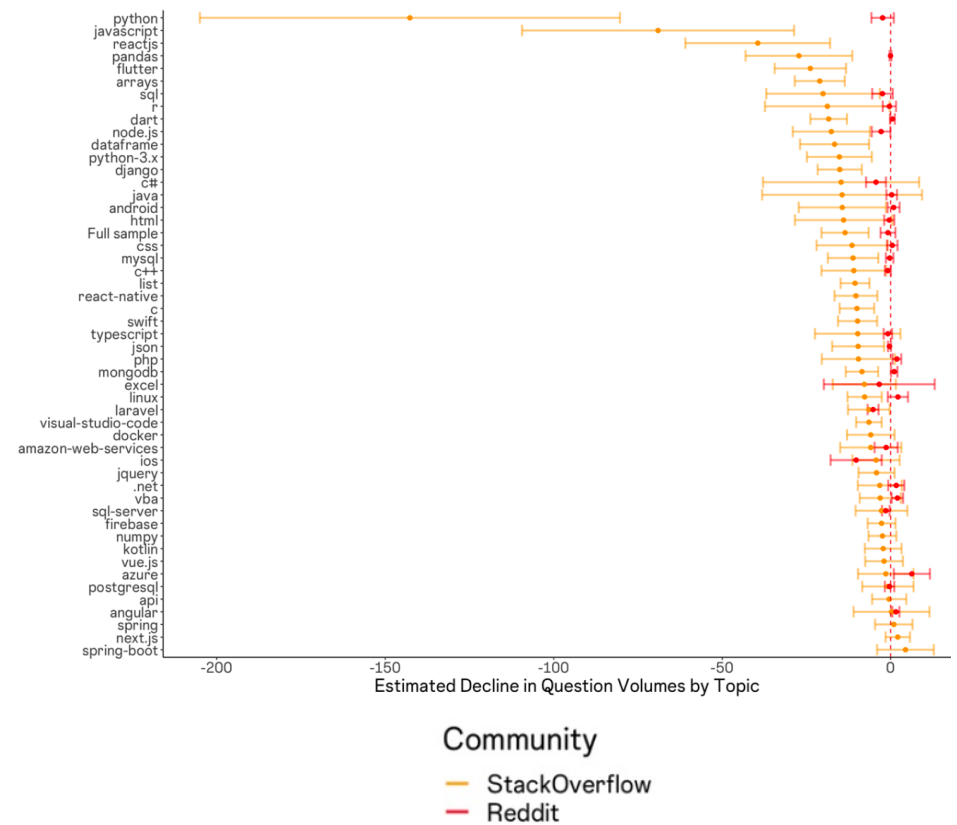
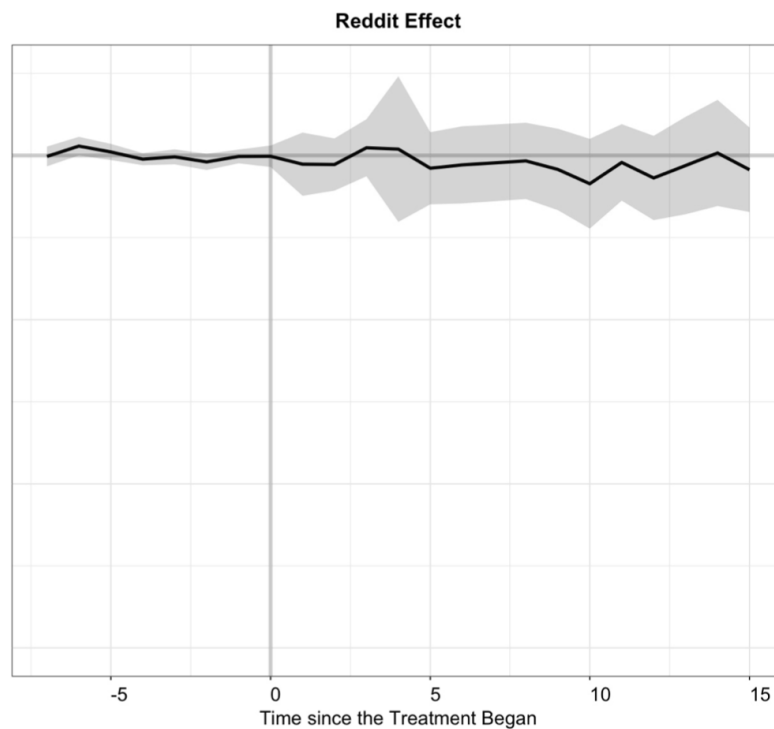
- **Data:**

- We consider posting activity at Reddit sub-communities focused on programming language / software development frameworks that appear in our Stack Overflow topics.
- We collect data on the volume of the posts per sub-Reddit and day from <https://subredditstats.com>

- **Method:**

- We estimate changes in posting volumes in these analogous sub-Reddits, over the same window of time, employing the same identification strategy.
- The rationale is that Reddit sub-communities are documented to be much more 'social' than Stack Overflow (Antelmi et al. 2023; Sengupta 2020).

Heterogeneous Effects: Which Communities?



Average Effects by Community

	Stack Overflow	Reddit
Treat	9.244*** (2.094)	1.830 (1.403)
Treat × Post	-13.554*** (3.117)	-0.701 (0.596)
Observations	16,100	8,316
R ² Adj.	0.908	0.844
FE: Topic	Yes	Yes
FE: Year-Week	Yes	Yes

Notes: (1) $+p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

(2) Standard errors in parentheses, clustered by topic.

(3) We are able to identify sub-reddits for approximately 50% of topics.

Table S1.

The average effect of ChatGPT's release on daily, per topic, question-posting volumes at Stack Overflow and daily per sub-reddit posting volumes at Reddit. We observe a statistically significant decline in daily question volumes per topic tag, on average, at Stack Overflow. By contrast, though we observe a negative coefficient in the case of Reddit, the estimate is not statistically significant. The estimate associated with Stack Overflow suggests an average daily decline of approximately 14 questions per topic.

Heterogeneous Effects: Which Users?

- **Data:**

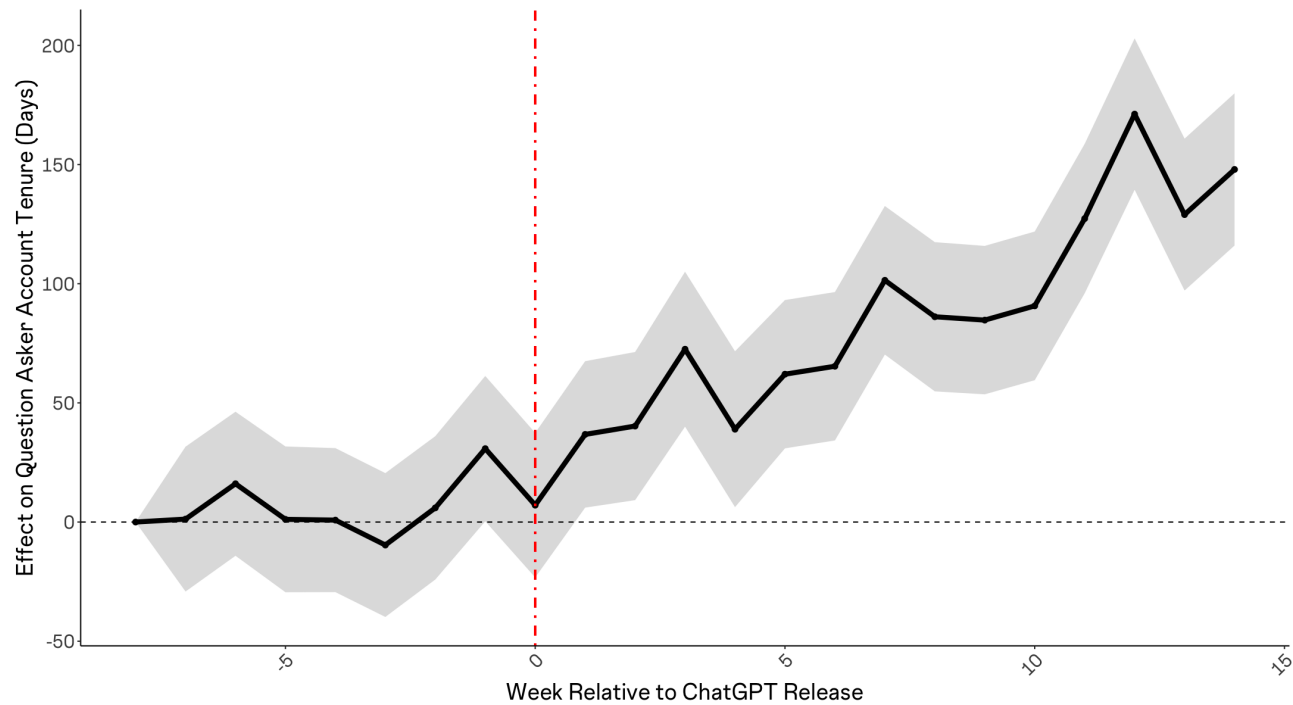
- We consider data at the question-level, taking the asking user's account tenure (in days) as our dependent variable.

- **Method:**

- We use the same strategy as before, and we estimate changes in the average tenure of users associated with posted questions at Stack Overflow.
- This is a question-level estimation, so we have no FEs in this specification.

$$Age_{c,t} = Treat_t + Treat_t \cdot RelWeek_c + W_c + e_{c,t}$$

Heterogeneous Effects: Which Users?



Heterogeneous Effects: Types of Knowledge Sought?

- **Data:**

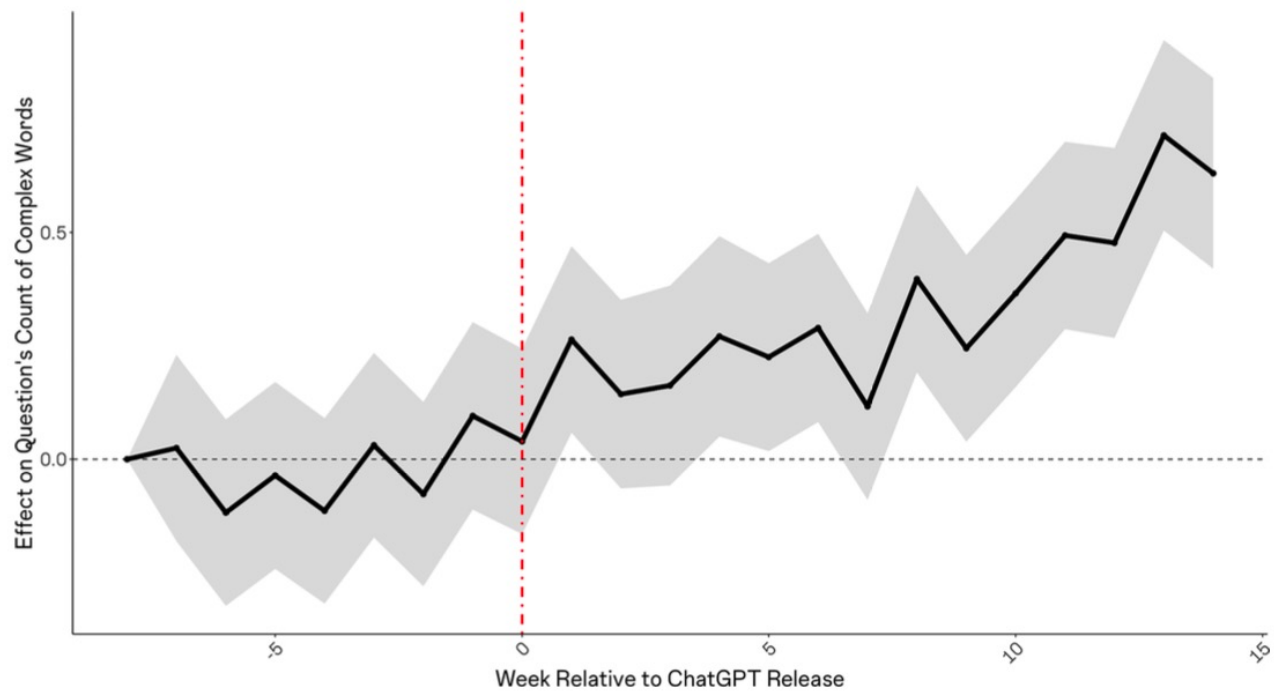
- Again, we consider data at the question-level, taking the volume of ‘big words’ in the Q as our dependent variable (6 letters or more). Proxy for complexity / sophistication.

- **Method:**

- We use the same strategy as before, and we estimate changes in the average count of big words associated with posted questions at Stack Overflow.
- This is a question-level estimation, so we have no FEs in this specification.

$$BigWords_{c,t} = Treat_t + Treat_t \cdot RelWeek_c + W_c + \epsilon_{c,t}$$

Question Effects: Types of Knowledge Sought



Second-Order Effects: Quality of Knowledge

- **Data:**

- We consider answer-level data, to understand changes in answer quality. Our measure of quality is a binary indicator of net up vs. down voting that an answer receives. We omit answers that attract no votes.

- **Method:**

- We estimate the relationship between our ChatGPT indicator and our binary indicator of net downvote, conditioning on the question being answered (question FE). If we add a user FE, effects will disappear entirely.

$$NetNegative_{q,c,t} = Treat_t + Treat_t \cdot Post_c + Q_q + W_c + e_{q,c,t}$$

$$NetNegative_{q,u,c,t} = Treat_t + Post_c + Treat_t \cdot Post_c + Q_q + W_c + \gamma_u + \mu_{q,u,c,t}$$

Second-Order Effects: Quality of Knowledge

Answers posted to StackOverflow, conditional on the question being answered, rise in their probability of being net-downvoted after ChatGPT emerges.

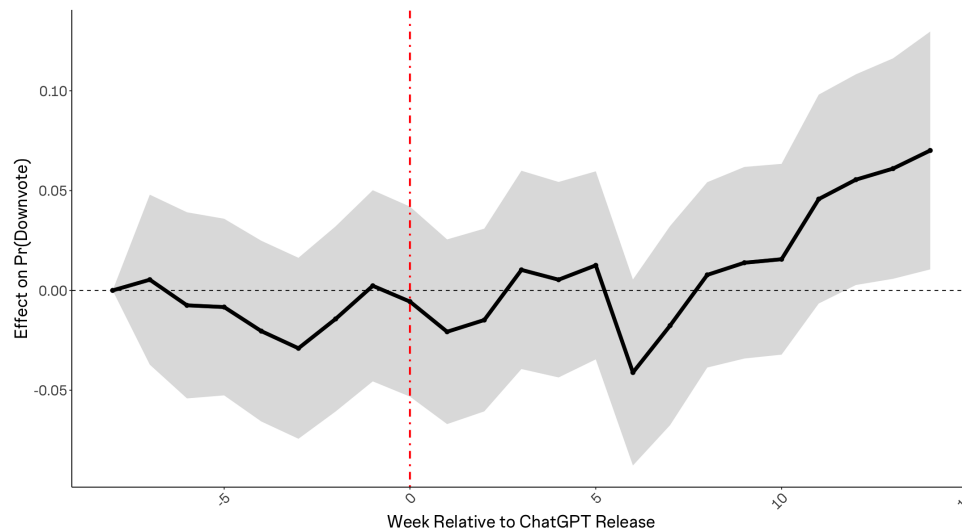
	(No User FE)	(User FE)
Treat	0.035*** (0.007)	0.015 (0.030)
Treat × Post	0.017* (0.009)	-0.005 (0.032)
Observations	710,509	710,509
R ₂	0.846	0.972
FE: Question	Yes	Yes
FE: Calendar Week	Yes	Yes
FE: User	No	Yes

Notes: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard errors in parentheses, clustered by question.

Second-Order Effects: Quality of Knowledge

Answers posted to StackOverflow, conditional on the question being answered, rise in their probability of being net-downvoted once ChatGPT emerges.



Summary & Implications

To what extent do users disengage from knowledge exchange?

Answer: Web traffic and question posting volumes decline markedly at Stack Overflow.

In what scenarios will those effects be largest?

Answer: Effects are largest on popular topics, they don't manifest on Reddit (social), and they are mostly driven by 'younger' / 'newer' users departing.

Are there second-order declines in quality of knowledge sharing?

Answer: Yes. We see a systematic rise in net-down-voting for Stack Overflow answers, which appears to be driven by a shrinking crowd.

What to Do? Design for Collaboration.

Design to Encourage Group and Team Usage.

[Decades of research and practice](#) speak to the importance of user interface design as a driver of how technology is used. A stark example of this contrast is apparent in how Open AI has provided users with access to ChatGPT, i.e., a single-user conversational interface, versus the way that Midjourney has deployed its generative image technology, i.e., a bot on Discord that is accessed within a chatroom. Whereas ChatGPTs interface encourages one-on-one usage, [Midjourney's interface encourages community](#) by having users engage with the technology in a chat-based group environment. New users can observe peers engaging with the tool, enabling them to connect with and learn from likeminded users. The above contrast suggests the value of designing the user interface of a generative AI tool to be oriented toward group and team usage, rather than individual usage.

What to Do? Target Social Contexts.

Deploy in Contexts Where Collaboration Is Already the Focus.

Beyond the interface to the technology itself, consider integrating the technology into venues where collaboration is already the focus. As a very simple example, LLMs can be integrated directly into online knowledge platforms like wikis and discussion forums. This is notably a step that Stack Overflow itself has recently taken, with its [announcement](#) of Overflow AI, an LLM trained on Stack Overflow's own data, which will be made available to users within the Stack Overflow community interface.

What to Do? Incentivize Connection.

Incentivize Employees to Use Gen AI for Collaboration.

[LinkedIn](#) has deployed an internal peer-recognition program that allows employees to celebrate peers within the organization for behaviors that encourage relationships and social connections. This has led greater worker productivity and improved employee retention. To counter social erosion that may occur with reliance on generative AI tools, create explicit incentives for employees, e.g., within broader employee recognition programs, that reward peer connection and desirable (collaborative) patterns of AI use. For example, employees can be rewarded for verifying and quality-assuring deliverables that peers may have produced with the support of generative AI tools. Recognizing employees for these sorts of activities has the dual benefit of preventing errors and hallucinations in AI output from creeping into workers' output, while also encouraging peer connection.

What to Do?

Embed Opportunities for Connection.

Supplement LLM Output with Peer Expert Recommendations.

Generative AI's output to a user can be supplemented to enable new connections with peers who can complement the AI-generated content. Consider supplementing output to include recommendations of peer experts within the firm who may have relevant knowledge, based on the user's initial prompt or query. Expert recommendations can be drawn from internal [expertise search tools](#) that many organizations now have in place.

Thank you.

gburtch@bu.edu

<https://gburtch.github.io>

