ChatGPT Is A User-Generated Knowledge-Sharing Killer

Junzhi Xue
School of Management, xuejunzhi@mail.ustc.edu.cn

Lizheng Wang
University of Science and Technology of China, lzwang@mail.ustc.edu.cn

Jinyang Zheng
Purdue University, zhengjy@purdue.edu

yongjun li
University of Science and Technology of China, lionli@ustc.edu.cn

Yong Tan
University of Washington, ytan@uw.edu

ChatGPT Is A User-Generated Knowledge-Sharing Killer

Completed Research Paper

Junzhi Xue
University of Science and Technology of China
Hefei, Anhui, China
xuejunzhi@mail.ustc.edu.cn

Lizheng Wang
University of Science and Technology of China
Hefei, Anhui, China
lzwang@mail.ustc.edu.cn

Jinyang Zheng
Purdue University
West Lafayette, Indiana, United States
zhengjy@purdue.edu

Yongjun Li
University of Science and Technology of China
Hefei, Anhui, China
lionli@ustc.edu.cn

Yong Tan
University of Washington
Seattle, Washington, United States
ytan@uw.edu

Abstract

Large Language Models (LLMs), e.g., ChatGPT, is expected to reshape a broad spectrum of domains. This study examines the impact of ChatGPT on question aksing in Q&A communitits via the natural experiment. Safe-guided by supporting evidence of parallel trends, a difference-in-difference (DID) analysis suggests the launching trigger an average 2.6% reduction of question-asking on Stack Overflow, confirming a lower-search-cost-enabled substitution. Our further analysis suggests that, this substitution effect has resulted in more longer, less readable and less cognitive and hence more sophisticated questions on average. Finally, the insignificant change in the score given by viewers per question suggests no improvement in the question quality and decreased platform-wide engagement. Our moderation analysis further ascertain the types of individuals who are more susceptible to ChatGPT. Taken together, our paper suggests LLMs may threaten the survival of user-generated knowledge-sharing communities, which may further threaten the sustainable learning and long-run improvement of LLMs.

Keywords: ChatGPT, Q&A community, question-asking

Introduction

Groundbreaking development in the field of artificial intelligence (AI) based on large language models (LLMs) (Mitchell and Krakauer 2023) has given rise to a language generation tool such as ChatGPT. It revolutionized communication with machines, empowering individuals to interact with AI systems in natural language and execute complex tasks at low costs. Beyond its intelligent content output, powerful natural language understanding capabilities, and continuous learning and improvement (e.g., GPT-4) (H. Holden Thorp 2023; van Dis et al. 2023), attention and discussions have been given to how this innovative technology could disrupt or reshape various important economic sectors, such as education (Cotton et al. 2023), healthcare (Cascella et al. 2023), business (Dowling and Lucey 2023), scientific research (van Dis et al. 2023), technology (H. Holden Thorp 2023; van Dis et al. 2023), content production (Kar et al. 2023), etc.

Among the broad spectrum of domains, the relationship between ChatGPT and user-generated knowledge-sharing on platforms such as Stack Overflow and GitHub is perhaps of one of the greatest interests to researchers and practitioners (Meghmala 2023). While the fluent and clear Q&A of ChatGPT may substitute the user-generated Q&A via lowered search cost, thereby dampening the engagement on user platforms (Dave 2023), the cost saving by AI may allow users to engage in more creative activities and hence provide high-quality questions and answers. Also, the generative AI nature of ChatGPT implies assistance in content creation, leading to a surge in the quantity and quality of user-generated content (UGC) like the Q&A. Knowing this relationship would further imply LLMs' continued learning and future improvement, given the essential learning source role of the user-generated knowledge for the training of LLMs (Dave 2023).

This study investigates how ChatGPT's launching disrupts user-generated knowledge-sharing. Utilizing a full dataset from the user-generated Q&A community for programmers, Stack Overflow, spanning three months, both before and after the release of ChatGPT, and a same structured dataset but lagged by a year as control, this study employs the Difference-in-Difference (DID) method (Mitze et al. 2020) to quantify the quantitative and qualitative changes in the user-generated Q&A caused by the launching of ChatGPT. While the choices of control may render the design confounding concerns of other intertemporal policy changes correlated to the focal launching, we carefully examine the policy change history on Stack Overflow, make according empirical corrections, and test on the presence of remaining confounders via examining the violation of parallel trends in order to justify the design validity.

The findings suggest that the release of ChatGPT results in a reduction of question-asking among platform users by 2.63%, supporting the hypothesis that users are opting for ChatGPT as an alternate source of information- and knowledge-sharing. Qualitatively, while the length of asked questions, on average, increased by approximately 2.60%, the readability and level of cognition involved in the text reduced by 2.56% and 0.62%, respectively. To the disappointment of our hope, the subjective quality of questions raised, measured by a score function of up-votes and down-votes on the questions, does not change significantly, indicating that question raising does not benefit from the AI assistance as well as the search cost saved by ChatGPT.

The mechanism analyses further reveal that users qualitatively adjust their questions to be longer, less readable, and less cognitive questions when they do not change the number of questions asked. This rules out the alternative explanation as users more lessen the shorter but more readable and highly cognitive questions for the qualitative changes in the main results; instead, the saved search cost allows users to ask questions longer and in a manner that is hard for ChatGPT to process. Moreover, this study analyses the heterogeneity of the impact of ChatGPT on different user types, showing that new and low-reputation users experienced a greater increase in the complexity of questions and that LLMs exhibit stronger substitutability in terms of knowledge acquisition for them.

This study provides the first empirical evidence to quantitatively measure the impact of LLMs technology on knowledge-based UGC platforms, offering significant insights to relevant stakeholders.

On the practical front, the findings caution on the sustainability of user-generated know-sharing platforms (communities). Specially, we find there is no quality improvement in questions implied by the unchanged subjective quality. Such unchanged quality improvement together with decreased question number will hurt platform-wide engagement. This further cautions on the diminishing improvement and future learning of

LLMs because of a downward spiral triggered by diminishing user-generated learning sources.

Context and Data

We leverage the natural experiment of the release of ChatGPT on November 30, 2023, and treat the user-generated knowledge sharing on Stack Overflow as the objects (i.e., units) in this experiment. Stack Overflow is one of the largest and most representative online communities for programming Q&A exchanges and knowledge sharing. Thousands of questions and responses are generated in the community every day. As of March 2022, the community has over 20 million registered users, over 24 million questions and 35 million answers (Wikipedia contributors 2023). Registered users are free to ask programming-related questions and volunteer to answer other users' questions.

To measure user-generated knowledge sharing on Stack Overflow, we collect a full sample of question-level data from the community's quarterly dump. This dump covers all user-generated content, shown as a snapshot of the main relevant data in the community, including anonymized users' profiles, posts, votes, etc. The posts include questions, answers, tag wiki, etc., among which questions and answers account for the vast majority. The votes contain viewers' every record of upvote and downvote. The data utilized in this study is the newest release, which was made available in March 2023 (Stack Exchange Community 2023).

We focus on the data associated with questions in the dump contains detailed records of all undeleted questions from July 31, 2008, to March 4, 2023[1], including *time stamp*, *owner user ID*, *title*, *body*, *tags*, *upvotes*, *downvotes*. An example of a question is shown in Figure 1. We then organize this data as panel data, defining each day as a time and each question class in Stack Overflow's question classification scheme as a cross-sectional unit. Stack Overflow uses the tags for question classifications, and a question may simultaneously relate to multiple tags. We choose the first tag of each question, usually the primary one, as the classification standard. Doing so avoids duplicated counting of question numbers - the total number of questions across different cross-sectional units equals the total number of questions on the platform, which aligns our result interpretation to the platform-wide change. Further, we notice that some tags can be very rare, which will result in some very sparse cross-sectional units. We, therefore, combine this part of tags into a new class named 'others'. In the final dataset, 5% of questions fall in this 'others' class.

To reflect the quantitative change of user-generated knowledge sharing, we measure *the number of questions* for each class and period. For the qualitative change operationalization, we extract all text strings from the body of a question and calculate question-level characteristics. The first characteristic is *the length of question*, which is the word count of the question body. In general, if a question takes more words to explain, it could contain more details and complexity. The same logic applies to *the number of tags*, so be our second characteristic. Third, among many popular NLP approaches to text processing, we chose two metrics that are both simple to calculate and relevant to our research context: *SMOG* and *Cognition*. The SMOG measures the readability of the question content, a question with a higher SMOG is more difficult to understand and requires a longer education experience (Deng et al. 2022); The Cognition, calculated based on Linguistic Inquiry and Word Count (LIWC), reflects the degree of cognition efforts put into constructing the question (Tausczik and Pennebaker 2010). Finally, we introduce an external metric: *Score*, defined as the difference between upvotes and downvotes. As the upvotes and downvotes reflect the subjective quality assessment of peer users, this variable serves as a subjective measure of question quality. A question with a high score should gain the approval of other users in the community. A statistical description of pre-described variables is shown in Table 1.

Another concerned policy potentially related to our research question is that Stack Overflow banned the use of ChatGPT for answering questions on December 5 and removed suspected ChatGPT-generated answers. This ban, however, is not a reversal of the launch of ChatGPT we are interested in. In specific, it couldn't prohibit the use of Chat-GPT beyond Stack Overflow, thereby blocking the potential substitution. It could neither stop users from leveraging ChatGPT to compose a question. Also, knowing that our control group also has no use of ChatGPT for answering questions, any treatment effect, if identified, shall not attribute to this ban.

[1]Community has policies to remove questions that don't qualify (Stack Overflow 2023).

Variable	Sample size	Mean	SD	Min	Max
Questions	234033	6.36	36.84	0	920
Length	1487880	89.59	63.08	0	4154
SMOG	1487880	8.02	4.26	0.00	53.50
Tags	1487880	3.04	1.24	1	5
Cognition	1487880	15.22	5.65	0.00	100.00
Score	1487880	0.24	1.25	-17	328

Table 1. Variable statistics

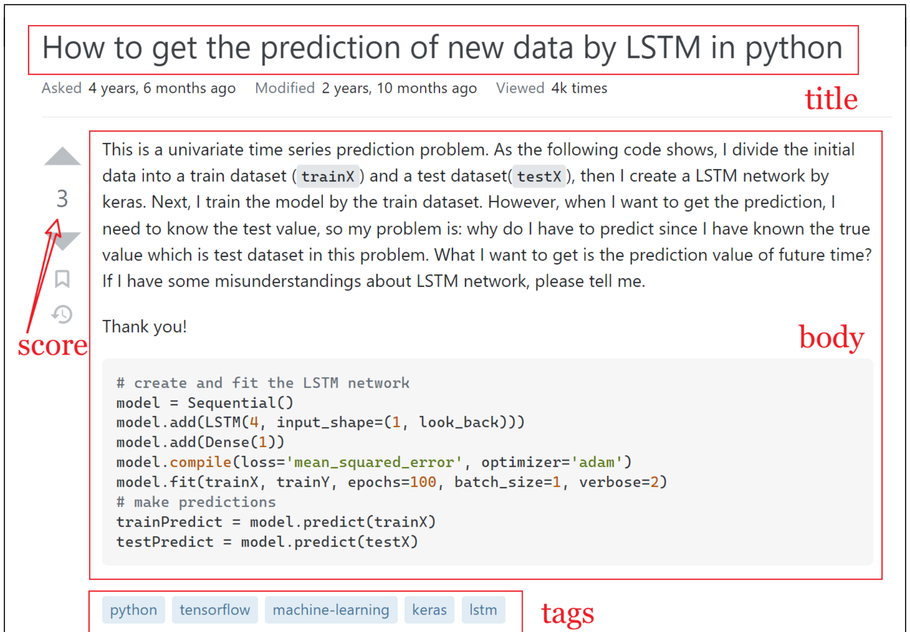

Figure 1. Components of A Question

Model-free Evidence

In the left top of Figure 2, we compare the difference in the number of daily questions asked between the two groups, using September 1, 2021, to February 28, 2022, as the control group and September 2022 to February 28, 2023, as the treatment group. Notably, a dip in the number of queries can be observed towards the end of December for both groups, which is attributed to the Christmas holiday period. It is also worth mentioning that a cyclical pattern is discernible in the graph, reflecting the difference between weekdays and weekends. It is evident that the introduction of ChatGPT had a discernible impact on the volume of new questions submitted by users, with a notable decrease in this metric compared to pre-deployment levels. However, this change was not replicated in the control group, indicating that it was indeed ChatGPT that influenced the observed decrease in question-asking.

Similar temporal descriptions of the mean level of question characteristics are shown in the rest of Figure 2. We noticed that question characteristics exhibit similar temporal trends before the release of ChatGPT, except that, there is a rapid growth in the mean value of Length, SMOG, Tags and Cognition after October 25, 2022. Such a steep increase can be attributed to the 'Ask Wizard' functionality implemented in the whole platform. The 'Ask Wizard' feature contains a series of prompts to help users better describe their questions (Stack Overflow 2022). This feature is mandatory for first-time users who ask questions and is available as an option for users who have asked questions before. We have controlled the impact of this functionality in our models below, and the validation of the parallel trend before the release of ChatGPT discussed on the robustness check demonstrates the static effect of this functionality, which further helps to distinguish the change of question-asking after the release of ChatGPT when controlling the impact of 'Ask Wizard'.

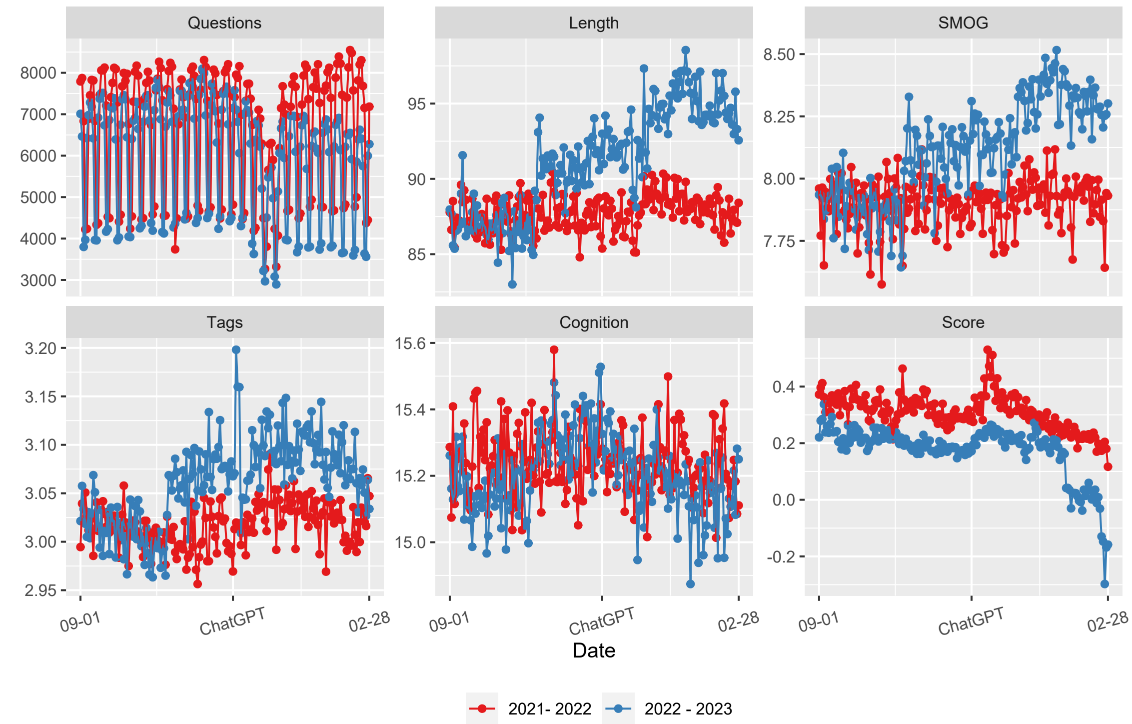

Figure 2. Comparation of Questions Asked Between 2021-09-01~2022-02-28 and 2022-09-01~2023-02-28

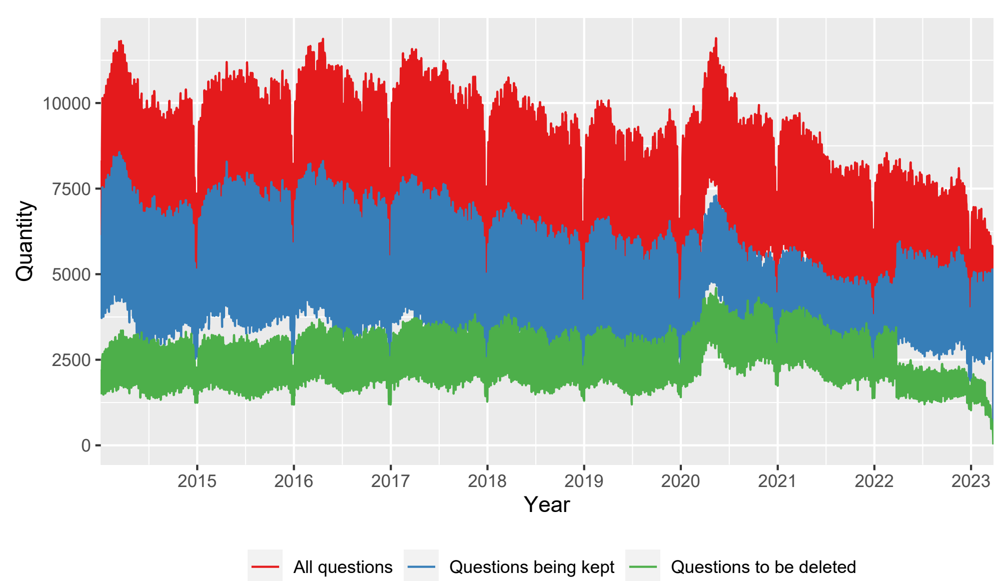

Figure 3. Daily Number of Questions Since 2014

Empirical Design and Results

Given that this natural experiment suffers from the absence of a proper control group during the same time window, as the release of ChatGPT can affect every user and Q&A task, as well as the motivation in the model-free evidence, we use the same data during the same time window of the last year as the control group (Wang and Overby 2022). These two time windows where our panel data are extracted from are then aligned as parallel, and the pre- or post-treatment time are defined as times before or after November 30 in both windows (i.e., years). Because of the exactly same cross-sectional units and similar natural seasonality patterns, these lagged data could serve as an exchangeable control group under the condition of an absence of other time-specific exogenous shocks that largely impact the content generation. This condition, however, is likely to hold according to the historical major changes on Stack Overflow, as shown in Figure 3, if we observe for a longer period of time, we can see that the total number of questions rebounded after the Christmas holiday each year but declined after the Christmas holiday in 2022, hence justifying our model using one-year lagged data as control group and demonstrating the impact of ChatGPT.

The Number of Questions

The availability of a control group enables a classical Difference-in-Difference design to identify the impact of ChatGPT on the number of questions asked, after teasing out persistent confounders and time-specific confounders. Following the standard practice, we specify the following equation:

$$\log(1 + \text{Questions}_{it}) = \beta_1 \text{Treat}_i \times \text{After}_t + \text{AskWizard}_t + u_i + T_t + d_{it} + h_{it} + \epsilon_{it}, \tag{1}$$

where Questions_{it} is the number of questions asked under tag i on date t, Treat_i equals 1 if tag i belongs to treatment group and 0 otherwise, After_t equals 1 if date t exceeds November 30 and 0 otherwise. We take the logarithm form of the dependent variable to address any skewness issues, so the coefficient of the interacted term β_1 captures the percentage impact of ChatGPT on the number of questions. Further, we add AskWizard_t to our model, it equals 1 if date t is greater than or equal to October 25, 2022, and 0 otherwise. Finally, u_i and T_t are tag and date level fixed effects respectively, d_{it} controls day of week effects, and h_{it} controls the holiday effect.

	log(1+Questions)
After × Treat	-0.0267*** (0.0056)
Day of week FE	Yes
Day FE	Yes
Tag FE	Yes
Observations	234,033
Adjusted R^2	0.84481
Clustered (Tag) standard-errors in parentheses; FE: fixed effects	
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05	

Table 2. The Change of Question Quantities

Result: Table 2 shows that the coefficient of interaction is significantly negative (-0.0267, p < 0.001), that is, the release of ChatGPT has significantly reduced the number of newly posted questions by 2.6% ($100 \times (e^{-0.0267} - 1)$%). The results suggest a decline in users' tendency to post questions on knowledge-based UGC platforms. This raises a new research question: what kinds of questions on the platform have been reduced? A plausible hypothesis is that ChatGPT could act as an alternative source of knowledge for individuals seeking straightforward and general question's answers. To validate this claim, we explore the change in question characteristics asked after the release of ChatGPT.

The Characteristics of Questions

The substitution by ChatGPT, implied by the lessened questions raised on Stack Overflow, is not necessarily a threat to the sustainability of knowledge-sharing communities and hence LLMs. The saved search cost may reallocate to asking a smaller set of questions that is more engaging and of higher quality. Also, AI assistance may facilitate the composition of high-quality questions, which drives up the user-engagement and participation. The increased engagement per question may offset the engagement loss due to fewer questions, and the quality improvement can benefit LLMs' future learning. To provide a full-spectrum picture of the ChatGPT release's consequences and hence test the above hypothetical expectation, we expand to examine qualitative changes in the questions caused by ChatGPT release. Following a similar identifying strategy, we specify:

$$\text{Characteristic}_{ijt} = \beta_1 \text{Treat}_{ij} \times \text{After}_t + \beta_2 \text{AskWizard}_t + u_i + T_t + d_{it} + h_{it} + \epsilon_{ijt}, \tag{2}$$

where $\text{Characteristic}_{ijt}$ is a proxy of the characteristic of question j asked under tag i on date t. Treat_{ij} equals 1 if question j under tag i belongs to treatment group and 0 otherwise. The rest variables keep in line with the definitions in Equation 1.

Result: In Table 3, we find that ChatGPT has significantly increased the length of questions by 2.63%, decreased the readability of questions by 2.60% and decreased the level of cognition by 0.62%. Such results suggest that ChatGPT has stimulated users to ask longer, less readable and less cognitive questions, in other words, those shorter, more readable and more cognitive questions which are convenient to handle by ChatGPT, are replaced by ChatGPT, and ChatGPT has enabled users to allocate additional time and effort towards asking more complex and in-depth questions.

Further, we investigate the change in the number of tags, and results suggest no significant difference in the number of tags after the release of ChatGPT, one possible explanation is that user questions are often focused, and a small number of tags is enough to locate a question, and there is a community limit on the maximum number of tags a question can have (≤ 5). Finally, the subjective quality of questions created, measured by a score function of up-votes and down-votes on the questions, does not change significantly, indicating that question raising does not benefit from the AI assistance as well as the search cost saved by ChatGPT.

	log(1+Length)	log(1+SMOG)	log(1+Cognition)	log(1+Tags)	Score
After $\times$ Treat	0.0263***	0.0259***	-0.0062**	0.0011	-0.0159
	(0.0033)	(0.0050)	(0.0020)	(0.0021)	(0.0134)
Day of week FE	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes
Tag FE	Yes	Yes	Yes	Yes	Yes
Observations	1,487,880	1,487,880	1,487,880	1,487,880	1,487,880
Adjusted R^2	0.01777	0.01178	0.00728	0.15108	0.01997
Clustered (Tag) standard-errors in parentheses; FE: fixed effects					
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05					

Table 3. The Change of Question Body Characteristics

Noteworthy, the estimated impact of ChatGPT on the platform represents a conservative lower bound, as not all users may be aware of its release or possess the proficiency to effectively utilize its capabilities. As the diffusion of LLMs technology progresses, the impact on the platform is likely to expand gradually. Additionally, the statistically significant positive impact of the release of ChatGPT on the length of questions suggests that users are turning to ChatGPT as an alternative source of knowledge for generic and simple questions, rather than posting those questions on knowledge-based UGC platforms. It is expected that platforms will react proactively to this emerging challenge, as they seek to safeguard their interests, thereby creating a more complex and dynamic future that is difficult to predict with certainty.

Mechanism Analysis

One plausible explanation behind our main findings on the qualitative changes in questions asked is the direct qualitative change in formulating new questions. Such an improvement is economically plausible, as the substitution of ChatGPT implies search and time cost savings. Such savings may spill over to empower users to formulate more sophisticated, and more advanced questions. Intuitively, ChatGPT's ability to provide quick and immediate answers may free up users' time, enabling them to invest more in formulating more sophisticated questions.

The identification of this mechanism is equivalent to testing the remaining qualitative change caused by ChatGPT's launch after teasing out the substitution's effect. This adjustment can be achieved following the backdoor criteria (Pearl 2009), through blocking the causal path via the quantity change (reduction) that reflects the substitution, and can be implemented via the following equation:

$$\text{Characteristic}_{ijt} = \beta_1 \text{Treat}_{ij} \times \text{After}_t + \beta_2 \text{Questions}_{it} + \beta_3 \text{AskWizard}_t + u_i + T_t + d_{it} + h_{it} + \epsilon_{ijt}, \quad (3)$$

where Questions_{it} is the number of questions asked under tag i on day t. The interpretation of the parameter of interest is now updated as the average qualitative change when the quantity of questions is controlled/unchanged, which specifically attributes to the complexity and nature change of the questions that is not due to the quantity reduction (and hence substitution).

Results in Table 4 show that, users still tend to ask longer, less readable, and less cognitive questions when controlling the reduction of question quantity. These findings suggest the presence of a direct qualitative change in question asking is not via the reduction in question quantity (i.e., substitution mechanism). The presence of a direct qualitative change further implies the externality of ChatGPT's saved search and time cost on the qualitative change of questions. While LLMs may substitute human question answering, they may free the time for humans to formulate intricate and challenging questions on human Q&A.

	log(1+Length)	log(1+SMOG)	log(1+Cognition)	log(1+Tags)	Score
Questions	$3.71\times 10^{-5**}$	3.05×10^{-5}	-7.34×10^{-6}	4.17×10^{-6}	2.62×10^{-5}
	(1.17×10^{-5})	(1.72×10^{-5})	(7.73×10^{-6})	(7.59×10^{-6})	(4.23×10^{-5})
After × Treat	0.0273^{***}	0.0267^{***}	-0.0064^{***}	0.0012	-0.0154
	(0.0032)	(0.0051)	(0.0019)	(0.0021)	(0.0134)
Day of week FE	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes
Tag FE	Yes	Yes	Yes	Yes	Yes
Observations	1,487,880	1,487,880	1,487,880	1,487,880	1,487,880
Adjusted R^2	0.01778	0.01178	0.00693	0.15108	0.01997
Clustered (Tag) standard-errors in parentheses; FE: fixed effects					
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05					

Table 4. Mechanism Analysis Results

Heterogeneity Analysis

The above analysis confirms that LLM-based tools have a substitution effect on knowledge-sharing platforms for low-complexity problems, as evidenced by a decrease in user activity on UGC platforms. However, is this substitution effect different across users, and what types of users are more susceptible to the impact of LLMs from the perspective of the primary stakeholders - the UGC platform?

	log(1+Questions)	log(1+Length)	log(1+SMOG)	log(1+Cognition)
After × Treat × New	-0.0011	0.1741^{***}	0.2083^{***}	0.0269^{***}
	(0.0052)	(0.0067)	(0.0083)	(0.0034)
Day of week FE	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes
Tag FE	Yes	Yes	Yes	Yes
New FE	Yes			
Observations	471,686	1,468,948	1,468,948	1,468,948
Adjusted R^2	0.81225	0.02301	0.01571	0.00777
Clustered (Tag) standard-errors in parentheses; FE: fixed effects				
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05				

Table 5. The Heterogeneity Effects of ChatGPT on New and Experienced Users

New vs. Experienced Users

In this section, we classified users into two categories based on the considerations of platform switching costs (Uncapher and Wagner 2018): new users versus experienced users. The former represents users with lower switching costs or weaker social ties, while the latter represents users with higher switching costs or stronger social ties. Users registered within one year before the latest date of the treatment group and the control group are new users, otherwise, they are experienced users, that's to say, users in the treatment group whose registration date is less than February 28, 2022, are experienced users, while those in the control group are February 28, 2021. We use dummy variable New_k to indicate whether user k is a new user (New_k = 1) or not (New_k = 0).

First, we explore the heterogeneous effect of ChatGPT on the number of questions asked by new users and experienced users. Following the standard practice of Difference-in-Difference, we split the daily number of questions under tag i into two groups: asked by new users and asked by experienced users, and run a Difference-in-Difference-in-Difference (DDD) model, i.e.,

$$\begin{aligned} \log(1 + \text{Questions}_{ikt}) = \alpha_k + \beta_1 \text{Treat}_i \times \text{After}_t + \beta_2 \text{New}_k \times \text{Treat}_i \times \text{After}_t + \\ \beta_3 \text{New}_k \times \text{Treat}_i + \beta_4 \text{New}_k \times \text{After}_t + \text{AskWizard}_{it} + u_i + T_t + d_{it} + h_{it} + \epsilon_{ikt}. \quad (4) \end{aligned}$$

where Questions_{ikt} is the daily number of questions under tag i asked by group k, New_k is a dummy variable denoting whether group k is the new user group(one if yes, zero otherwise), α_k is the fix effects of new user group (k = 1) and experienced (k = 0). The key variable we care about is the coefficient of the triple interaction term β_2, which represents the difference in the impact of ChatGPT on new and old users. As shown in Table 4, ChatGPT does not affect new and old users significantly differently in the number of questions asked.

Second, we explore the heterogeneous effect of ChatGPT on the characteristics of questions asked, specially,

$$\begin{aligned} \text{Characteristic}_{ijt} = \beta_1 \text{Treat}_{ij} \times \text{After}_t + \beta_2 \text{New}_j \times \text{Treat}_{ij} \times \text{After}_t + \\ \beta_3 \text{New}_j \times \text{Treat}_i + \beta_4 \text{New}_j \times \text{After}_t + \beta_6 \text{AskWizard}_t + u_i + T_t + d_{it} + h_{it} + \epsilon_{ijt}, \quad (5) \end{aligned}$$

where New_j is a dummy variable indicating whether question j is asked by a new user (one if yes, zero otherwise). As shown in Table 5, New users are more likely to be influenced by ChatGPT to ask longer and less readable questions on the community than experienced users when considering the complexity of the question's body text. One possible explanation is that new users generally possess less programming experience than their experienced counterparts, which renders ChatGPT more effective at solving low-complexity questions of new users, and those who continue to ask questions in the community may be running into sophisticated questions.

	log(1 + Questions)	log(1+Length)	log(1+SMOG)	log(1+Cognition)
After × Treat × HighRepu	-0.0085	-0.2049***	-0.2034***	-0.0533***
	(0.0050)	(0.0090)	(0.0103)	(0.0029)
Day of week FE	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes
Tag FE	Yes	Yes	Yes	Yes
HighRepu	Yes			
Observations	471,686	1,471,288	1,471,288	1,471,288
Adjusted R^2	0.81630	0.02298	0.01533	0.00818
Clustered (Tag) standard-errors in parentheses				
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05				

Table 6. The Heterogeneity Effects of ChatGPT on High-reputation and Low-reputation Users

High-reputation vs. Low-reputation Users

To further explore the heterogeneous effect of ChatGPT, we classify users into high-reputation users and low-reputation users based on their relative level of reputation. Users earn their reputation points by the peer-users' upvotes to their questions and answers. In other words, a user with a higher reputation usually produces more questions and answers from the perspective of question quantity and produces higher-quality questions and answers from the perspective of the nature of posts. Practically, a user in treatment (control) group is seen as a high-reputation user if his/her reputation at the end of treatment (control) horizon exceeds the mean level. Let HighRepu_k indicate whether group k is the high-reputation group or not (one if yes, zero otherwise), and HighRepu_j denotes whether question j is asked by a high-reputation user (one if yes, zero otherwise). By substituting New_k in Equation (4) by HighRepu_k and New_j in Equation (5) by HighRepu_j, we can gain insights of the heterogeneous effect of ChatGPT on high-reputation users and low-reputation users.

Column (2) reported in Table 6 suggest that ChatGPT does not affect high-reputation and low-reputation users differently in the number of question asked. As for question characteristics, columns (3)-(5) in Table 6 indicate that questions asked by users with high reputation are less affected by ChatGPT compared with questions asked by users with low reputation. The explanation for such results is similar to the explanation for the heterogeneous effect on new and experienced users, that is, users with low reputaion are usually asking low-complexity questions which might be answered by ChatGPT, and consequently, those questions asked by low-reputation users and still shown in the platform are more likely to be sophisticated questions.

Robustness Check

Parallel Trend

A standard DID approach requires the parallel trend before treatment, that is, control and treatment groups have parallel trends in their number of questions and question's content features. Following the standard practice in literature, the parallel trend can be tested by the relative time model. Specially, for the dependent variable being the number of questions, the relative time model can be formulated as

$$\log(1+\text{Questions}_{it}) = \beta_1 \text{Treat}_i \times \text{After}_t + \beta_2 \times \text{AskWizard}_t + \sum_{\tau=-2}^{-16} \gamma_\tau \text{Pre}_{it}(\tau) + u_i + T_t + d_{it} + h_{it} + \epsilon_{it}, \quad (6)$$

where $\text{Pre}_{it}(\tau)$ is a dummy variable, which represents whether day t is $|\tau|$ days before the treatment. For instance, Preo_{it}(-2) represents whether day t is two days before the treatment. Following the practice in

(Deng et al. 2022; Guan et al. 2023), we use dummy variable Pre_{it}(-16) to represent pre-treatment periods that are greater than or equal to 16 days prior to treatment. We omit Pre_{it}(-1) as baseline, which is the day before ChatGPT was released, If γ_τ equals 0 significantly for τ = -2, ..., -16, then there's no significant difference between treatment and control group before the release of ChatGPT, the parallel trend is satisfied. For outcome variables being question's body features, the form of regression is very similar,

$$\text{Characteristic}_{ijt} = \beta_1 \text{Treat}_{ij} \times \text{After}_t + \beta_2 \text{AskWizard}_t + \sum_{\tau=-2}^{-16} \gamma_\tau \text{Pre}_{it}(\tau) + u_i + T_t + d_{it} + h_{it} + \epsilon_{ijt}. \quad (7)$$

We report the results of all parallel trend tests in Table 7. It suggests that, prior to the release of ChatGPT, there was no substantial difference between the treatment group and the control group in terms of both the number of questions and question's content features, suggesting that the treatment and control groups were comparable before the intervention.

	log(1+Questions)	log(1+Length)	log(1+SMOG)	log(1+Cognition)	log(1+Tags)	Score
Pre(-16)	0.0086 (0.0234)	-0.0233 (0.0132)	-0.0362 (0.0213)	0.0122 (0.0086)	-0.0060 (0.0062)	-0.0236 (0.0225)
Pre(-15)	-0.0085 (0.0332)	-0.0028 (0.0161)	-0.0380 (0.0257)	0.0150 (0.0117)	-0.0046 (0.0083)	0.0204 (0.0312)
Pre(-14)	0.0192 (0.0325)	-0.0403* (0.0179)	-0.0293 (0.0288)	0.0105 (0.0119)	-0.0089 (0.0075)	-0.0259 (0.0327)
Pre(-13)	0.0582 (0.0335)	-0.0269 (0.0178)	-0.0248 (0.0264)	0.0149 (0.0118)	-0.0103 (0.0082)	-0.0207 (0.0381)
Pre(-12)	0.0323 (0.0331)	-0.0184 (0.0189)	-0.0702* (0.0274)	0.0190 (0.0120)	-0.0035 (0.0074)	0.0408 (0.0270)
Pre(-11)	0.0292 (0.0316)	-0.0243 (0.0218)	-0.0312 (0.0314)	0.0232 (0.0130)	-9.33×10^{-5}(0.0096)	0.0280 (0.0294)
Pre(-10)	0.0128 (0.0309)	0.0134 (0.0193)	-0.0262 (0.0334)	0.0256 (0.0144)	0.0128 (0.0096)	0.0658 (0.0402)
Pre(-9)	0.0352 (0.0309)	-0.0122 (0.0219)	-0.0400 (0.0356)	0.0025 (0.0131)	-0.0147 (0.0102)	0.0195 (0.0384)
Pre(-8)	-0.0004 (0.0322)	-0.0254 (0.0180)	-0.0351 (0.0243)	0.0250* (0.0116)	-0.0099 (0.0081)	0.0355 (0.0303)
Pre(-7)	-0.0270 (0.0329)	-0.0106 (0.0163)	-0.0081 (0.0286)	0.0315** (0.0118)	-0.0037 (0.0082)	0.0186 (0.0333)
Pre(-6)	-0.0214 (0.0338)	-0.0072 (0.0175)	-0.0321 (0.0290)	0.0123 (0.0124)	0.0017 (0.0089)	0.0003 (0.0308)
Pre(-5)	-0.0255 (0.0338)	-0.0226 (0.0211)	-0.0295 (0.0315)	0.0167 (0.0130)	-9.98×10^{-5}(0.0086)	-0.0128 (0.0381)
Pre(-4)	0.0293 (0.0309)	0.0146 (0.0182)	-0.0272 (0.0236)	0.0229 (0.0134)	-0.0015 (0.0082)	-0.0004 (0.0252)
Pre(-3)	0.0236 (0.0300)	-0.0305 (0.0211)	-0.0649 (0.0388)	0.0095 (0.0142)	-0.0048 (0.0104)	0.0230 (0.0364)
Pre(-2)	0.0033 (0.0316)	-0.0034 (0.0218)	-0.0352 (0.0308)	0.0185 (0.0133)	-0.0076 (0.0089)	-0.0173 (0.0378)
Pre(-1)	baseline					
Day of week FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Tag FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	234,033	1,487,880	1,487,880	1,487,880	1,487,880	1,487,880
Adjusted R^2	0.84481	0.01777	0.01178	0.00693	0.15108	0.01997
Clustered (Tag) standard-errors in parentheses; FE: fixed effects						
Signif. Codes: ***: 0.001, **: 0.01, *: 0.05						

Table 7. Parallel Trend Test Results

Questions Deleted

Another potential concern is that the decline in questions is due to an increase in the number of deleted questions, as the data used in the main analysis contains only the qualified questions. Those identified as unqualified questions by the platform's cleaning mechanism will be removed. Luckily, Stack Overflow has opened an interface on the internet (community wiki 2022), through which, those deleted questions can be queried, and the deletion date can be accessed, but key variables are missing such as *owner user ID* and *body*. With this data, we can at least test if the decline in questions is caused by an increase in the number of deleted questions, as the data used in our main analysis contains only the qualified questions which are not removed automatically by the platform's cleaning mechanism. Specifically, we collect additional data on the number of deleted questions provided by Stack Overflow and add them to the number of questions in the main analysis. We then re-perform the DID model defined in Equation (1). The results in Table 8 suggest that the negative effect of ChatGPT on question-asking holds and, hence, that the analysis on the quantitative changes of questions is robust to the inclusion of deleted questions.

Dependent Variable:	log(1+Questions)
After × Treat	-0.0419*** (0.0057)
Day of week FE	Yes
Day FE	Yes
TagFE	Yes
Observations	260,821
Adjusted R^2	0.84861
Clustered (Class) standard-errors in parentheses	
*Signif. Codes: ***: 0.001, **: 0.01, *: 0.05*	

Table 8. The Change of Question Quantities When including deleted questions

Discussion

This article examines the impact of LLMs on knowledge-oriented UGC platforms by analyzing changes in individual behavior in an online Q&A community before and after the introduction of ChatGPT. Our empirical results yield from the DID framework show that the release of ChatGPT has a negative impact on individual question-posting behavior on the knowledge-based UGC platform. The negative impact can be attributed to the fact that ChatGPT's responses to simple inquiries may have displaced individual's tendency to ask questions online. Our empirical findings provide support for this hypothesis, as we observed that individuals tend to pose longer and more complex questions, resulting in a 2.63% increase in question length, a 2.60% decrease in readability and a 0.62% decrease in cognition. This phenomenon is likely due to ChatGPT's ability to serve as an alternative source of knowledge, capable of satisfying individuals with low-complexity questions. However, it is important to note that ChatGPT has developmental limitations, particularly in its capacity to comprehend complex problems, which creates a disparity between its abilities and those of professional technical personnel. Nonetheless, this gap is expected to gradually narrow over time, and we can anticipate that the impact of LLMs will expand progressively. Besides, we find that the quality of questions measured by the score of viewers has no significant change after the release of ChatGPT.

Further, Our study has unveiled that the impact of LLMs is contingent upon user types. Specifically, our empirical results demonstrate that newly registered and low-reputation users are particularly susceptible to the effects of LLMs on the complexity of questions. This finding suggests that the adoption of LLMs enables users to allocate additional time and effort toward asking more complex and in-depth questions.

Finally, the impact of ChatGPT on relevant stakeholders has been notable, particularly in terms of the welfare of individuals, UGC platforms and ChatGPT itself. For individuals, ChatGPT has been a game-changer, providing access to knowledge at lower time costs, which has translated into a significant welfare improvement. The positive externalities of this phenomenon are also noteworthy, as individuals can now focus on more complex problem-solving approaches, contributing to overall social welfare. Although AI has the potential to bring about positive social impacts, it also entails certain drawbacks. One of the most notable ones is the uneven distribution of access to AI opportunities and proficiency levels among individuals, which can exacerbate existing societal inequalities. For UGC platforms, according to the conclusions of this paper, the number of questions has decreased, but the quality of questions has not increased, suggesting that UGC platforms are at some risk of being replaced. For ChatGPT itself, failure to improve the overall quality of the questions may also affect the sustainable learning of ChatGPT in the future, thus limiting its own development and long-run improvement.

Despite concerns about the potential negative effects of LLMs on UGC platforms, there are also new opportunities to be explored. Rather than simply prohibiting the use of LLM-based tools, the focus should be on leveraging these powerful tools to produce highly customized and interactive content for users. At the same time, it's crucial to ensure that the information generated by LLMs is accurate and unbiased. The key lies in adopting policies that balance the advantages of LLM-based tools with the need to maintain the trust and engagement of users.

Limitations and Future Research Directions

Firstly, the conclusions of this paper are limited to the impact of LLMs on Q&A communities and are not applicable to the wider digital content industry, especially the entertainment-oriented digital content field. This also motivates researchers to explore more widely the impact of LLMs on the digital industry, in order to better respond to the challenges and opportunities brought by this new far-reaching technology. Secondly, this study is limited to measuring the impact of LLMs technology on Q&A communities from the perspective of communities. It cannot observe from a micro perspective whether users are using LLMs and how they switch between LLMs and Q&A communities. Future research will include individual-level adoption of LLMs information into observations to explain how individuals optimize their behavior through LLMs-based applications to achieve productivity improvement.

References

Cascella, M., Montomoli, J., Bellini, V., and Bignami, E. 2023. "Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios". *Journal of Medical Systems* (47:1), p. 33.

community wiki 2022. *Database Schema Documentation for the Public Data Dump and SEDE*. URL: https://meta.stackexchange.com/questions/2677/database-schema-documentation-for-the-public-data-dump-and-sede/2678#2678 (visited on May 4, 2023).

Cotton, D. R. E., Cotton, P. A., and Shipway, J. R. 2023. "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT". *Innovations in Education and Teaching International* (0:0), pp. 1–12.

Dave, P. 2023. *Stack Overflow Will Charge AI Giants for Training Data*. URL: https://www.wired.com/story/stack-overflow-will-charge-ai-giants-for-training-data/ (visited on May 4, 2023).

Deng, Y., Zheng, J., Khern-am-nuai, W., and Kannan, K. 2022. "More than the Quantity: The Value of Editorial Reviews for a User-Generated Content Platform". *Management Science* (68:9), pp. 6865–6888.

Dowling, M. and Lucey, B. 2023. "ChatGPT for (Finance) Research: The Bananarama Conjecture ". *Finance Research Letters* (53), p. 103662.

Guan, Y., Tan, Y., Wei, Q., and Chen, G. 2023. "When Images Backfire: The Effect of Customer-Generated Images on Product Rating Dynamics". *Information Systems Research* (), isre.2023.1201.

H. Holden Thorp 2023. "ChatGPT Is Fun, but Not an Author". *Science* (379:6630), pp. 313–313.

Kar, A. et al. 2023. ""So What If ChatGPT Wrote It?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy". *International Journal of Information Management* (71), p. 102642.

Meghmala 2023. *Will ChatGPT be the Stack Overflow Killer?* URL: https://techcabal.com/2023/01/31/stack-overflow-chat-gpt/ (visited on May 4, 2023).

Mitchell, M. and Krakauer, D. C. 2023. "The Debate over Understanding in AI's Large Language Models". *Proceedings of the National Academy of Sciences* (120:13), e2215907120.

Mitze, T., Kosfeld, R., Rode, J., and Wälde, K. 2020. "Face Masks Considerably Reduce COVID-19 Cases in Germany". *Proceedings of the National Academy of Sciences* (117:51), pp. 32293–32301.

Pearl, J. 2009. *Causality*. Cambridge university press.

Stack Exchange Community 2023. *dataStack Exchange Data Dump*. URL: https://archive.org/details/stackexchange (visited on Mar. 9, 2023).

Stack Overflow 2022. *What is the Ask Wizard?* URL: https://stackoverflow.com/help/askwizard (visited on May 4, 2023).

Stack Overflow 2023. *The Community User Deleted My Question! What Gives?* URL: https://stackoverflow.com/help/auto-deleted-questions (visited on Mar. 24, 2023).

Tausczik, Y. R. and Pennebaker, J. W. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods". *Journal of Language and Social Psychology* (29:1), pp. 24–54.

Uncapher, M. R. and Wagner, A. D. 2018. "Minds and Brains of Media Multitaskers: Current Findings and Future Directions". *Proceedings of the National Academy of Sciences* (115:40), pp. 9889–9896.

Van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., and Bockting, C. L. 2023. "ChatGPT: Five Priorities for Research". *Nature* (614:7947), pp. 224–226.

Wang, H. and Overby, E. M. 2022. "How Does Online Lending Influence Bankruptcy Filings?" *Management Science* (68:5), pp. 3309–3329.

Wikipedia contributors 2023. *Stack Overflow — Wikipedia, The Free Encyclopedia*. URL: https://en.wikipedia.org/w/index.php?title=Stack_Overflow&oldid=1151449153 (visited on May 4, 2023).