https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Articles in Advance, pp. 1–21
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Examining the Impact of Generative AI on Users' Voluntary Knowledge Contribution: Evidence from A Natural Experiment on Stack Overflow

Guohou Shan,[a] Liangfei Qiu[b,*]

[a] D'Amore-McKim School of Business, Northeastern University, Boston, Massachusetts 02115; [b] Warrington College of Business, University of Florida, Gainesville, Florida 32611

*Corresponding author

Contact: g.shan@northeastern.edu, https://orcid.org/0000-0002-3268-6348 (GS); liangfei.qiu@warrington.ufl.edu, https://orcid.org/0000-0002-8771-9389 (LQ)

Received: May 30, 2023
Revised: May 15, 2024; December 29, 2024; June 22, 2025
Accepted: August 26, 2025
Published Online in Articles in Advance: September 24, 2025

https://doi.org/10.1287/isre.2023.0332

Copyright: © 2025 INFORMS

Abstract. Voluntary knowledge contribution on online platforms holds significant value for users, platforms, and firms. Rapid advancements in generative artificial intelligence (AI) techniques have facilitated the automatic generation of knowledge on question-and-answer (Q&A) platforms. However, the impact of generative AI on users' voluntary knowledge contributions remains an empirical question. On the one hand, users may learn from generative AI, improving their answers by providing more organized and logical responses. On the other hand, generative AI can produce fabricated answers, and the accelerated pace of responding with AI assistance may impose additional cognitive burdens for comprehending the outputs, potentially reducing overall contributions. Our study examines the effects of generative AI, specifically ChatGPT, on users' voluntary knowledge contributions on Stack Overflow, one of the largest Q&A platforms. Utilizing a natural experiment, we employ difference-in-differences (DID) estimation to investigate the effects of generative AI on both the quantity and quality of user contributions, measured by the number of answers generated per day, answer length, and readability. Our findings reveal that the use of generative AI correlates with an increased number of answers generated by users, and these answers tend to be shorter in length and easier to read. We further explore the moderating effects of cumulative usage and usage intensity on the impacts of generative AI to test the mechanisms of learning and cognitive load. Our results indicate that users are learning from generative AI, enabling them to answer more questions while producing shorter and more digestible responses. Conversely, the additional cognitive burden associated with intensive AI usage negatively affects its impact on answer quantity. The implications of this study are both theoretical and practical. Theoretically, we contribute to the Information Systems (IS) literature by examining the influence of generative AI on users' voluntary knowledge contributions within the context of Q&A platforms. Practically, our findings provide platform owners and managers with insights into how generative AI affects users' knowledge contribution behavior, guiding decision-making and strategic development for integrating generative AI into their platforms.

History: Wonseok Oh, Senior Editor; Khim Yong Goh, Associate Editor.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2023.0332.

Keywords: knowledge contribution • generative AI • voluntary Q&A platforms • answer quantity • answer quality • difference-in-differences

1. Introduction

Voluntary knowledge contribution on online platforms has become increasingly important for users, platforms, and firms. However, the 90-9-1 rule highlights a key challenge: Only 1% of users actively create content, 9% of users respond to others' generated content without creating it, and 90% of users are just lurkers (Elder 2025). Addressing this disparity and these challenges, recent advances in generative AI—notably large language models (LLMs)—offer promising avenues for enhancing content creation. Among these tools, ChatGPT, developed by OpenAI and released in November 2022, reached 100 million users just two months after its launch and secured a $10 billion investment for further development (Nerdynav 2023). Built on LLMs like GPT-3.5 and refined using reinforcement learning from human feedback, ChatGPT is capable of mimicking human behavior and answering questions in a conversational manner (OpenAI 2022).

Despite the increasing prevalence of these tools, their impact on voluntary knowledge contribution behavior and outcomes—particularly on Q&A platforms—

remains unexplored. On the one hand, generative AI can produce highly accurate answers, supported through continuous training on diverse data and refinement based on human feedback (Abdullah 2023). Related studies have found that generative AI tools can generate organized, concise responses and effective summaries of complex information (Guo et al. 2023, Tian et al. 2024). Moreover, interacting with these tools can help users improve their own responses in terms of generating succinct code and delivering concise yet digestible explanations. Jia et al. (2024) found that AI can enhance an individual's ability to answer questions, particularly by handling repetitive questions, allowing humans to focus on generating more organized and readable responses. Therefore, the use of generative AI in generating answers may improve the *quality* of users' voluntary knowledge contributions (i.e., *contribution quality*).

Furthermore, generative AI tools like ChatGPT can help users contribute a large volume of voluntary knowledge (e.g., providing answers) because AI acts as an effective assistant that enhances individuals' abilities to respond to questions (Jia et al. 2024). Built on large language models (see, e.g., OpenAI 2022), generative AI can automate the answer-generation process (Guo et al. 2023). Users who utilize ChatGPT could learn and improve their ability to generate more answers by automating this process, as suggested by social learning theory (Bandura and Walters 1977). Additionally, the learning science literature highlights ChatGPT as an effective educational tool (see, e.g., Baker and Reimann 2024). Users leveraging ChatGPT to generate answers can learn how to answer questions effectively and expand their domain knowledge, as noted by Zhang et al. (2025). This type of learning enables users to answer more questions, thereby increasing their answer quantity. Consequently, the use of generative AI may also boost the quantity of users' voluntary knowledge contributions (i.e., *contribution quantity*).

However, there are also potential downsides to tool use. Some research has found that intensively interacting with technology can increase cognitive load, taxing users' working memory (Paas et al. 2003), suggesting that intensively using generative AI tools can increase cognitive load because of the additional time and effort required to comprehend the content generated by these tools. As users invest mental energy in understanding and adapting AI-generated content to ensure its accuracy and relevance to the specific question, their available mental resources for generating a large number of answers are reduced. Likewise, recent studies have revealed that generative AI systems can exhibit institutionalizing bias, produce hallucinations, and exhibit limited interpretability (see, e.g., Susarla et al. 2023). Users of generative AI tools may inadvertently compromise the quality of the answers they generate (e.g., being unorganized, lengthy, and harder to read) by introducing new biases from generative AI tools. In short, it is possible that the use of generative AI may lead to a decline in either or both the *quality* and *quantity* of voluntary contributions.

Understanding the potential implications of employing generative AI to answer questions on users' behavior within Q&A platforms is paramount, given its ramifications for users, platforms, and firms (see, e.g., Huang et al. 2022, Wang et al. 2022). Although previous studies have examined different factors that drive users to contribute to knowledge sharing, the impact of generative AI on users' answering behavior on Q&A platforms remains unclear. Therefore, we aim to address the following research question:

RQ1. *Does generative AI use affect users' voluntary knowledge contribution regarding the number of answers, answer length, and readability on Q&A platforms?*

The answer to this question holds great relevance to industry practices. Q&A knowledge platforms, such as Stack Overflow, have been implementing policies to restrict the use of generative AI in providing answers. Nevertheless, it remains unknown whether generative AI can enhance users' knowledge contribution. Through the investigation of RQ1, we can offer actionable insights to guide the management and implementation of policies concerning the utilization of generative AI on knowledge platforms.

Regarding RQ1, our findings indicate that users' usage of generative AI correlates with their increased answer generation, characterized by shorter and more readable content. This finding implies that the utilization of generative AI can enhance the ability of human users to address questions effectively, as proposed by Jia et al. (2024), where users can generate more answers. Additionally, these answers can be composed in a more concise and easily comprehensible format.

Furthermore, it is unclear how using generative AI affects users' voluntary knowledge contribution on Q&A platforms. If users indeed glean knowledge from these tools during their interaction, then the influence of generative AI may be more pronounced with cumulative usage because it facilitates the knowledge acquisition and learning process (Li et al. 2013). However, if generative AI induces cognitive overload in its users, then more intense use could diminish the impact of generative AI. Therefore, to determine how generative AI affects users' voluntary knowledge contribution, we propose RQ2 as follows:

RQ2. *How do cumulative ChatGPT usage and usage intensity moderate the impact of generative AI on users' voluntary knowledge contributions in terms of the number of answers, answer length, and readability on Q&A platforms?*

By addressing RQ2, we offer valuable managerial insights for knowledge platform managers on effectively harnessing generative AI to facilitate and enhance users' engagement on their platforms. Specifically, our findings provide guidance on determining the appropriate circumstances and methods for permitting the use of generative AI to enhance users' contributions.

Regarding RQ2, our findings indicate that cumulative usage enhances users' learning outcomes. Specifically, users who cumulatively use generative AI tools more tend to generate even more answers, and these answers tend to be even shorter and more digestible. These results suggest that users learn from their interactions with generative AI tools, leading to more concise and easily comprehensible answer generation, as proposed by Kung et al. (2023). These findings are consistent with the principles of social learning theory and existing literature on learning through IT interactions (see, e.g., Barki et al. 2007).

Furthermore, concerning RQ2, our research reveals that intensive usage moderates the impact of generative AI on users' voluntary knowledge contribution. Specifically, higher usage intensity reduces the impact on answer quantity, suggesting a mechanism related to cognitive load. In particular, users experience cognitive burden when intensively interacting with and using generative AI tools. Intensive usage of these tools can increase cognitive load, leading users to adopt mental shortcuts, as noted by Adams et al. (2021), thereby resulting in the generation of fewer answers.

Our research makes several key contributions. Theoretically, we extend the literature on voluntary knowledge contribution in Information Systems (IS) (see, e.g., Qiu et al. 2017, Wang et al. 2022, Liu et al. 2025) by examining how generative AI shapes voluntary knowledge contribution on Q&A platforms. To our knowledge, ours is the first study to explore the impact of generative AI users' voluntary knowledge contribution using the lens of learning processes and social learning theory. We also contribute to the emerging literature on how AI impacts human behavior (c.f. Xue et al. 2022) by proposing two theoretical mechanisms, learning and cognitive load, that underlie these effects. This type of understanding enriches future research endeavors to further understand the nuanced ways AI impacts human behavior.

Furthermore, we contribute to the emerging literature examining the impact of generative AI on voluntary knowledge contribution (see, e.g., Burtch et al. 2024, Quinn and Gutt 2025). Specifically, we examine how users learn from generative AI usage and how this learning influences their voluntary knowledge contributions on Q&A platforms. Practically, our research offers several actionable managerial implications for the design and regulation of generative AI use on knowledge-sharing platforms. Specifically, we believe that these insights will help create policies that effectively utilize generative AI, encouraging users to contribute high-quality content on their platforms.

2. Literature Review

In this section, we summarize previous relevant literature on voluntary knowledge contribution and the impact of AI adoption on user behavior to establish the theoretical foundations for our study. Figure A.1 in Online Appendix A positions our research in existing literature.

2.1. Voluntary Knowledge Contribution

The practice of voluntary knowledge contribution is of key importance for both users and firms (Moqri et al. 2018, Shi et al. 2021, Wang et al. 2022). In online communities, it plays a crucial role in sustaining users' participation, a frequent challenge (Chen et al. 2018, Cao et al. 2024). Moreover, voluntary knowledge contribution has implications for individual users' job-hopping and turnover tendencies (see, e.g., Huang and Zhang 2016, Chen et al. 2022) and firm productivity (see, e.g., Huang et al. 2022). These factors underscore the critical role of voluntary knowledge contribution and its far-reaching effects on both individual users and organizations within the online platform ecosystem.

Previous IS research in this domain has examined largely voluntary knowledge contribution on digital platforms and can consequently be categorized into three distinct streams: the impact of knowledge contribution, assessment of contribution quality, and identification of antecedents influencing the contribution process. For instance, Huang and Zhang (2016) discovered that knowledge contribution is associated with a greater likelihood of job-hopping, highlighting the potential impact of knowledge contribution on individuals' career choices. Furthermore, Liu et al. (2020) developed a text analytic framework aimed specifically at predicting the usefulness of answers, providing insights into assessing the value and relevance of contributed knowledge.

Explorations of the antecedents of knowledge contribution are common in prior literature, with many focusing on intrinsic and extrinsic motivations, including career concerns (see, e.g., Xu et al. 2020, Pu et al. 2022), reciprocity (Yan et al. 2022), peer recognition (Kumar and Qiu 2021, Shi et al. 2024), self-image (see, e.g., Qiu and Kumar 2017, Chen et al. 2018), identity verification (see, e.g., Ma and Agarwal 2007, Pu et al. 2020), incentives (see, e.g., Wang et al. 2022, Wang et al. 2024, Zhang et al. 2025), sponsor investment (Huang et al. 2018), and attention and appreciation (see, e.g., Tan et al. 2022). However, few prior studies have investigated the impact of adopting generative AI on voluntary knowledge contribution. Our study fills that research gap. In addition, we highlight recent great research efforts investigating the impact of generative AI on voluntary knowledge contribution, including users' content generation (see, e.g., Sanatizadeh et al. 2025; Su et al. 2025a, b), questioning behavior (see, e.g., Xue et al. 2023), users' participation and content generation quality in the online community (see, e.g., Burtch et al. 2024), and question-posting activity on StackExchange (see, e.g., Quinn and Gutt 2025), as well as the impact of banning artificial intelligence-generated content (AIGC) on the community-level knowledge demand, provision, and provision efficiency (see, e.g., Borwankar et al. 2023).

Our study contributes to this body of research in several novel ways. First, we focus on understanding the subsequent individual-level impact of prior experience with generative AI usage, whereas current working papers primarily examine effects at the platform level. Second, our study specifically examines the influence of ChatGPT on answer generation, contrasting with existing research, which predominantly explores its impact on questioning behavior or the consequences of its exclusion. Lastly, our study adopts social learning and cognitive burden as theoretical mechanisms to explain the usage of generative AI in affecting individual users' knowledge contribution, whereas previous studies either did not employ a specific theory (Xue et al. 2023) or utilized alternative theories such as information foraging theory (Quinn and Gutt 2025, Sanatizadeh et al. 2025), protection motivation theory (Borwankar et al. 2023), online information content demand and supply (Burtch et al. 2024), and social interactivity theory.

Additionally, we compare our paper with Burtch et al. (2024) in detail because both of us have focused on examining the impact of ChatGPT on knowledge provision (e.g., answer generation). Burtch et al. (2024) examined a similar question; however, our study diverges from theirs in several key aspects, including research focus and findings. First, whereas our paper investigates the cognitive and behavioral adaptations of users as they engage with generative AI, Burtch et al. (2024) primarily explored the implications of AI on content generation and community engagement. This distinction is significant because it underscores two complementary but distinct dimensions of AI interaction; our study sheds light on how generative AI influences individual learning processes and behavioral shifts, whereas Burtch et al. (2024) focused on the broader structural transformations in online communities resulting from AI integration. Second, our research emphasizes how prior interactions with ChatGPT refine user output by fostering brevity and readability. In contrast, Burtch et al. (2024) centered their analysis on how AI impacts user participation patterns and the overall dynamics of online communities. Together, these perspectives provide a more comprehensive understanding of generative AI's multifaceted influence—ours at the individual level and theirs at the community level. A more detailed discussion of these distinctions is summarized in Table A.1 of Online Appendix A and illustrated in Online Appendix A.

2.2. Impact of AI Adoption on Human Behavior

The advancement of AI has profoundly influenced human behavior. Its impact is broad and growing, having influenced domains such as trading behavior (Dempster and Leemans 2006), financial investments (Ge et al. 2021), customer interactions (Schanke et al. 2021), medical diagnosis decisions (Jussupow et al. 2021), and user behaviors on online crowdsourcing platforms (Lysyakov and Viswanathan 2022).

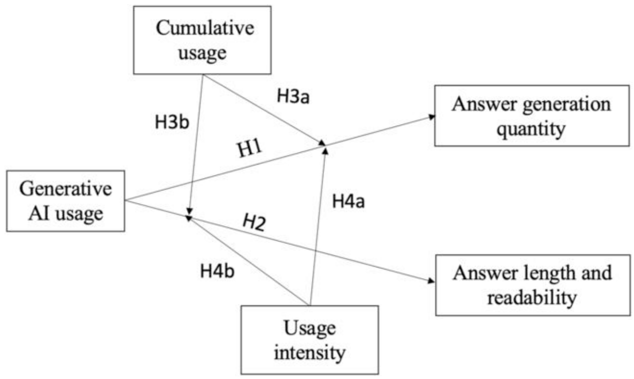
Previous research examining the behavioral effects of AI can be classified into two categories: studies of nongenerative AI and studies of generative AI. Within the domain of nongenerative AI research, research has focused on three main areas: the impacts of nonexplainable AI (Ge et al. 2021, Lysyakov and Viswanathan 2022, Boyaci et al. 2024), explainable AI (Hong et al. 2021, Bauer et al. 2023), and the comparison between AI and human impacts (Fügener et al. 2022, You et al. 2022). For instance, Boyaci et al. (2024) explored the influence of AI input on human decision-making in scenarios where cognitive resources are limited. They posited that although AI can deliver rapid but incomplete information at minimal cognitive cost, human decision-making still demands significant cognitive effort. Their findings indicated that AI predictions can improve decision accuracy, but this often requires individuals to exert more cognitive effort, thereby increasing cognitive load. Moreover, they highlighted that this heightened cognitive burden becomes especially pronounced in situations where cognitive constraints are most severe, such as under time pressure or during multitasking.

In parallel, the rise of generative AI has sparked rapid research interest. Various domains, including education (Baidoo-Anu and Owusu Ansah 2023), healthcare (Vaishya et al. 2023), finance (Dowling and Lucey 2023), labor markets (Yilmaz et al. 2023), and worker productivity (Brynjolfsson et al. 2025), have been influenced. For instance, Yilmaz et al. (2023) demonstrated that generative AI is more suitable for replacing human labor in analytical tasks rather than cultural and emotional ones. These studies highlight the broad range of impacts and applications of generative AI across different domains.

3. Research Model and Hypothesis Development

In this section, we present and discuss our research model (depicted in Figure 1) and develop hypotheses

Figure 1. Research Model



based on social learning theory and cognitive load theory.

3.1. Theory of Social Learning and Cognitive Load of Users' Generative AI Usage

When users employ generative AI tools to respond to questions on voluntary knowledge platforms, the social learning mechanism may shape their behavior.

Users can potentially learn from the generative AI during the answer-generation process by observing the tool's approach. This is consistent with social learning theory, which asserts that individuals can acquire knowledge and skills through observation (Bandura and Walters 1977, Stewart and Jordan 2017). The learning science literature similarly identifies the value of ChatGPT as an effective educational tool (see, e.g., Baker and Reimann 2024). As users generate responses with ChatGPT, they learn how to structure their responses and expand their domain knowledge (Zhang et al. 2025). This learning can enhance users' ability to answer more questions, thereby increasing their overall answer quantity. Moreover, existing research suggests that users' engagement with IT, including generative AI tools, is inherently tied to their learning experiences (Huang et al. 2024).

In our context, users who employ generative AI tools for answering questions can observe how these tools formulate responses, gaining insights that improve their own answer-generation skills. Additionally, generative AI, built on large language models, facilitates quick and efficient response generation (Guo et al. 2023). Research has shown that generative AI boosts productivity by increasing the frequency and quantity of responses to customer queries (Brynjolfsson et al. 2025). As a result, users who are exposed to generative AI can learn from its question-answering processes, enhancing their proficiency in producing responses more efficiently. Hence, users tend to generate more answers on a Q&A knowledge platform after using generative AI. In light of this, we propose the following:

Hypothesis 1. *Users generate more answers on a Q&A knowledge platform after using generative AI.*

The use of generative AI could lead to the creation of shorter and more digestible answers on Q&A platforms. Tools like ChatGPT are trained on extensive data sets with human feedback, enabling them to generate organized, coherent, and concise responses (Guo et al. 2023). In addition, generative AI tools like ChatGPT have demonstrated a strong capacity for producing succinct yet comprehensive summaries (Tian et al. 2024). This capability is especially valuable in the context of coding-related questions on platforms like Stack Overflow, where precision, clarity, and conciseness are paramount. In such cases, users are often tasked with providing solutions that must be both technically accurate and easy to understand. Generative AI assists by generating clear, concise code snippets and explanations, which can significantly reduce the complexity of presenting technical solutions. As users interact with these AI tools, they can learn how to better structure their own responses—both in terms of generating succinct code and delivering concise yet digestible explanations. In addition, learning science literature reveals that ChatGPT is an effective education tool (see, e.g., Baker and Reimann 2024). Users using ChatGPT to generate answers can gain an opportunity to learn from ChatGPT in terms of generating questions in a succinct yet digestible manner, as suggested by Zhang et al. (2025). Furthermore, this learning process is particularly critical for coding questions because Stack Overflow users typically seek direct, efficient answers that solve specific programming issues. By engaging with AI-generated code and explanations, users can improve their ability to highlight the most relevant details, offer cleaner code, and simplify complex technical problems for others. Over time, this enhanced skill set allows them to provide more succinct, structured answers, thereby making them generate shorter and more digestible answers on the platform. The ability of generative AI to streamline complex coding concepts not only helps users refine their cognitive processes but also ensures that their answers are concise, technically sound, and more digestible. This reflects an important feedback loop between tool use and cognitive skill development, as suggested by Barki et al. (2007) and Huang et al. (2024). Therefore, we hypothesize the following:

Hypothesis 2. *Users generate shorter and more digestible answers on a Q&A knowledge platform after using generative AI.*

3.2. Moderating Mechanisms from Cumulative Usage and Usage Intensity

To unpack mechanisms, we explore how the cumulative usage and intensity of generative AI usage moderate its impact on voluntary knowledge contributions. Cumulative usage is expected to amplify the positive effects of generative AI. Repeated interactions reinforce users' familiarity with the AI's functionalities while also facilitating a deeper learning process. As highlighted by Huang et al. (2024), users actively observe and internalize the strategies employed by the AI in generating responses, a concept rooted in social learning theory articulated by Bandura and Walters (1977) and further supported by Stewart and Jordan (2017). Research indicates that the cumulative use of information technology (IT) tools significantly enhances user learning, thereby enriching cognitive processes and improving users' understanding of how these tools operate (Li et al. 2013). Moreover, learning science literature underscores the effectiveness of cumulative learning and repetition as a means to reinforce learning outcomes, allowing learners to practice consistent patterns and enhance their performance (Marton 2006).

In the context of the frequent use of generative AI tools, users can frequently observe how swiftly the AI generates responses and the organization and coherence of the answers produced. Studies by Brynjolfsson et al. (2025) and Guo et al. (2023) further illustrate that this ongoing interaction both familiarizes users with the AI's capabilities and fosters an environment conducive to learning. Consequently, users who cumulatively engage with generative AI tools are likely to strengthen their learning processes, thereby amplifying the overall impact of the technology on their knowledge contribution behaviors. Based on these assertions, we propose the following hypotheses:

Hypothesis 3a. *Cumulative usage enhances the impact of generative AI on users' knowledge contribution quantity.*

Hypothesis 3b. *Cumulative usage enhances the impact of generative AI on users' knowledge contribution quality in terms of the answer length and readability.*

In addition to learning, intensively engaging with generative AI tools could introduce a cognitive burden on users, particularly on Q&A platforms where users must invest additional mental effort to comprehend, refine, and ensure the accuracy of the generated content before posting. As users intensively utilize generative AI, this cognitive burden can become exacerbated; they must allocate additional mental resources to interact with the tool effectively and process the information it provides (Paas et al. 2003). Research suggests that although AI can deliver information rapidly at minimal cognitive cost, it often presents incomplete or contextually ambiguous data that requires users to exert additional cognitive effort to accurately interpret and comprehend this information (Boyaci et al. 2024). Moreover, this additional effort compounds with frequent use, leading to more mental energy expended in adapting the AI-generated content to enhance its accuracy and relevance, potentially diminishing the overall quantity of answers they produce. Simultaneously, the heightened cognitive burden associated with intensive usage can undermine answer quality. High-quality responses typically require additional time and mental effort to polish and refine (Blohm et al. 2016, Zhai et al. 2025). However, when users are subjected to significant cognitive strain because of the intensive use of generative AI, they may find it challenging to dedicate adequate time and effort toward improving the quality of their answers. Therefore, we propose the following:

Hypothesis 4a. *Usage intensity mitigates the impact of generative AI on users' knowledge contribution quantity.*

Hypothesis 4b. *Usage intensity mitigates the impact of generative AI on users' knowledge contribution quality in terms of length and readability.*

4. Research Context and Data

4.1. Research Context

We use Stack Overflow as the context to investigate our research questions and research model. Stack Overflow is a well-established Q&A platform for developers, launched in 2008. It has a user base of 14 million registered users and attracting 100 million monthly visitors in 2022 (David 2023). Given its prominence and relevance to our research, Stack Overflow provides an ideal setting to examine the impact of generative AI on users' voluntary knowledge contribution. Ever since the launch of OpenAI's ChatGPT on November 30, 2022, there has been widespread adoption of this technology for answering questions on the Stack Overflow platform. The developer community, including individuals like Andrew Shearer, has shared their experiences utilizing ChatGPT to provide answers on Stack Overflow (Shearer 2022). As a response to this trend, the platform implemented a policy banning (see Online Appendix B for details) the usage of ChatGPT on December 5, 2022 (StackOverflow 2022).

This brief but pivotal period—from launch to ban—constitutes a natural experiment. In our study, we identified users who generated answers similar to those produced by ChatGPT as treated users, whereas users who did not employ ChatGPT served as control users. By leveraging this natural experiment, we can gain insights into how generative AI affects users' knowledge contribution.

4.2. Treatment Identification

To identify treated users, we utilize the launch of ChatGPT as a natural experiment. Various tools are available to detect whether the answer is generated using generative AI, including the AI text classifier and

GPT-2 Output Detector developed by OpenAI, GPTZeroX developed by Edward Tian (Princeton), and DetectGPT developed by Stanford University (Alcántara 2023, Bowman 2023, Kirchner 2023). Each method has its pros and cons. For instance, the OpenAI text classifier is effective at detecting text generated from GPT3.5, but it may be unreliable for shorter texts below 1,000 characters. The GPT-2 Output Detector, GPTZeroX, and DetectGPT can detect AI-generated text with shorter content, but they require a minimum length. Because the answers on Stack Overflow are usually short (our data have an average of 150 characters), we use the OpenAI GPT-2 Output Detector to identify answers from ChatGPT. This method has been effective in achieving high accuracy, up to 99.3% (Solaiman et al. 2019).

Furthermore, it is essential to acknowledge that our research relied primarily on the OpenAI GPT-2 Output Detector for identifying cases where users generated answers similar to those produced by a generative AI tool or utilized ChatGPT. However, there might be a possibility of misidentifications. To complement the treatment identification process, we employ the ChatGPT API (Brockman et al. 2024) to generate answers for the same posts that users in our data set had responded to. To assess the similarity between the generated ChatGPT answers and the existing answers in our data set for the corresponding posts, we adopt the approach used by Temizkan et al. (2017) and Adamopoulos et al. (2018) to measure the similarity (i.e., Jaccard). Answers with a similarity score exceeding 0.9 were classified as being from ChatGPT. We used 0.9 as the threshold instead of 1 (100% the same), allowing users to modify answers from GPT. The results, which are presented in Table B.1 in Online Appendix B, consistently support the findings obtained from using the GPT-2 Output Detector to identify the treatment of generative AI usage.

To understand our treatment identification accuracy, we also surveyed the Stack Overflow users in our sample, and we found that our treatment identification accuracy was 83.02%, which aligns closely with the treatment classification accuracy of the GPT-2 Output Detector reported by Burtch et al. (2024). More detailed results are put in Online Appendix B.

4.3. Data

To investigate our research questions, we collected data from Stack Overflow three months prior to the launch of ChatGPT on November 30, 2022, until its ban on December 5, 2022. We also used data from one month prior to the launch of ChatGPT and found consistent results (shown in Online Appendix C). Our data were drawn directly from the Stack Overflow platform between September 1, 2022, and December 4, 2022. The data set collected contains 3,238,381 questions and 1,254,841 answers from 223,696 users. During this period, there were no other policy changes on the platform, providing a pure natural experiment to examine the effects of generative AI on knowledge contribution.

In this research, we focus on two dimensions of knowledge contribution: quantity and quality. We measure contribution quantity using the number of answers generated. Content generation frequency has been commonly used as a performance metric in other user-generated contexts (see, e.g., Burtch et al. 2022). To measure the quality of contribution, we use answer length and readability, which are commonly used in prior research on knowledge contribution (see, e.g., Qiao et al. 2020). We also use answer scores, measured by the number of upvotes, which indicate the quality of answers evaluated by other users (Lee et al. 2019). We put our results associated with the answer score in Online Appendix D. Our data are composed of those variables shown in Table 1. To gain a better understanding of our data, we provide data summary statistics and a correlation matrix of variables in Tables 2 and 3.

From Table 2, the mean value of treatment is 0.051, indicating that 5.1% of observations (=29,478) involve using generative AI to answer questions. The average tenure is 2,099.131, indicating that users exist for an average of 2,099.131 days when they write an answer. Other summary statistics related to the answer generation quantity and quality are also generated. Table 3 shows that most variables considered have weak or small correlations (the absolute coefficient is less than 0.3) (Cohen 1988).

5. Estimation Strategy

This study is chiefly interested in examining the impact of using generative AI on voluntary knowledge contribution on Q&A platforms. To achieve this, we employ a difference-in-differences (DID) estimation strategy, taking advantage of the launch of ChatGPT as a natural experiment. Specifically, the launch of ChatGPT provides a potential treatment for every user on the platform to be treated (e.g., using ChatGPT to generate answers). To structure the data, we use a panel format, considering individual users as the unit of analysis and days as the time unit. Our panel is unbalanced because we exclude periods when users did not generate any answers. Given that the usage of generative AI can occur at various times, our DID analysis is a special case that accommodates the switching of treatment for individual users, as outlined by Athey and Imbens (2022). Furthermore, our panel data exhibit a general panel treatment structure, allowing for the possibility of treatment reversals (Liu et al. 2024). To understand whether the launch of ChatGPT is possibly exogenous and whether the parallel trend assumption of the DID

Table 1. Variables and Definitions

Category	Variables	Definitions
Independent Variable	$Treatment_{it}$	It shows whether user i generates answers similar to those produced by generative AI tools on day t.
Dependent Variables	$NumAnswer_{it}$	It represents the number of answers generated by user i on day t.
	$Length_{it}$	It represents the average length in terms of the number of words of answers generated by user i on day t.
	$Readability_{it}$	It represents the average readability of answers generated by user i on day t. We used gunning-fog measure, indicating the years of formal education needed to understand the answers, following a group of others such as the Automated Readability Index (ARI) and Flesch–Kincaid Reading Ease (FRE) test, following Johnson et al. (2015). We put the results of ARI and FRE in Online Appendix D.
	$Score_{it}$	It represents the average upvotes of answers generated by user i on day t.
Control Variables and Moderators	$Tenure_{it}$	It represents the existing duration of user i on the Stack Overflow platform until day t, measured by the number of days that passed since the launch of the user account.
	$Upvote_{it}$	It represents the ratio of questions answered by user i on day t, which received upvotes from other users.
	$Question_{it}$	It represents the number of questions asked by user i on day t.
	$Badge_{it}$	It represents the number of badges user i received on day t.
	$Bounty_{it}$	It represents the ratio of questions containing bounty answered by user i on day t.
	$Cumu\_usage_{it}$	It quantifies the cumulative ChatGPT usage and is measured by the log transformation of the cumulative answers generated using ChatGPT by user i until day t.
	Intensity_{it}	It measures how intensively user i used ChatGPT in answer generation on day t. It is measured by the usage frequency divided by the average time duration among those answers generated by ChatGPT by user i on a particular day t. Specifically, if user i used ChatGPT to generate two answers, A and B, on day t, then we define $Intensity_{it}$ as the number of answers divided by the time duration (in hours) between answers A and B. If the user generated more than two answers on the same day, then the denominator becomes the average time interval (in hours) between consecutive answers, and the numerator is the total number of answers generated. In cases where the user generated only one answer on day t, we set $Intensity_{it}$ to 1 divided by 24, treating 24 hours as the default time window. However, in our robustness checks, we also experimented with alternative denominators of 0.5, 1, 2, 3, 4, 5, 6, 7, and 8 hours to assess the sensitivity of our measure. If no answers were generated on a given day, then $Intensity_{it}$ is assigned a value of zero.

Notes. The unit of analysis of this study is a user and day. In other words, all of our variables are measured at the daily level for each user in our sample. To avoid potential data-skewed distribution, we use the log transformation of NumAnswer, named as Log(NumAnswer), following Song et al. (2019).

model holds, we also conduct lead and lag models to examine the parallel trends. However, there are several challenges associated with employing the DID estimation in our context, as detailed in Online Appendix E. Those challenges include unobserved time-varying confounders and treatment heterogeneity, potential parallel trend violation, self-selection on the observable factors, self-selection process, reverse causality, answers from generative AI that may affect the results, false significance, and treatment misclassification. Our strategies for addressing potential endogeneity concerns are summarized in Table 4.

6. Empirical Results

6.1. Baseline Results

In this research, we aim mainly to investigate how generative AI affects users' voluntary answer generation on knowledge contribution platforms. Our main dependent variables are the number of answers generated, the answer length, and readability. We use a two-way fixed effects model to empirically examine the main impact as our baseline results, and the model specifications are shown here in Equation (1).

$$Y_{it} = \alpha_i + \beta_0 + \beta_1 \; Treatment_{it} + \beta_2 \; Controls + \gamma_t + \varepsilon_{it}, \tag{1}$$

Table 2. Variable Summary Statistics

Variables	Obs	Mean	SD	Min	Max
Treatment	578,480	0.051	0.220	0	1
NumAnswer	578,480	2.163	3.715	1	297
Log(NumAnswer)	578,480	0.984	0.475	0.693	5.697
Length	578,480	26.639	19.162	1	453
Readability	578,480	10.118	6.489	0	280.400
Score	578,480	0.179	0.501	0	78
Tenure	578,480	2,099.131	1,544.297	0	5,240
Upvote	578,480	0.222	0.364	0	1
Question	578,480	2.733	5.030	1	332
Badge	578,480	0.186	0.612	0	155
Bounty	578,480	0.001	0.031	0	1
Cumu_usage	578,480	2.148	1.527	0.693	9.760
Intensity	578,480	0.011	0.143	0	6.903

Table 3. Variable Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treatment (1)	1.00										
Log(NumAnswer) (2)	0.02	1.00									
Length (3)	−0.02	0.04	1.00								
Readability (4)	−0.01	0.04	0.27	1.00							
Score (5)	−0.01	0.07	0.02	0.03	1.00						
Tenure (6)	−0.004	0.05	0.02	0.07	0.13	1.00					
Upvote (7)	−0.02	0.03	−0.03	−0.02	0.07	0.08	1.00				
Question (8)	0.02	***0.74***	0.02	0.02	0.09	0.07	0.09	1.00			
Badge (9)	−0.004	0.06	0.002	−0.02	−0.02	−0.19	0.05	0.05	1.00		
Bounty (10)	−0.0002	−0.001	−0.001	0.001	0.001	0.01	−0.01	−0.001	0.01		
Cumu_usage (11)	0.02	***0.56***	0.03	0.06	0.21	0.22	0.03	***0.45***	−0.07	0.002	
Intensity (12)	0.32	0.08	−0.01	−0.01	−0.01	−0.01	−0.01	0.05	0.01	−0.001	0.07

Notes. The values in Table 3 represent the Pearson correlation coefficient. Strong correlations (0.60–0.79) and moderate correlations (0.40–0.59) are shown in bold and italics.

where Y_{it} indicates a dependent variable of user i at a specific day t; $Treatment_{it}$ is a dummy variable, representing whether answers from user i are similar to those produced by the generative AI at a particular day t; *Controls* reveals mainly the time-variant control variables, which include $Tenure_{it}$, $Upvote_{it}$, $Question_{it}$, $Badge_{it}$, and $Bounty_{it}$. Their definitions are stated in Table 1.

Table 5 reveals that using generative AI tools significantly correlates with users' knowledge generation. Specifically, we observe that using generative AI tools correlates with users' generation of 16.77% (=exp(0.155) − 1) more answers every day, on average. However, the answer length decreases by 22.64, whereas answers will be easier to read by reducing the 2.165 years of education needed to understand the answers, on average. Hence, Hypothesis 1 and Hypothesis 2 are supported. In addition, we follow Athey and Imbens (2022) and arrange the treatment as a staggered adoption DID, getting consistent findings (Table F.1 in Online Appendix F).

These empirical findings address the interests of RQ1, Hypothesis 1, and Hypothesis 2. Our findings indicate that the usage of generative AI tools does indeed correlate with users' voluntary knowledge contribution behavior on Q&A platforms. Specifically, we observe that users generate more answers with shorter and more readable content when using generative AI. One potential mechanism underlying these results is the influence of social learning (Bandura and Walters 1977) because users tend to learn and mimic the answers from generative AI tools. This learning process enables users to generate answers more quickly with more concise and easier-to-read content (Tian et al. 2024).

6.2. Main Results and Checks to Address Potential Endogeneity Concerns

6.2.1. Imputation Estimator and Parallel Trend Investigation. According to Goodman-Bacon (2021), treatment can be heterogeneous at either the individual or

Table 4. Main Identification Strategy and Robustness Checks

Potential issues	Adopted methods	Findings
Unobserved time-varying confounders and treatment heterogeneity	Imputation estimator	Unobserved time-varying confounders and treatment heterogeneity do not affect the main findings.
Parallel trend violation	Generalized synthetic control method	The main results are consistent.
Self-selection on the observable factors	PSM + DID	Self-selection on observational factors does not affect the main findings.
Self-selection on the observable factors	CEM + DID	Self-selection on observational factors does not affect the main findings.
Self-selection process	Two-stage Heckman correction	Self-selection process does not affect the main findings.
Reverse causality	Remove observations generating the last answer via generative AI of a day; Lag the treatment	Future usage will not affect the main findings.
Answers from generative AI may affect the results	Only focus on the organic answers	The results are robust.
False significance	Random implementation tests	False significance is not an issue.
Treatment misclassification	Permutation tests	Our results are robust in terms of the treatment misclassification.

Table 5. Estimation Results of the Baseline DID Model

Variables	(1) Log(NumAnswer)	(2) Length	(3) Readability
Treatment	0.155***	−22.64***	−2.165***
	(0.01)	(0.75)	(0.21)
Tenure	−0.00197***	0.109***	−0.244***
	(0.0001)	(0.01)	(0.01)
Upvote	−0.0349***	−0.422***	−0.203***
	(0.01)	(0.09)	(0.03)
Question	0.0701***	0.0643***	0.0106***
	(0.004)	(0.01)	(0.002)
Badge	0.0183***	−0.0745	−0.0190
	(0.0048)	(0.05)	(0.01)
Bounty	−0.0300**	1.241	0.181
	(0.01)	(0.83)	(0.31)
Constant	4.856***	−199.8***	518.4***
	(0.72)	(28.74)	(10.60)
User fixed effects	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes
Observation	578,480	578,480	578,480
R^2	0.50	0.01	0.001

Notes. User cluster standard errors in parentheses. The readability (gunning fog index) means the years of formal education needed to understand the generated answers. We use log transformation of NumAnswer, and the result is consistent by using NumAnswer, shown in Table D.2 in Online Appendix D. We also include the number of cumulative questions asked by the user as a control with getting consistent results, shown in the Table D.3 of Online Appendix D. Witnessing the unbalanced panel data, we also organized a balanced panel by including the zero-answer days for examining the impact on the number of answers with getting consistent results on the impact on number of answers, shown in Table D.4 of Online Appendix D. We also use the log transformation for other outcome variables with consistent findings shown in Table D.5 of Online Appendix D. In addition, we examine the impact on answer scores with insignificant findings (shown in Table D.6 in Online Appendix D).
**$p < 0.01$; ***$p < 0.001$.

time level or both, which can make the DID estimation biased. In our research setting, the treatment (e.g., using a generative AI tool to answer questions) occurs at different times for different treated units (a general panel treatment structure) because users can decide to use the AI tool on different days for different questions. Thus, the traditional DID estimation may bias our results.

Recent literature has proposed several methods to address treatment heterogeneity in the context of the general panel treatment structure (Cengiz et al. 2019, Baker et al. 2022, De Chaisemartin and D'Haultfoeuille 2022, Liu et al. 2024). These methods can be categorized into three groups: imputation estimator (see, e.g., Liu et al. 2024), Cohort-Specific Average Treatment Effects on the Treated (CATT) estimator (see, e.g., De Chaisemartin and D'Haultfoeuille 2022), and stacked regression estimator (see, e.g., Cengiz et al. 2019). In our research, we primarily employ the imputation estimator to address treatment heterogeneity. This choice is motivated by several factors. First, the stacked regression estimator is inconsistent with the sample-average treatment effect, which makes it less suitable for our analysis (Baker et al. 2022). Second, by employing the imputation estimator, we can effectively address the potential occurrence of treatment reversals. This estimator enables us to account for situations where the treatment may be switched back and forth during the study period (Liu et al. 2024). Third, the imputation estimator facilitates the examination of parallel trends, an important consideration in our study (Pan and Qiu 2022, Liu et al. 2024).

For the imputation estimator, we follow Liu et al. (2024) and construct the imputation estimator. The imputation estimator we employed possesses the capability to address a certain degree of unobserved time-varying confounders, which is achieved through the utilization of a factor-augmented model to estimate the counterfactual potential outcomes for the untreated units. As outlined by Liu et al. (2024), this modeling approach has the potential to mitigate the impact of unobserved time-varying confounders on the differences observed between treated and untreated units. Our estimation results shown in Table 6 show that the estimations of the imputation estimator are consistent with the DID model.

To ensure the validity of our DID estimator, we thoroughly examine the parallel trend assumption. Failing to satisfy this assumption would undermine the reliability of our analysis. To assess the parallel trend assumption, we employ leads and lags models, as suggested by Alyakoob and Rahman (2022) and Liu et al. (2024). Liu et al. (2024) cautioned that the coefficients and p-values of the treatment indicator and dummy variables in traditional leads and lags models may be biased because of their dependence on the selected baseline category. To overcome this limitation, we follow their approach and plot the average differences between the outcomes and the predicted counterfactual for each user in period t (relative to the treatment) in the

Table 6. Estimation Main Results of Imputation Estimator

Variables	(1) Log(NumAnswer)	(2) Length	(3) Readability
Treatment	0.182***	−22.228***	−2.122***
	(0.01)	(0.96)	(0.32)
Tenure	7.806e-12	3.853e-11	−8.457e-11
	(4.48e-12)	(1.44e-10)	(6.96e-11)
Upvote	−0.039***	−0.224*	−0.091*
	(0.01)	(0.11)	(0.04)
Question	0.068***	0.501***	0.007
	(0.004)	(0.006)	(0.002)
Badge	0.018**	0.094	−0.010
	(0.01)	(0.05)	(0.02)
Bounty	−0.048***	1.152	0.462
	(0.02)	(0.91)	(0.37)
Constant	0.793***	27.382***	10.244***
	(0.01)	(0.09)	(0.03)

Note. Standard errors in parentheses.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

treatment groups, utilizing the imputation estimator. These plots are presented in Figure 2 for the user-level dependent variables. By examining the average pretreatment residuals, we can determine whether they converge to zero, which indicates the presence of a parallel trend.

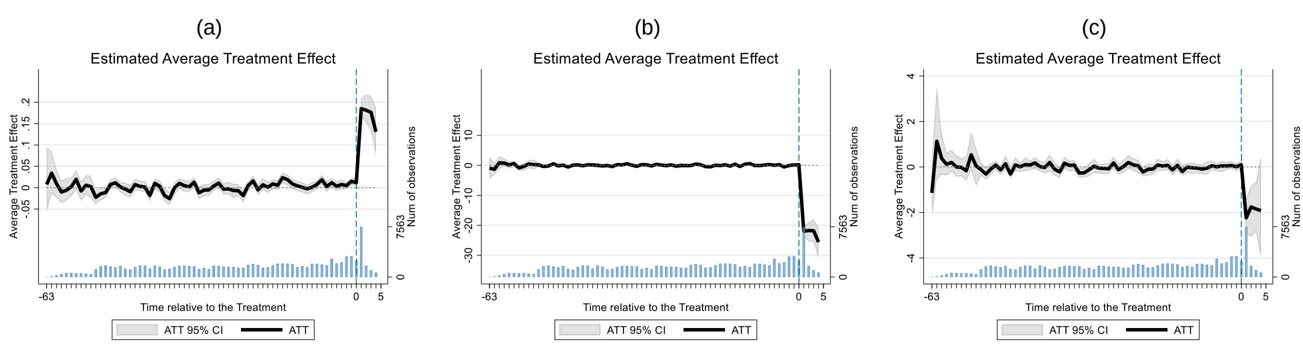
Figure 2 shows that the average pretreatment residuals (i.e., the shaded area before period 0, relative to the treatment) converge to zero across different outcome variables (Figure 2, (a)–(c)), indicating that the parallel trend assumption holds. Additionally, after period 0, we observe a positive impact on the *Log(NumAnswer)* and a negative impact on *Length* and *Readability*. These findings provide suggestive evidence supporting our main conclusions regarding the impact of generative usage.

6.2.2. Generalized Synthetic Control Method. To further enhance the causal identification, we followed Xu (2017) and conducted the generalized synthetic control method (GSCM), mitigating the systematic differences between users using and not using ChatGPT, by constructing weighted control units that match the outcome variable of the treated units in the pretreatment period. Our results shown in Table G.1 of Online Appendix G reveal consistent findings.

6.2.3. DID With Matching Techniques. Self-selection of using generative AI tools (e.g., ChatGPT) may bias the results. The decision to adopt generative AI tools may depend on user-level characteristics (such as the tenure of users in answering questions and the cumulated number of questions they ask) and question-level characteristics (such as the question length, polarity, subjectivity, readability, and topics obtained by running LDA with the detailed process shown in Online Appendix H). To minimize the effect of those observable characteristics (Li 2016, Kumar et al. 2018), we used PSM to realign the users who used generative AI tools versus those who did not on a specific day. Following the literature, we prepared the control group (users who did not use generative AI tools) matching the treatment (users who used generative AI tools) by a one-to-one, nearest-neighbor selection process. We located the best-matching users from the control group for each unit in the treatment group in terms of observed characteristics. While running the PSM, we followed John et al. (2020) by using the caliper matching in searching for the nearest neighbor matched sample. After matching, the usable sample size decreased from 578,480 to 58,512. Our PSM matches users at a day level, which is a dynamic matching technique (Aral et al. 2009). According to Aral et al. (2009), in the dynamic matching process, the propensity to be treated for a user at a particular time period can be explained by users' observable characteristics and behaviors. We used the dynamic matching technique because of the nature of our general panel treatment structure, where the treatment is allowed to switch back and forth over time. Matching users at the day level enables us to appropriately account for the dynamic treatment patterns and ensure accurate comparisons between treated and control groups throughout the study period.

The PSM performance and statistics reported in Figure H.2(a) in Online Appendix H show that the imbalance of all covariates was significantly reduced after matching, establishing that the treatment and control groups were well matched in terms of observable characteristics. Figure H.2(b) in Online Appendix H reports the propensity score histogram, noting overlapping distributions between treatment and control groups. Table H.2 in Online Appendix H reports the mean difference tests on the matched observed characteristics across control and treatment groups, which confirms that the mean differences of all covariates become insignificant after matching. We reran the regression analysis across this modified sample, estimating Equation (1). The results reported in Table H.3 of Online Appendix H show that all our findings are robust; that is, using generative AI tools is associated

Figure 2. (Color online) Parallel Trend



Note. (a) *Log(NumAnswer)*; (b) *Length*; (c) *Readability*.

with an increased number of generated answers, shorter answer length, and easier-to-read answer content.

In addition, to test the robustness of the matching techniques, we also used traditional static matching, where we matched the users based on their characteristics, including their self-disclosed experience of using a large language model, their programmer identity, and gender before the launch of the ChatGPT on the Stack Overflow platform. Our matching results shown in Table H.4 reveal consistent findings.

Aside from using PSM, we also conduct CEM to mitigate users' self-selection of using generative AI. Compared with PSM, CEM can produce matched samples with lower covariate imbalance (Ge et al. 2021). Specifically, CEM creates bins for each covariate into meaningful bins, matches the treatment and control samples based on the bins, and keeps the original values of the covariates for analysis (Blackwell et al. 2009). Adopting this procedure by following Ge et al. (2021), we obtained 573,862 observations. We reran the regression analysis across this modified sample, estimating Equation (1). The results reported in Table H.5 in Online Appendix H show that all of our findings remain robust.

6.2.4. DID With Heckman Correction Model. Given that users have the choice to use generative AI tools when answering questions, their decisions may be influenced by various factors, including their willingness to participate and their overall mood, as observed in other contexts like the online review generation process (Yin et al. 2021). To address concerns related to self-selection bias in using generative AI tools, we employ a two-stage Heckman correction model (Heckman 1979, Gopal and Sivaramakrishnan 2008). In the first stage, we incorporate observable characteristics such as users' tenure and the number of questions they ask per day to satisfy the inclusion condition. Additionally, we include a plausibly exogenous variable, the cumulated number of answers generated by other users using ChatGPT (CNAGOC), to meet the exclusion condition. Because CNAGOC measures the cumulative ChatGPT usage by other users, which does not directly influence the answer generation of focal users but rather impacts their own ChatGPT usage, the exclusion condition is satisfied. Using a Probit model, we calculated the inverse Mills ratio (IMR) to account for the self-selection process of using generative AI tools. We then added the calculated IMR to the main model shown in Equation (1).

Panel A of Table I.1 in Online Appendix I reports the results of the first stage. Panel B of Table I.1 in Online Appendix I reports the results of the second stage of this analysis and shows that the main effects of generative AI are robust and consistent after accounting for the potential self-selection bias.

6.2.5. Simultaneity. In our econometric model, we treat users using ChatGPT to generate an answer as being treated. However, the answer could be the last answer of the day, which makes the estimation using future usage to predict historical behavior. To address this potential simultaneous concern, we checked our data and found that 4.07% of the users generated the last answer on a particular day using ChatGPT. To test the robustness of our finding here, we removed those periods and reran our main model. The results shown in Table J.1 of Online Appendix J reveal consistent findings. To further address the issue of simultaneity, we introduced a one-period lag in the treatment, analyzing how generative AI usage in the previous period affects users' responses in the current period. Our results shown in Tables J.2 and J.3 of Online Appendix J reveal consistent findings. Career-related motives can also affect users' decisions to use ChatGPT. To mitigate this potential concern, we ran subgroup analyses within different careers, such as programmers or software developers, disclosed by users themselves, finding consistent results shown in Tables J.4–J.6 of Online Appendix J.

6.2.6. Falsification Tests. We conducted two falsification tests to examine the impact on organic answering behavior and assess the potential for false significance. In our analysis of users' behavior, we did not distinguish between answers from ChatGPT and those directly from human users, which could introduce bias in our results regarding the impact on users' answering behavior. The rationale behind this approach is grounded in the notion that if users genuinely learned answer generation methodology from ChatGPT, their ability to generate answers independently should be influenced. Existing literature supports the idea that users' learning can affect their subsequent behaviors (Yaraghi et al. 2015). Therefore, if users learned from ChatGPT in generating answers, it should impact their organic answering behavior. Consequently, we removed answers generated by ChatGPT when measuring the outcome variables and reran our models. The results, presented in Table J.7 in Online Appendix J, demonstrate that our main findings remain robust.

Additionally, in our DID estimation, a potential concern arises regarding significant effects stemming from false significance because of spurious correlations in dependent variables. To address it, we followed the literature (see, e.g., Burtch et al. 2018, Cheng et al. 2020, Pamuru et al. 2021) and conducted a falsification test. This involved randomly implementing the treatment to each individual during a specific day (e.g., randomly generating and assigning the treatment). By using pseudo-treatments for each user, acting as placebos, we conducted regressions (Equation (1)) to obtain treatment estimates. This falsification test was replicated 1,000 times. The results, presented in Table J.8 in Online Appendix J, indicate that the mean value of the random treatments is close to zero and significantly different from our estimated effect. This suggests that the impacts of our treatments are unlikely to be driven by spurious relationships or serial correlations.

6.2.7. Permutation Tests. Our treatment identification may not be perfect because of the potential misclassification of some users. To ensure the robustness of our findings, we adopted the same procedure used by Zhou and Lee (2024), who faced similar challenges in identifying generative AI usage, and conducted permutation tests to validate the results. In these tests, we randomly assigned treatment to a sample of users in the control group and performed simulations to assess the sensitivity of the impacts of generative AI usage to potential measurement errors in treatment identification. Specifically, we executed 50 simulations, increasing the proportion of units in the control group exposed to measurement error by 2% for every five iterations. Consistent with the approach taken by Zhou and Lee (2024), we reported the results of the worst-case scenario, where 10% of the control group was mislabeled, in Table J.9. Additionally, we visualized all the findings in Figure J.1 in Online Appendix J, which corroborate these results.

6.3. Moderation Mechanisms

To dissect the mechanisms through which the use of generative AI tools correlates with users' voluntary Q&A contributions, we explore moderating factors such as cumulative ChatGPT usage and usage intensity. These factors help us test the proposed learning and cognitive burden mechanisms. Because our treatment for a user is reversible—allowing it to switch on and off, as outlined by Liu et al. (2024)—the learning and cognitive load associated with ChatGPT usage occurs primarily when users engage with ChatGPT to generate an answer (i.e., when the treatment is active).

The interaction between users and IT influences their learning process (Barki et al. 2007, Huang et al. 2024). In our case, users employing ChatGPT to generate answers can learn from observing how responses are generated by the tool. Moreover, existing literature suggests that cumulative usage enhances users' learning processes by solidifying their understanding of the tool's functionalities (Li et al. 2013). Hence, we utilize the cumulative ChatGPT usage (*Cumu_usage*) to delve into the underlying learning mechanism (Kwon et al. 2023). It quantifies users' cumulative ChatGPT usage. Specifically, we measure it by using the log transformation of the number of answers generated by a user using ChatGPT until a particular day.

Furthermore, intensive ChatGPT usage can impose a cognitive burden on users, particularly when they need to process and potentially modify numerous answers generated by ChatGPT during intensive usage periods. This cognitive burden may result in users producing fewer answers and generating content of lower quality, as suggested by Blohm et al. (2016). Usage intensity (*Intensity*) assesses how frequently a user employs ChatGPT in generating answers. It measures the frequency and concentration of AI tool usage within a short timeframe, with the detailed measurement shown in Table 1. To examine the moderating effects, we employ the following regression model to uncover the underlying mechanism:

$$Y_{it} = \alpha_i + \beta_0 + \beta_1\ Treatment_{it} \times Moderator_{it} + \beta_2\ Treatment_{it} + Controls + \lambda_t + \varepsilon_{it}. \tag{2}$$

The meanings of the variables are the same as in Equation (1). $Moderator_{it}$ refers to *Cumu_usage* or *Intensity* by user i on day t. We use the interaction term ($Treatment_{it} \times Moderator_{it}$) to investigate the moderation effects. We used control variables, including $Tenure_{it}$, $Upvote_{it}$, $Question_{it}$, $Badge_{it}$, and $Bounty_{it}$. The results of the matched sample are shown in Tables 7 and 8.

From Table 7, we observe that cumulative usage indeed moderates the main effect, indicating a notable influence on user engagement. Specifically, when users have a higher cumulative usage, we observe a notable increase in the impact on the logarithm of the number of generated answers, with an average increase of 0.296. Moreover, the impact on the length and readability of the generated answers also sees an accentuation of 2.582 and 0.490, respectively. Consequently, we find support for Hypothesis 3a and Hypothesis 3b.

We also present the results associated with the scores in Table D.8 in Online Appendix D. We find that the treatment effect on scores becomes negative after the interaction effect is added, which means that when there is no cumulative usage, the impact of generative AI usage on the answer score is negative. A plausible explanation is that when users employ generative AI tools, their responses may resemble AI-generated content, which often lacks the interactive and dialogic qualities characteristic of human contributions. Although AI-generated content (or content resembling it) may demonstrate higher technical accuracy, it often feels more transactional and less engaging. This absence of a human element may diminish its ability to foster interaction or resonate with the community, potentially leading to fewer upvotes (Lee et al. 2019). To ensure the robustness of the findings associated with the learning, we also introduced two further moderators related to users' answering experience and expertise, with findings displayed in Tables K.2 and K.3 in Online Appendix K, which support the learning mechanism.

Table 7. Interaction Effect of Cumulative Usage on the Matched Sample

Variables	(1) Log(NumAnswer)	(2) Length	(3) Readability
Treatment × Cumu_usage	0.296***	−2.582***	−0.490***
	(0.00509)	(0.232)	(0.0898)
Treatment	0.534+	−42.13**	−21.37***
	(0.303)	(13.84)	(5.347)
Tenure	−0.00582	0.445**	0.227***
	(0.00322)	(0.147)	(0.0569)
Upvote	−0.110***	0.0223	−0.0680
	(0.00681)	(0.311)	(0.120)
Question	0.0461***	0.124***	0.0289***
	(0.000343)	(0.0157)	(0.00605)
Badge	−0.0144**	0.231	−0.152
	(0.00523)	(0.239)	(0.0922)
Bounty	−0.0348	5.130	−0.0269
	(0.0725)	(3.310)	(1.279)
Cumu_usage	−0.00100	−0.125	0.0462
	(0.00405)	(0.185)	(0.0716)
Constant	12.36	−863.9**	−445.1***
	(6.453)	(294.8)	(113.9)
User fixed effects	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes
Observation	59,194	59,194	59,194
R^2	0.90	0.84	0.81

Notes. User cluster standard errors in parentheses. Cumu_usage is measured by the log transformation of the cumulative answers generated using ChatGPT by user *i* until day *t*.
$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Furthermore, to test the robustness of the results of the learning mechanism (Kwon et al. 2023), we designed a repetitive ChatGPT usage (*Repetitive_Usage*) to check whether users engage with ChatGPT multiple times on a specific day. In general, it measures the frequency of AI tool engagement over a broader time span. If users utilize ChatGPT at least twice within the same day, their repetitive usage is recorded as one, indicating that they have engaged in repeated interactions with the tool on that day. We ran the moderation analysis using *Reptitive_Usage* and found that using ChatGPT repetitively can enhance the generative AI impact, signaling the learning mechanisms. The results are shown in Table K.4 in Online Appendix K.

These empirical findings provide insights into RQ2 and Hypothesis 3, focusing on the moderating role of cumulative usage in the impact of generative AI usage. Overall, our results indicate that users who cumulatively use generative AI tools experience a more pronounced impact, suggesting a learning mechanism at play. This aligns with prior research highlighting the learning mechanism associated with IT usage and the enhancement of learning through cumulative engagement (Barki et al. 2007, Li et al. 2013).

From Table 8, intensity indeed moderates this impact. Specifically, for each unit increase in intensity, we observe a decrease of 0.021 in the impact on the logarithm of the number of generated answers, on average. However, the coefficients of the interaction terms are not significant for length and readability. Consequently, we find support for Hypothesis 4a but not Hypothesis 4b.

These empirical findings shed light on RQ2 and Hypothesis 4, which delve into the moderating role of usage intensity in the impact of generative AI on users' voluntary knowledge contribution on Q&A platforms. Specifically, our results indicate that more intensive usage of generative AI tools leads to a less pronounced impact on answer quantity. This suggests that users experience cognitive burdens when engaging with generative AI tools intensively, resulting in the generation of fewer answers. These findings align with literature highlighting the negative impact of cognitive burden on user performance (Blohm et al. 2016).

In Table 8, we used 24, represented as a 24-hour period, as the denominator when we calculated the usage intensity for those cases where only one usage was on a particular day. However, it may not be precise because a normal user cannot stay on Stack Overflow for 24 hours. To ensure that our results are robust, we also examined using 0.5h, 1h, 2h, 3h, 4h, 5h, 6h, 7h, 8h, and maximum time duration as the denominator, which resulted in consistent findings presented in Tables K.5–K.14 in Online Appendix K. Additionally, we focus on the organic answers to understand whether there is a substitution effect. Our results, shown in Table K.15 in Online Appendix K, reveal that the intensive usage of ChatGPT negatively impacts the

Table 8. Interaction Effect of the Usage Intensity on the Matched Sample

Variables	(1) Log(NumAnswer)	(2) Length	(3) Readability
Treatment × Intensity	−0.0210***	0.0574	−0.0636
	(0.00537)	(0.249)	(0.0961)
Treatment	0.610*	−43.50**	−21.59***
	(0.299)	(13.86)	(5.349)
Tenure	−0.0397*	−0.00586	0.449**
	(0.0184)	(0.00318)	(0.148)
Upvote	−0.859***	−0.117***	0.0447
	(0.0389)	(0.00672)	(0.312)
Question	0.549***	0.0470***	0.0859***
	(0.00192)	(0.000331)	(0.0153)
Badge	−0.0632*	−0.0199***	0.235
	(0.0299)	(0.00516)	(0.239)
Bounty	−0.309	−0.0909	5.364
	(0.414)	(0.0715)	(3.317)
Repetitive_usage	0.443***	−1.641***	−0.303*
	(0.00816)	(0.379)	(0.146)
Constant	12.44	−872.1**	−446.2***
	(6.369)	(295.4)	(114.0)
User fixed effects	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes
Observation	59,194	59,194	59,194
R^2	0.90	0.84	0.81

Notes. User cluster standard errors in parentheses. Usage intensity is measured by the usage frequency divided by the average time duration among those answers generated by ChatGPT by user i on a particular day t. Repetitive_usage is measured by whether the user i used generative AI to generate answers at least twice on a particular day t.
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

number of organic answers, which suggests a substitution effect associated with the usage of ChatGPT.

To further disentangle the substitution effect from the cognitive load effect, we refined our measurement of usage intensity. Specifically, we focused on the intensive use of ChatGPT to generate answers for more complex questions—operationalized by question length—under the assumption that longer questions demand greater cognitive effort. Such questions typically require users to spend more time digesting the content and evaluating whether the AI-generated responses adequately address the underlying issues, thereby increasing cognitive load. To systematically identify complex questions, we applied several threshold criteria based on question length, including the mean, median, and top third of the distribution within our data set. Using these thresholds, we distinguished between complex and noncomplex questions. We then conducted a moderation analysis similar to our earlier approach but used the number of ChatGPT-generated answers as the outcome variable. This choice allowed us to isolate the cognitive load mechanism by excluding organic answers and thereby minimizing the influence of substitution effects. Our results, presented in Table K.16 in Online Appendix K, indicate that intensive use of ChatGPT on longer, more complex questions is associated with a reduction in the number of subsequent ChatGPT-generated answers. This finding suggests that increased cognitive load, rather than substitution, is likely responsible for the observed decline in usage. This refined analysis provides stronger empirical support for the role of cognitive load and further reinforces the distinction between cognitive and behavioral mechanisms in AI-assisted content generation.

6.4. Additional Mechanisms Checks

To further investigate the mechanisms of learning and cognitive burden underlying the impact of generative AI usage, we conducted additional tests focusing on: (1) inter-answer duration using a hazard model, (2) answer speed, (3) writing style, (4) additional surveys, and (5) suggestive evidence of users' writing patterns.

Following the approach of He et al. (2022) and Li et al. (2010), we employed a Cox proportional hazard model, with inter-answer duration as the dependent variable and user treatment as the independent variable, controlling for the same variables used in our DID models. Our results, presented in Table L.1 in Online Appendix L, indicate that treated users who adopted generative AI are more likely to generate answers faster. Therefore, it appears that users become quicker and more frequent in generating answers after adopting generative AI.

Furthermore, we examined the impact of ChatGPT usage on users'writing style changes. The detailed results are shown in Online Appendix L.2. The results presented in Table L.2 in Online Appendix L demonstrate that the use of ChatGPT significantly influences users' writing styles, resulting in a closer alignment between their own writing and the style of ChatGPT-generated responses (as indicated by smaller stylistic differences). Specifically, when users have previously utilized ChatGPT, the length, polarity, subjectivity, and readability of their answers—after ChatGPT usage is restricted—show a measurable shift toward the style of ChatGPT-generated content. Compared with their pre-ChatGPT writing, the differences in these metrics are reduced by 4.807 (length), 0.0579 (polarity), 0.100 (subjectivity), and 1.868 (readability), indicating a lasting impact of ChatGPT on their writing style even after its usage is restricted. This suggests that users effectively learn writing styles from ChatGPT.

To further substantiate the learning and cognitive burden mechanisms, we also sent surveys to Stack Overflow users to understand their interactions with ChatGPT. The detailed process is shown in Online Appendix L.3. Our findings, presented in Table L.3, indicate that more than 75% of Stack Overflow users and more than 54% of Amazon Mechanical Turk users responded affirmatively to experiencing learning effects and cognitive burden either "Always" or "Often." These results offer robust confirmation of both the learning and cognitive burden mechanisms associated with ChatGPT usage.

Finally, we examined users' writing patterns across periods before the ChatGPT launch, during usage, and post-ban. The results, presented in Figure L.1 in Online Appendix L, indicate notable shifts in users' writing patterns, such as answer length and subjectivity after the ban, which tend to align more closely with their patterns before the ChatGPT launch, albeit with larger values mirroring those observed during ChatGPT usage. Additionally, answers' polarity and readability post-ban also shifted toward patterns exhibited during ChatGPT usage. These findings strongly suggest that users learn from ChatGPT in generating answers, as evidenced by the observed alterations in their writing patterns over time.

6.5. Additional Analysis

In addition to the major findings presented in this paper, we conducted further analyses, including an examination of the impact of the ChatGPT usage ban policy on user behavior and the incorporation of additional measures for assessing answer quality. Detailed information is provided in Online Appendix M. Specifically, our findings indicate that generative AI usage leads users to produce a greater volume of answers characterized by shorter and more easily readable content. Although the use of generative AI tools does not significantly alter the sentiment of the generated answers, it does have a notable impact on their subjectivity and the use of political language as well as on subsequent generative AI usage.

7. Discussion

Voluntary knowledge of online platforms is crucial for users, platforms, and firms. With the recent advances in generative AI techniques, automatically generated knowledge has become possible. The development of generative AI chatbots like ChatGPT, which are built on large language models (e.g., GPT-3.5) and reinforced by human feedback, has made it possible to generate answers automatically on Q&A platforms. However, it is unclear whether and how the use of generative AI affects users' knowledge contribution on Q&A platforms. This study examines the impact of prior experience with generative AI on users' subsequent voluntary knowledge contribution in terms of two dimensions: quantity and quality.

Using a data set from one of the biggest Q&A platforms, Stack Overflow, and employing DID econometric techniques with various methods to examine the impact, we find that the usage of generative AI does affect users' voluntary knowledge contribution. Specifically, using generative AI correlates with users uploading 16.77% more answers every day, on average. Notably, these answers are also shorter and easier to read. Additionally, we find that the cumulative usage and usage intensity can moderate the generative AI impact to different degrees, confirming the mechanism of learning and cognitive load, respectively.

7.1. Theoretical Contribution and Managerial Implications

Our research contributes to multiple streams of IS literature. First, we enhance the understanding of voluntary knowledge contribution within the IS literature. Specifically, we are among the pioneers in exploring the effects of generative AI on users' answering behavior on knowledge platforms through the learning and cognitive load mechanisms. Second, we expand the knowledge within the IS literature concerning the impact of AI on human behaviors. In particular, we focus on comprehending how the emergence of new generative AI technologies influences users' learning processes. In addition, we contribute to the emerging literature examining the impact of generative AI on voluntary knowledge contribution (see, e.g., Burtch et al. 2024, Quinn and Gutt 2025) by understanding users' learning from using ChatGPT. Hence, our research advances the understanding of the complex interplay between generative AI, voluntary knowledge contribution, and human behaviors in the context of Q&A platforms.

In addition to contributing to the academic literature, our research also holds significant managerial implications. The findings of our study can assist knowledge platform owners and managers in formulating effective policies to manage users' utilization of generative AI and improve knowledge contribution on their platforms. We have examined the impact of generative AI on both the quantity and quality of users' knowledge contributions, taking into account factors such as cumulative usage and usage intensity as moderators. These findings offer valuable insights and guidance for knowledge platform owners and managers, informing their decision-making processes and strategic approaches. By understanding the implications of generative AI and its interaction with user factors, platform stakeholders can optimize their platforms' performance and foster a more productive and engaging knowledge-sharing environment.

Our findings provide evidence that the usage of generative AI on knowledge platforms significantly influences users' voluntary knowledge contribution. We assert that knowledge platform owners and managers should acknowledge the impact of generative AI on both the quantity and quality of users' voluntary knowledge contribution. Furthermore, it is crucial for them to recognize that the impact of generative AI can vary, depending on users' cumulative usage and usage intensity. By understanding these dynamics, knowledge platform owners and managers can make informed decisions regarding policies governing the utilization of generative AI on their platforms, thereby fostering an environment that encourages and enhances users' knowledge contribution.

Our research offers several actionable managerial insights for managers of voluntary knowledge platforms. First, our findings suggest that platforms like Stack Overflow may want to reconsider their current policies banning the use of generative AI tools such as ChatGPT. Our research demonstrates that these tools can significantly enhance user engagement and content generation by streamlining the process of answering questions. Rather than prohibiting the use of generative AI tools outright, platforms could benefit from adopting more flexible policies that permit their usage under controlled conditions. For example, managers could implement guidelines that allow users to leverage ChatGPT while maintaining clear standards for content quality and originality.

Second, our study underscores the importance of the learning process in users' effective utilization of generative AI tools. Our results indicate that when users learn how to use these tools thoughtfully, they can produce answers that are not only concise but also easier to understand, which benefits the entire community. To capitalize on this, platform managers should cultivate a culture of learning among users. Managers might create initiatives encouraging users to share their experiences and best practices for utilizing AI tools, thereby helping others to learn more efficiently. This peer-learning approach could lead to a virtuous cycle of improved answer quality and greater user engagement.

Third, our research highlights the cognitive load that can be associated with the intensive use of generative AI tools, which may lead to diminished user performance over time. To mitigate these potential negative effects, platform managers should encourage moderation in the use of AI tools like ChatGPT. Rather than imposing a blanket ban, platforms such as Stack Overflow could implement guidelines that encourage users to reduce their daily usage intensity of generative AI tools, thus preventing cognitive fatigue and promoting sustained, high-quality contributions. For instance, platforms could introduce gentle reminders that nudge users toward mindful, moderated interaction with AI tools.

In conclusion, although Stack Overflow currently prohibits the use of ChatGPT, this policy may inadvertently stifle users' opportunities to learn and improve their responses through AI-assisted tools. Instead of a complete ban, platforms could strike a balance by allowing AI usage while promoting responsible and moderated use. This approach would enable users to benefit from the efficiency and learning opportunities while ensuring that content quality and user cognition are not compromised.

7.2. Limitations and Future Research

We conducted a study to examine the impact of generative AI on users' voluntary contribution quantity and quality on the Stack Overflow platform. However, our research does have certain limitations that should be acknowledged. First, our study focused solely on Stack Overflow, which might restrict the generalizability of our findings to other knowledge platforms. Future research endeavors could incorporate data from diverse knowledge platforms that have implemented generative AI over a more extended period. This would provide a broader perspective and facilitate a more comprehensive understanding of the effects of generative AI on users' answer contributions.

Second, we explored the impact of generative AI usage on users' knowledge contribution behavior on the platform. Future research could delve deeper into the differences between generative AI-generated and human-generated answers through an answer-level analysis.

Third, whereas our research focused primarily on the adoption of generative AI in generating answers, future studies could investigate its adoption in generating questions and its potential effects on users' questioning behavior on voluntary knowledge platforms. As an initial step, we examined the impact of generative AI adoption in generating answers on questioning behavior in Online Appendix N. The results indicate that using ChatGPT to generate answers significantly increases the number and length of questions asked but decreases question bookmark votes and does not affect question readability.

Fourth, ChatGPT was permitted for use for only five days on the Stack Overflow platforms, which may impose limitations. The effects we observed may be transient, and the learning mechanisms we elucidated may apply only within this short timeframe. Therefore, future research could validate our findings and mechanisms using a longer treatment period.

Fifth, our treatment identification algorithm achieved an accuracy of 83.02%, falling short of 100%. Hence, future research could employ randomized controlled experiments to manipulate generative AI usage and corroborate the impacts and findings from our study. Additionally, we identified ChatGPT-generated answers by comparing user-submitted answers with responses generated by GPT using a similarity score. Although this method offers a practical approach for large-scale detection, we recognize that it has inherent limitations. In particular, users may significantly modify AI-generated content before posting it, which can reduce the similarity score and make detection more difficult. Although we employed a 0.9 similarity threshold to allow for reasonable edits while still capturing substantial reliance on GPT-generated content, we acknowledge that this approach may not detect all instances of AI-assisted writing, especially when users make extensive revisions. Therefore, we consider this a methodological limitation of our study and encourage future research to develop more sophisticated and precise techniques for identifying AI-generated or AI-assisted content. Such methods could incorporate linguistic patterns, behavioral signals, or metadata to more accurately distinguish between original and AI-influenced responses.

Sixth, it is important to acknowledge that in our study, we focused on examining the moderating effects of cumulative usage and usage intensity to unpack the mechanisms of learning and cognitive load. However, there may be other potential mechanisms that could explain the impact of generative AI on users' knowledge contribution. For instance, users' answering habits could be a moderator to explore in future research. Investigating additional moderators would provide a more comprehensive understanding of the underlying mechanisms shaping the impact of generative AI. By considering a wider range of moderators, we can gain deeper insights into the factors that influence users' knowledge contribution when utilizing generative AI.

Seventh, this paper employed usage intensity to investigate the effects of cognitive load. Future research could design laboratory experiments utilizing physiological measures, such as eye-tracking or electrodermal activity devices, to gain deeper insights into users' cognitive activity while interacting with ChatGPT. An alternative way to strengthen the measurement of intensity would be to identify periods during which users relied exclusively on ChatGPT without submitting any organic (i.e., manually written) answers. By focusing on these periods, researchers could more precisely isolate the effects of intensive ChatGPT usage. Demonstrating that higher intensity within such periods is associated with a reduction in the number of answers would provide stronger evidence that short-term, high-frequency reliance on ChatGPT may diminish overall productivity, potentially because of increased cognitive load or fatigue.

Eighth, our findings could establish stronger causality by fully addressing the simultaneity issue, where users may adjust their ChatGPT usage based on their answering behavior on a given day. Therefore, we encourage future research to explore this issue further to enhance the robustness and causality of our findings. Ninth, it is important to acknowledge that some users may directly copy and paste answers generated by GPT onto Stack Overflow without significant modification. In such cases, it becomes challenging to observe or measure any genuine learning on the part of the user because the content reflects AI output rather than the user's own understanding or knowledge acquisition. Therefore, the learning mechanism examined in this study is most applicable to scenarios where users actively adapt, modify, or build upon GPT-generated answers. We recognize this limitation and encourage future research to investigate the learning processes associated with users who directly copy and paste GPT-generated content. Exploring how such behavior influences user learning, community dynamics, and content quality could provide valuable insights into the broader implications of AI-assisted answer generation on Q&A platforms. Lastly, we acknowledge a limitation of our current study concerning the cognitive load mechanism. This research focuses primarily on hypothesizing and testing the increased cognitive load associated with extensive ChatGPT usage. However, as an assistant, AI can also reduce users' cognitive load (Jia et al. 2024). Therefore, we encourage future research to investigate the mechanisms underlying cognitive load reduction resulting from ChatGPT use.

Acknowledgments

The authors sincerely appreciate the thoughtful feedback and constructive comments provided by the editors (Prof. Wonseok Oh and Prof. Khim Yong Goh) and three anonymous reviewers. Additionally, the authors thank Prof. Jingchuan Pu for reviewing their first version and providing helpful comments.

References

Abdullah U (2023) How accurate is ChatGPT in generating responses? PC guide (March 7), https://www.pcguide.com/apps/how-accurate-is-chat-gpt/.

Adamopoulos P, Ghose A, Todri V (2018) The impact of user personality traits on word of mouth: Text-mining social media platforms. *Inform. Systems Res.* 29(3):612–640.

Adams GS, Converse BA, Hales AH, Klotz LE (2021) People systematically overlook subtractive changes. *Nature* 592(7853):258–261.

Alcántara A (2023) Is it human or AI? New tools help you spot the bots. *Wall Street Journal* (January 10), https://www.wsj.com/articles/is-it-human-or-ai-new-tools-help-you-spot-the-bots-11673356404.

Alyakoob M, Rahman MS (2022) Shared prosperity (or lack thereof) in the sharing economy. *Inform. Systems Res.* 33(2):638–658.

Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci USA* 106(51):21544–21549.

Athey S, Imbens GW (2022) Design-based analysis in difference-in-differences settings with staggered adoption. *J. Econom.* 226(1):62–79.

Baidoo-Anu D, Owusu Ansah L (2023) Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Preprint, submitted April 8, http://dx.doi.org/10.2139/ssrn.4337484.

Baker M, Reimann P (2024) Editorial notes: Revitalising foundational concepts. *Int. J. Comput. Support. Collab. Learn.* 19(4):395–400.

Baker AC, Larcker DF, Wang CC (2022) How much should we trust staggered difference-in-differences estimates? *J. Financial Econom.* 144(2):370–395.

Bandura A, Walters RH (1977) *Social Learning Theory*, vol. 1 (Prentice Hall, Englewood Cliffs, NJ).

Barki H, Titah R, Boffo C (2007) Information system use–Related activity: An expanded behavioral conceptualization of individual-level information system use. *Inform. Systems Res.* 18(2):173–192.

Bauer K, von Zahn M, Hinz O (2023) Expl (AI) ned: The impact of explainable artificial intelligence on users' information processing. *Inform. Systems Res.* 34(4):1582–1602.

Blackwell M, Iacus S, King G, Porro G (2009) CEM: Coarsened exact matching in Stata. *Stata J.* 9(4):524–546.
Blohm I, Riedl C, Füller J, Leimeister JM (2016) Rate or trade? Identifying winning ideas in open idea sourcing. *Inform. Systems Res.* 27(1):27–48.
Borwankar S, Khern-Am-Nuai W, Kannan K (2023) Unraveling the impact: An empirical investigation of ChatGPT's exclusion from Stack Overflow. Preprint, submitted October 4, https://ssrn.com/abstract=4481959.
Bowman E (2023) A college student created an app that can tell whether AI wrote an essay. *NPR* (January 9), https://www.npr.org/2023/01/09/1147549845/gptzero-ai-chatgpt-edward-tian-plagiarism.
Boyaci T, Canyakmaz C, de Véricourt F (2024) Human and machine: The impact of machine input on decision making under cognitive limitations. *Management Sci.* 70(2):1258–1275.
Brockman G, Eleti A, Georges E, Jang J, Kilpatrick L, Lim R, Miller L, Pokrass M (2024) Introducing ChatGPT and Whisper APIs. *OpenAI* (April 24), https://openai.com/blog/introducing-chatgpt-and-whisper-apis.
Brynjolfsson E, Li D, Raymond LR (2025) Generative AI at work. *Quart. J. Econom.* 140(2):889–942.
Burtch G, Carnahan S, Greenwood BN (2018) Can you gig it? An empirical examination of the gig economy and entrepreneurial activity. *Management Sci.* 64(12):5497–5520.
Burtch G, Lee D, Chen Z (2024) The consequences of generative AI for online knowledge communities. *Sci. Rep.* 14(1):10413.
Burtch G, He Q, Hong Y, Lee D (2022) How do peer awards motivate creative content? Experimental evidence from Reddit. *Management Sci.* 68(5):3488–3506.
Cao Z, Zhu Y, Li G, Qiu L (2024) Consequences of information feed integration on user engagement and contribution: A natural experiment in an online knowledge-sharing community. *Inform. Systems Res*. 35(3):1114–1136.
Cengiz D, Dube A, Lindner A, Zipperer B (2019) The effect of minimum wages on low-wage jobs. *Qurart. J. Econom.* 134(3):1405–1454.
Chen W, Wei X, Zhu KX (2018) Engaging voluntary contributions in online communities: A hidden Markov model. *MIS Quart.* 42(1):83–100.
Chen Y, Cheng HK, Liu Y, Pu J, Qiu L, Wang N (2022) Knowledge-sharing ties and equivalence in corporate online communities: A novel source to understand voluntary turnover. *Production Oper. Management* 31(10):3896–3913.
Cheng Z, Pang MS, Pavlou PA (2020) Mitigating traffic congestion: The role of intelligent transportation systems. *Inform. Systems Res.* 31(3):653–674.
Cohen J (1988) Set correlation and contingency tables. *Appl. Psychol. Meas.* 12(4):425–434.
David C (2023) Stack Overflow growth and usage statistics (2023). *SignHouse* (January 17), https://www.usesignhouse.com/blog/stack-overflow-stats.
De Chaisemartin C, D'Haultfoeuille X (2022) Difference-in-differences estimators of intertemporal treatment effects. National Bureau of Economic Research, Cambridge, MA.
Dempster MAH, Leemans V (2006) An automated FX trading system using adaptive reinforcement learning. *Expert Syst. Appl.* 30(3):543–552.
Dowling M, Lucey B (2023) ChatGPT for (finance) research: The Bananarama conjecture. *Financial Res. Lett.* 53(2023):103662–103662.
Elder D (2025) Balancing creators and consumers: Mastering the 1% rule in online spaces. *Bettermode* (March 14), https://bettermode.com/blog/one-percent-rule-online-communities.
Fügener A, Grahl J, Gupta A, Ketter W (2022) Cognitive challenges in Human–Artificial Intelligence Collaboration: Investigating the path toward productive delegation. *Inform. Systems Res.* 33(2):678–696.
Ge R, Zheng Z, Tian X, Liao L (2021) Human–Robot interaction: When investors adjust the usage of robo-advisors in peer-to-peer lending. *Inform. Systems Res.* 32(3):774–785.
Goodman-Bacon A (2021) Difference-in-differences with variation in treatment timing. *J. Econom.* 225(2):254–277.
Gopal A, Sivaramakrishnan K (2008) Research note—On vendor preferences for contract types in offshore software projects: The case of fixed price vs. time and materials contracts. *Inform. Systems Res.* 19(2):202–220.
Guo B, Zhang X, Wang Z, Jiang M, Nie J, Ding Y, Yue J, Wu Y (2023) How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection. Preprint, submitted January 18, https://arxiv.org/abs/2301.07597.
He S, Qiu L, Cheng X (2022) Surge pricing and short-term wage elasticity of labor supply in real-time ridesharing markets. *MIS Quart.* 46(1):193–228.
Heckman JJ (1979) Sample selection bias as a specification error. *Econom. J. Econometric Soc.* 47(1):153–161.
Hong X, Guan SU, Wong PW, Xue N, Man KL, Liu D (2021) Can AI teach humans? Humans AI collaboration for lifelong machine learning. *Int. Conf. Data Sci. Inform. Tech*. (Association for Computing Machinery, New York), 427–432.
Huang P, Zhang Z (2016) Participation in open knowledge communities and job-hopping: Evidence from enterprise software. *MIS Quart.* 40(3):785–806.
Huang P, Tafti A, Mithas S (2018) Platform sponsor investments and user contributions in knowledge communities: The role of knowledge seeding. *MIS Quart.* 42(1):213–240.
Huang P, Ceccagnoli M, Forman C, Wu DJ (2022) IT knowledge spillovers, absorptive capacity, and productivity: Evidence from enterprise software. *Inform. Systems Res.* 33(3):908–934.
Huang N, Wang L, Hong Y, Lin L, Guo X, Chen G (2024) When the clock strikes: A multimethod investigation of on-the-hour effects in online learning. *Inform. Systems Res.* 35(2):766–782.
Jia N, Luo X, Fang Z, Liao C (2024) When and how artificial intelligence augments employee creativity. *Acad. Management J.* 67(1):5–32.
John K, Qi Q, Wang J (2020) Bank integration and the market for corporate control: Evidence from cross-state acquisitions. *Management Sci.* 66(7):3277–3294.
Johnson SL, Safadi H, Faraj S (2015) The emergence of online community leadership. *Inform. Systems Res.* 26(1):165–187.
Jussupow E, Spohrer K, Heinzl A, Gawlitza J (2021) Augmenting medical diagnosis decisions? An investigation into physicians' decision-making process with artificial intelligence. *Inform. Systems Res.* 32(3):713–735.
Kirchner JH (2023) New AI classifier for indicating AI-written text. *OpenAI* (January 31), https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text.
Kumar S, Qiu L (2021) *Social Media Analytics and Practical Applications: The Change to the Competition Landscape* (CRC Press, Boca Raton, FL).
Kumar N, Qiu L, Kumar S (2018) Exit, voice, and response on digital platforms: An empirical investigation of online management response strategies. *Inform. Systems Res.* 29(4):849–870.
Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, et al. (2023) Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit. Health* 2(2):e0000198.
Kwon HE, Dewan S, Oh W, Kim T (2023) Self-regulation and external influence: The relative efficacy of mobile apps and offline channels for personal weight management. *Inform. Systems Res.* 34(1):50–66.
Lee SY, Rui H, Whinston AB (2019) Is best answer really the best answer? The politeness bias. *MIS Quart.* 43(2):579–600.
Li X (2016) Could deal promotion improve merchants' online reputations? The moderating role of prior reviews. *J. Management Inform. Systems* 33(1):171–201.

Li X, Hsieh JPA, Rai A (2013) Motivational differences across post-acceptance information system usage behaviors: An investigation in the business intelligence systems context. *Inform. Systems Res.* 24(3):659–682.

Li S, Shang J, Slaughter SA (2010) Why do software firms fail? Capabilities, competitive actions, and firm survival in the software industry from 1995 to 2007. *Inform. Systems Res.* 21(3):631–654.

Liu L, Wang Y, Xu Y (2024) A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *Amer. J. Political Sci.* 68(1):160–176.

Liu X, Wang GA, Fan W, Zhang Z (2020) Finding useful solutions in online knowledge communities: A theory-driven design and multilevel analysis. *Inform. Systems Res.* 31(3):731–752.

Liu Y, Pu J, Chen Y, Qiu L, Cheng HK (2025) Departmental boundaries and knowledge sharing in corporate online communities. *Inform. Systems Res.*, ePub ahead of print June 9, https://doi.org/10.1287/isre.2023.0086.

Lysyakov M, Viswanathan S (2022) Threatened by AI: Analyzing users' responses to the introduction of AI in a crowd-sourcing platform. *Inform. Systems Res.* 34(3):1191–1210.

Ma M, Agarwal R (2007) Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Inform. Systems Res.* 18(1):42–67.

Marton F (2006) Sameness and difference in transfer. *J. Learn. Sci.* 15(4):499–535.

Moqri M, Mei X, Qiu L, Bandyopadhyay S (2018) Effect of "following" on contributions to open source communities. *J. Management Inform. Systems* 35(4):1188–1217.

Nerdynav (2023) 91 important ChatGPT statistics & user numbers in April 2023 (GPT-4, plugins update). Accessed July 18, 2025, https://nerdynav.com/chatgpt-statistics/.

OpenAI (2022) Introducing ChatGPT. *OpenAI* (November 30), https://openai.com/blog/chatgpt.

Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: Recent developments. *Educ. Psychol.* 18(1):1–4.

Pamuru V, Khern-Am-Nuai W, Kannan K (2021) The impact of an augmented reality game on local businesses: A study of Pokémon Go on restaurants. *Inform. Systems Res.* 32(3):950–966.

Pan Y, Qiu L (2022) How ride-sharing is shaping public transit system: A counterfactual estimator approach. *Production Oper. Management* 31(3):906–927.

Pu J, Chen Y, Qiu L, Cheng HK (2020) Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Inform. Systems Res.* 31(2):297–322.

Pu J, Liu Y, Chen Y, Qiu L, Cheng HK (2022) What questions are you inclined to answer? Effects of hierarchy in corporate Q&A communities. *Inform. Systems Res.* 33(1):244–264.

Qiao D, Lee SY, Whinston AB, Wei Q (2020) Financial incentives dampen altruism in online prosocial contributions: A study of online reviews. *Inform. Systems Res.* 31(4):1361–1375.

Qiu L, Kumar S (2017) Understanding voluntary knowledge provision and content contribution through a social-media-based prediction market: A field experiment. *Inform. Systems Res.* 28(3):529–546.

Qiu L, Cheng HK, Pu J (2017) Hidden profiles in corporate prediction markets: The impact of public information precision and social interactions. *MIS Quart.* 41(4):1249–1273.

Quinn M, Gutt D (2025) Heterogeneous effects of generative AI on knowledge seeking in online communities. *J. Management Inform. Systems* 42(2):370–399.

Sanatizadeh A, Lu Y, Zhao K, Hu Y (2025) Information foraging in the era of AI: Exploring the effect of ChatGPT on digital Q&A platforms. *Inform. Management* 62(6):104178.

Schanke S, Burtch G, Ray G (2021) Estimating the impact of "humanizing" customer service chatbots. *Inform. Systems Res.* 32(3):736–751.

Shearer A (2022) I used ChatGPT to answer questions on Stack Overflow. *DEV* (December 4), https://dev.to/andrews1022/ive-been-using-chatgpt-to-answer-questions-on-stack-overflow-2ib3.

Shi C, Hu P, Fan W, Qiu L (2021) How learning effects influence knowledge contribution in online Q&A community? A social cognitive perspective. *Decision Support Systems* 149:113610.

Shi C, Hu P, Fan W, Qiu L (2024) Competitive peer influence on knowledge contribution behaviors in online Q&A communities: A social comparison perspective. *Internet Res.* 34(5):1577–1601.

Solaiman I, Brundage M, Clark J, Askell A, Herbert-Voss A, Wu J, Radford A, et al. (2019) Release strategies and the social impacts of language models. Preprint, submitted August 24, https://arxiv.org/abs/1908.09203.

Song T, Huang J, Tan Y, Yu Y (2019) Using user-and marketer-generated content for box office revenue prediction: Differences between microblogging and third-party platforms. *Inform. Systems Res.* 30(1):191–203.

StackOverflow (2022) Temporary policy: ChatGPT is banned. *Stack Overlfow* (December 5), https://meta.stackoverflow.com/questions/421831/temporary-policy-chatgpt-is-banned.

Stewart OG, Jordan ME (2017) "Some explanation here": A case study of learning opportunities and tensions in an informal science learning environment. *Instructional Sci.* 45(2):137–156.

Su Y, Wang Q, Qiu L, Chen R (2025a) Navigating the sea of reviews: Unveiling the effects of introducing AI-generated summaries in e-commerce. Preprint, submitted July 14, https://doi.org/10.2139/ssrn.4872205.

Su Y, Zhang K, Wang Q, Qiu L (2025b) Generative AI and human knowledge sharing: Evidence from a natural experiment. Preprint, submitted May 5, https://doi.org/10.2139/ssrn.4628786.

Susarla A, Gopal R, Thatcher JB, Sarker S (2023) The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in Information Systems. *Inform. Systems Res.* 34(2):399–408.

Tan X, Jin F, Dennis AR (2022) How appreciation and attention affect contributions to electronic networks of practice. *J. Management Inform. Systems* 39(4):1037–1063.

Temizkan O, Park S, Saydam C (2017) Software diversity for improved network security: Optimal distribution of software-based shared vulnerabilities. *Inform. Systems Res.* 28(4):828–849.

Tian S, Jin Q, Yeganova L, Lai PT, Zhu Q, Chen X, Lu Z (2024) Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief. Bioinform.* 25(1):bbad493.

Vaishya R, Misra A, Vaish A (2023) ChatGPT: Is this version good for healthcare and research? *Diabet. Metabol. Syndr. Clinical Res. Rev.* 17(4):102744.

Wang J, Li G, Hui KL (2022) Monetary incentives and knowledge spillover: Evidence from a natural experiment. *Management Sci.* 68(5):3549–3572.

Wang Q, Qiu L, Xu W (2024) Informal payments and doctor engagement in an online health community: An empirical investigation using generalized synthetic control. *Inform. Systems Res.* 35(2):706–726.

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Anal. (Oxford)* 25(1):57–76.

Xu L, Nian T, Cabral L (2020) What makes geeks tick? A study of Stack Overflow careers. *Management Sci.* 66(2):587–604.

Xue M, Cao X, Feng X, Gu B, Zhang Y (2022) Is college education less necessary with AI? Evidence from firm-level labor structure changes. *J. Management Inform. Systems* 39(3):865–905.

Xue J, Wang L, Zheng J, Li Y, Tan Y (2023) Can ChatGPT kill user-generated Q&A platforms? Preprint, submitted May 21, http://dx.doi.org/10.2139/ssrn.4448938.

Yan Z, Kuang L, Qiu L (2022) Prosocial behaviors and economic performance: Evidence from an online mental healthcare platform. *Production Oper. Management* 31(10):3859–3876.

Yaraghi N, Du AY, Sharman R, Gopal RD, Ramesh R (2015) Health information exchange as a multisided platform: Adoption, usage, and practice involvement in service co-production. *Inform. Systems Res.* 26(1):1–18.

Yilmaz ED, Naumovska I, Aggarwal VA (2023) AI-driven labor substitution: Evidence from Google translate and ChatGPT. Preprint, submitted June 7, http://dx.doi.org/10.2139/ssrn.4400516.

Yin D, Bond SD, Zhang H (2021) Anger in consumer reviews: Unhelpful but persuasive? *MIS Quart.* 45(3):1059–1086.

You S, Yang CL, Li X (2022) Algorithmic versus human advice: Does presenting prediction performance matter for algorithm appreciation? *J. Management Inform. Systems* 39(2):336–365.

Zhai X, Nyaaba M, Ma W (2025) Can generative AI and ChatGPT outperform humans on cognitive-demanding problem-solving tasks in science? *Sci. Ed.* 34(2):649–670.

Zhang W, Li AW, Wu C (2025) University students' perceptions of using generative AI in translation practices. *Instr. Sci.* 53(2025):633–655.

Zhang D, Jiang H, Qiang M, Zhang K, Qiu L (2025) Time to stop? An empirical investigation on the consequences of canceling monetary incentives on a digital platform. *Inform. Systems Res.* 36(2):781–801.

Zhou E, Lee D (2024) Generative artificial intelligence, human creativity, and art. *PNAS Nexus* 3(3):1–8.

Copyright of Information Systems Research is the property of INFORMS: Institute for Operations Research & the Management Sciences and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.