Online Appendix

Unraveling the Impact: An Empirical Investigation of ChatGPT's Exclusion from Stack Overflow

A Screenshots

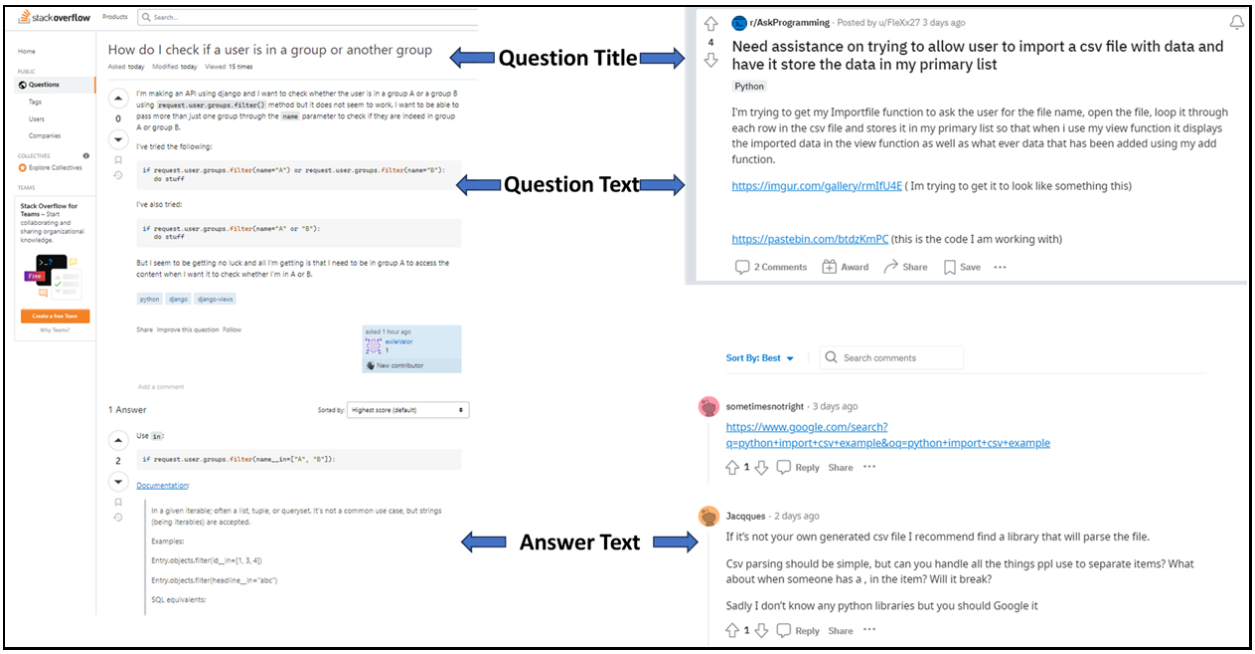


Figure A1: Screenshots of Stack Overflow (Left) and AskProgramming (Right)

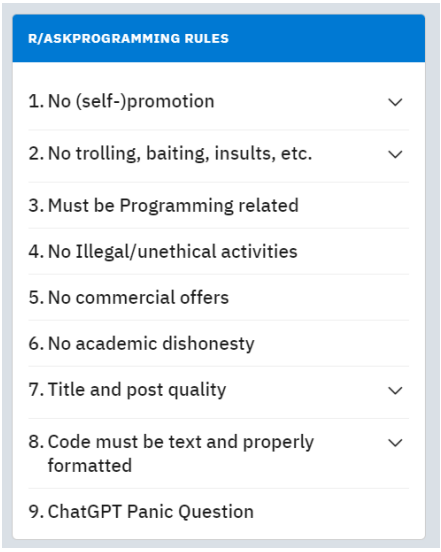
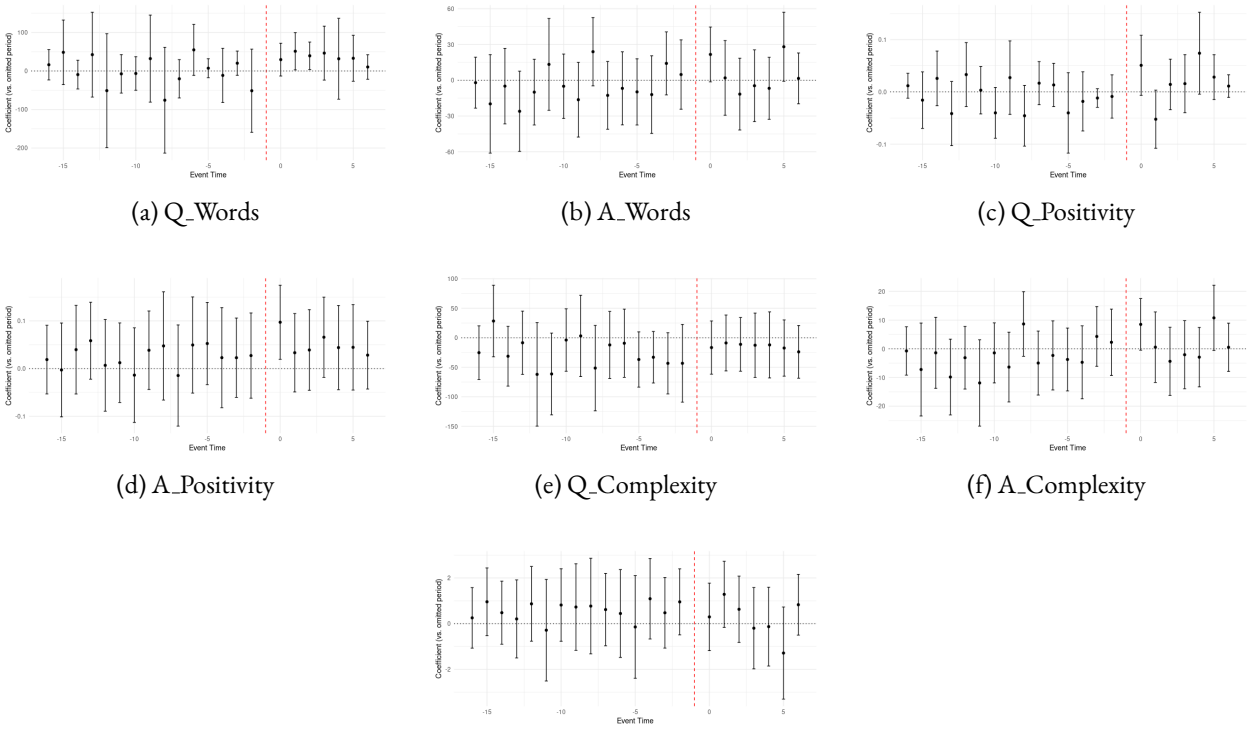


Figure A2: AskProgramming Submission Rules

B Parallel Trends



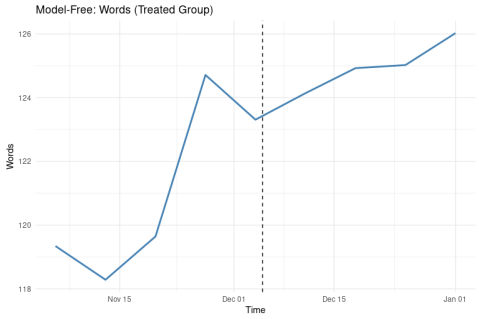
(a) Q_Words (b) A_Words (c) Q_Positivity

(d) A_Positivity (e) Q_Complexity (f) A_Complexity

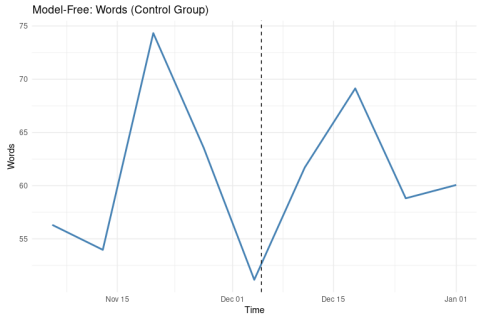
(g) A_Votes

Figure B3: Parallel Trends Across Outcome Variables

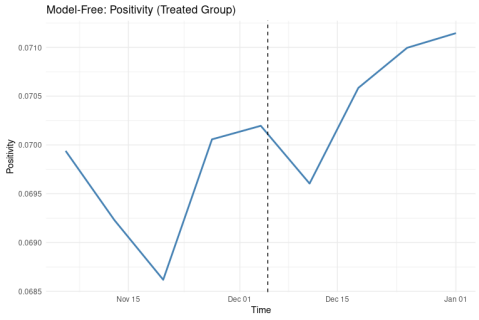
C Model Free Plots



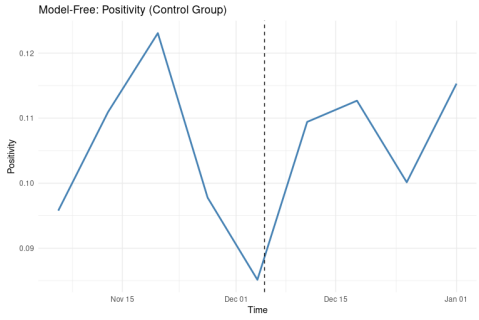
(a) Words - Treated



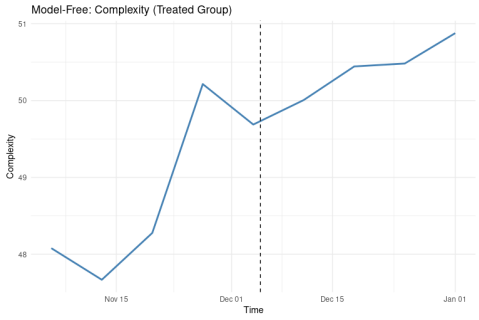
(b) Words - Control



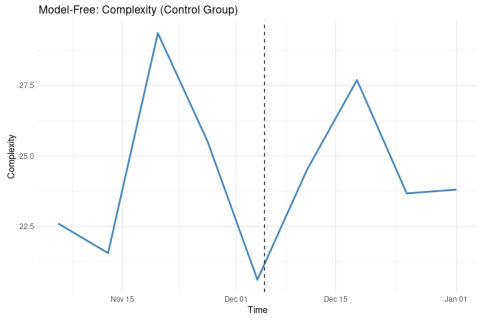
(c) Positivity - Treated



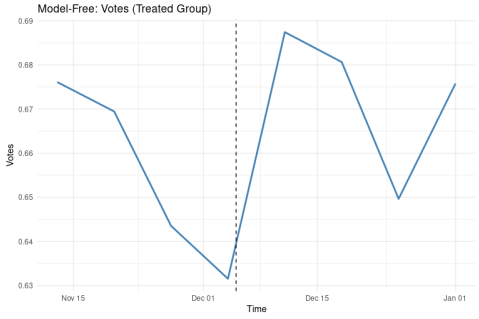
(d) Positivity - Control



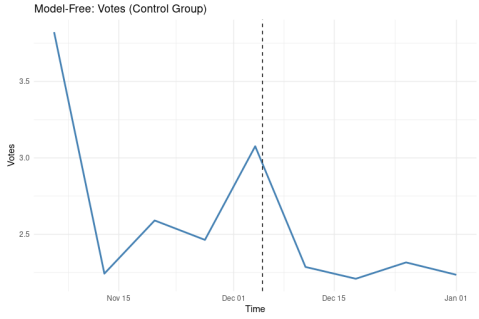
(e) Complexity - Treated



(f) Complexity - Control



(g) Votes - Treated



(h) Votes - Control

Figure C4: Model-Free Evidence for Treated and Control Groups Across Different Measures

D Result Tables

Table D1: Impact of User Credibility on Weekly Frequency

	Weekly Question Frequency	Weekly Answer Frequency
Treatment	-0.0072**	-0.0822**
	(0.0029)	(0.0406)
Reputation*Treatment	-0.0007	-0.0053***
	(0.0017)	(0.0010)
Fixed-Effects:		
Individual	Yes	Yes
Week	Yes	Yes
Observations	15,298,473	11,957,631
R^2	0.2608	0.7172

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust and clustered standard errors in parentheses.

Table D2: Robustness Results: Impact on Linguistic Characteristics 1

	Q_Words	A_Words	Q_Positivity	A_Positivity
Treatment	29.1955	0.3675***	0.0339	0.0266***
	(44.4681)	(0.1191)	(0.0931)	(0.0067)
Fixed-Effects:				
Individual	Yes	Yes	Yes	Yes
Day	Yes	Yes	Yes	Yes
Observations	2,068	877,220	2,068	877,220
R^2	0.6418	0.0618	0.4667	0.0912

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust and clustered standard errors in parentheses.

Table D3: Robustness Results: Impact on Linguistic Characteristics 2

	Q_Complexity	A_Complexity	A_Votes
Treatment	-1.2853	0.2362**	0.8073***
	(0.7826)	(0.1146)	(0.2299)
Fixed-Effects:			
Individual	Yes	Yes	Yes
Day	Yes	Yes	Yes
Observations	2,068	877,220	877,220
R^2	0.5463	0.0534	0.0756

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust and clustered standard errors in parentheses.

Table D4: Robustness Results - Impact on Weekly Frequency

	Weekly Questions	Weekly Answers
Treatment	-0.1066	-1.1999***
	(0.1057)	(0.1811)
Fixed-Effects:		
Individual	Yes	Yes
Week	Yes	Yes
Observations	4,255	143,463
R^2	0.7081	0.6636

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust and clustered standard errors in parentheses.

Table D5: Effect of Code Presence on Complexity and Votes

	A_Votes	A_Votes (Control)
Treatment	0.398** (0.177)	0.398** (0.177)
Code Presence		0.019*** (0.005)
Fixed-Effects:		
Individual	Yes	Yes
Day	Yes	Yes
Observations	541,897	541,897
R^2	0.4	0.4

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust and clustered standard errors in parentheses.

E Scenario-based Experiment

To further corroborate our findings and directly compare the effects across the two natural experiments, we conducted a scenario-based experiment. A scenario-based experiment is a research method where participants are presented with a realistic scenario and asked how they would respond or behave in that scenario. Because it allows controlled testing of specific variables while keeping the situation consistent for all participants, it has become a standard experimental approach to enhance the internal validity of observational studies in the recent IS literature (e.g., Liang et al., 2024; Schlackl et al., 2026; Zou et al., 2026).

In our context, the scenario-based experiment serves two specific purposes. First, it provides direct evidence that contributors genuinely adjust their writing behavior when AI-generated content is restricted by the platform, rather than the estimated improvement in our main analysis being solely a relative artifact of concurrent changes of the introduction of AI-generated content on the comparison platform. Second, we designed the experiment to test whether both the platform-specific ban (mirroring the Stack Overflow setting) and the full removal of access (mirroring the Italy setting) generate the same pattern of compensatory knowledge signaling in answer writing, or whether they operate through distinct mechanisms. The experiment includes two restriction scenarios designed to mirror our two natural experiments. The first corresponds to Stack Overflow's platform-level restriction, in

which AI-generated content is prohibited on a single platform while remaining available elsewhere. The second corresponds to Italy's nationwide restriction, in which ChatGPT is unavailable more broadly.

We collected responses from 440 participants on Prolific. Participants first read a scenario in which an online Q&A forum may either allow or prohibit the use of generative AI tools for content generation. To mirror the Italy natural experiment, half of the participants were asked to "Consider the same forum in both situations: 1) when the use of generative AI is allowed and 2) when it is not." To mirror the Stack Overflow natural experiment, the other half of participants were asked to "Consider the same forum in both situations: 1) When using generative AI is allowed on this forum (and on other platforms). 2) When using generative AI is NOT allowed on this forum, while it remains freely available and widely used on other platforms."

The participants in both scenarios were then asked to compare their behavior under the two policies. Specifically, they were asked to imagine answering a question on the forum after the policy change and indicate how it would affect the effort they invest, the level of friendliness of their response, and the depth of knowledge reflected in their answer. We use effort as a proxy for answer length, friendliness as a proxy for positive tone, and depth of knowledge as a proxy for answer complexity, as these constructs more naturally reflect how participants would describe their approach to responding. All questions used a 5-point Likert scale ranging from -2 (extremely decrease) to 2 (extremely increase), with a midpoint of 0 (no change). Participants were also given an opportunity to elaborate on their answers using an open-text format.

The results, presented in Table E1, show that across both scenarios, participants reported they would invest more effort in their responses, write answers that are more positive in tone and demonstrate greater depth of knowledge after the forum prohibits the use of generative AI ($p < 0.05$, one-tailed one-sample t-test). These findings align with our empirical results from both natural experiments, providing direct experimental support for the interpretation that contributors on restricted platforms genuinely alter their behavior in a manner consistent with compensatory knowledge signaling.

Table E1: Scenario Experiment Results: Participant Responses to Generative AI Forum Restrictions

Variable	Mean	Std	t-statistic	p-value
A_Words	0.62	0.801	16.249	$<$0.001
A_Positivity	0.38	0.748	10.585	$<$0.001
A_Complexity	0.11	0.870	2.577	0.005

We further tested whether the two scenarios, which mirror the two natural experiments, produce similar results using a two-sided independent sample t-test. The results are presented in Table E2 in which the "Effect" column summarizes whether the two restriction scenarios produce differences only in effect size ("Magnitude") or whether they differ in the direction or statistical significance of the observed effect. Consistent with the pattern observed across our natural experiments, we find no significant difference between scenarios on answer complexity and positivity. While effort showed a statistically significant difference, the difference was in magnitude rather than direction: participants in both conditions reported increased effort, but those in the Italy-mirroring condition reported a somewhat larger increase. This pattern suggests that whether AI is restricted on a single platform while remaining available elsewhere or removed entirely, the behavioral response of human contributors is qualitatively equivalent as one would expect based on our proposed mechanism of identity threat and compensatory knowledge

signaling.

Table E2: Comparison of Experiment Results Across Restriction Scenarios

Variable	Scenario	Mean	Std	t-statistic	p-value	Effect
A_Words	Platform Restriction Scenario	0.53	0.724	2.333	0.020	magnitude
	Full/Nationwide Restriction Scenario	0.71	0.864			
A_Positivity	Platform Restriction Scenario	0.35	0.655	0.765	0.222	none
	Full/Nationwide Restriction Scenario	0.40	0.830			
A_Complexity	Platform Restriction Scenario	0.10	0.737	0.274	0.392	none
	Full/Nationwide Restriction Scenario	0.12	0.986			

Finally, we conducted a thematic analysis of open-ended responses following the survey questions, with AI assistance (Claude, Anthropic) under human review, and with authorial responsibility for all final coding decisions. The most common themes reinforce the theoretical mechanisms proposed in our study. First, consistent with the knowledge signaling prediction, users reported investing more time to produce thoughtful, carefully researched answers (160 respondents or 36.4%). Second, respondents described making deliberate efforts to be kind, encouraging, and supportive, a tendency they noted as particularly salient in a human-only space (137 respondents or 31.1%). Third, providing direct evidence of AI identity threat, respondents asserted that their expertise is tool-independent, distancing their professional identity from AI (83 respondents or 18.9%).

Taken together, the scenario-based experiment strengthens the interpretation of our main findings in two ways. First, it provides direct evidence that the behavioral response we theorize, compensatory knowledge signaling in response to an AI restriction, operates as a genuine individual-level mechanism. Second, the equivalence of responses across the two scenario conditions addresses the concern regarding the differing natures of our two natural experiments. Despite the Stack Overflow setting involving a platform-specific ban near the introduction of ChatGPT while the Italy setting involving removal of an already-available tool, contributors exhibit qualitatively the same behavioral response in both contexts. This convergence supports the use of the Italy natural experiment as a valid complement to our main analysis.