



# Engage Wider Audience or Facilitate Quality Answers? a Mixed-methods Analysis of Questioning Strategies for Research Sensemaking on a Community Q&A Site

CHANGYANG HE, Hong Kong University of Science and Technology, China

YUE DENG, Hong Kong University of Science and Technology, China

LU HE, University of Wisconsin-Milwaukee, USA

QINGYU GUO, Hong Kong University of Science and Technology, China

YU ZHANG, City University of Hong Kong, China

ZHICONG LU, City University of Hong Kong, China

BO LI, Hong Kong University of Science and Technology, China

Discussing research-sensemaking questions on Community Question and Answering (CQA) platforms has been an increasingly common practice for the public to participate in science communication. Nonetheless, how users strategically craft research-sensemaking questions to engage public participation and facilitate knowledge construction is a significant yet less understood problem. To fill this gap, we collected 837 science-related questions and 157,684 answers from Zhihu, and conducted a mixed-methods study to explore user-developed strategies in proposing research-sensemaking questions, and their potential effects on public engagement and knowledge construction. Through open coding, we captured a comprehensive taxonomy of question-crafting strategies, such as eye-catching narratives with counter-intuitive claims and rigorous descriptions with data use. Regression analysis indicated that these strategies correlated with user engagement and answer construction in different ways (e.g., emotional questions attracted more views and answers), yet there existed a general divergence between wide participation and quality knowledge establishment, when most questioning strategies could not ensure both. Based on log analysis, we further found that collaborative editing afforded unique values in refining research-sensemaking questions regarding accuracy, rigor, comprehensiveness and attractiveness. We propose design implications to facilitate accessible, accurate and engaging science communication on CQA platforms.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: science communication, Q&A sites, language style, user engagement, knowledge construction

## ACM Reference Format:

Changyang He, Yue Deng, Lu He, Qingyu Guo, Yu Zhang, Zhicong Lu, and Bo Li. 2024. Engage Wider Audience or Facilitate Quality Answers? a Mixed-methods Analysis of Questioning Strategies for Research Sensemaking on a Community Q&A Site. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 50 (April 2024), 31 pages. <https://doi.org/10.1145/3637327>

Authors' addresses: Changyang He, Hong Kong University of Science and Technology, Hong Kong SAR, China, [cheai@cse.ust.hk](mailto:cheai@cse.ust.hk); Yue Deng, Hong Kong University of Science and Technology, Hong Kong SAR, China, [y dengbi@cse.ust.hk](mailto:y dengbi@cse.ust.hk); Lu He, University of Wisconsin-Milwaukee, Milwaukee, USA, [he32@uwm.edu](mailto:he32@uwm.edu); Qingyu Guo, Hong Kong University of Science and Technology, Hong Kong SAR, China, [qguoag@connect.ust.hk](mailto:qguoag@connect.ust.hk); Yu Zhang, City University of Hong Kong, Hong Kong, China, [yui.zhang@my.cityu.edu.hk](mailto:yui.zhang@my.cityu.edu.hk); Zhicong Lu, City University of Hong Kong, Hong Kong, China, [zhicong.lu@cityu.edu.hk](mailto:zhicong.lu@cityu.edu.hk); Bo Li, Hong Kong University of Science and Technology, Hong Kong SAR, China, [bli@cse.ust.hk](mailto:bli@cse.ust.hk).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2573-0142/2024/4-ART50

<https://doi.org/10.1145/3637327>

## 1 INTRODUCTION

Science communication is the appropriate use of communication techniques to produce personal awareness, enjoyment, interest, opinion-forming, or understanding of science [10]. It provides a valuable means for the public to make sense of research progress and improve decision-making in everyday life [85]. Recently, the development of social media accelerated scientific knowledge dissemination and exchange. There have been some successful practices of *science communication on social media*, such as expert-written explanatory tweets [39] and science-centered videos [91, 94].

Effective science communication has increasingly put an emphasis on the participatory practice with public engagement, rather than one-way information transmission, to facilitate the knowledge construction and comprehension among the public [13, 71]. To this end, *public making sense of research through social media* is an increasingly popular and crucial process in effective and engaging science communication. Users gather to discuss and develop an understanding of research in general social media sites such as Twitter [22, 65], science-centered communities such as r/science [52] and r/AskHistorians [40], and comment threads of scientific articles [89]. With a wide user base mixing domain experts and the general public, people could broadly exchange information [52] and learn from other users' opinions [89] to conceptualize the research and draw implications from it to guide daily life (e.g., the wide COVID-19-related e-print discussion [93]).

In particular, Community Question and Answering (CQA) platforms afford a valuable channel for public sensemaking of specific research with focus questions. In this setting, *askers* propose questions regarding particular research such as its practical implications or method details, *answerers* comprising domain experts and the general public voluntarily share their interpretations with domain expertise and personal opinions, and *viewers* manage to develop their understanding from such QA-based knowledge construction [64]. We define such focus questions regarding specific research<sup>1</sup> as “**research-sensemaking questions**”, under which users could participate in the research-centered discussion and establish their understanding of the work. Such QA-based discussions naturally contextualize the focus problem with curated questions and descriptions, and facilitate knowledge exchange among askers, answerers and viewers [51]. Besides, with question-centered discussions less constrained to social networks and communities, users in different domains and with diverse levels of expertise could contribute to and benefit from QA-based science knowledge exchange [11]. On Zhihu, one of the largest CQA platforms in China [49], some research-sensemaking questions received thousands of answers and upvotes and millions of views [11].

As the component initiating discussion, the art (and science) of crafting questions in CQA platforms is significant to elicit high-quality discourse [80]. An engaging research-sensemaking question plays a crucial role in attracting public participation and promoting effective knowledge construction [62, 64]. Moreover, the establishment of research-sensemaking questions might inherit critical challenges of science communication that hinder engaging a broad audience or ensuring the discussion quality [4, 90], which further necessitates strategic and careful questioning. On the one hand, research-centered discussions typically include specialized linguistic use (e.g., terminologies and hedge words such as “likely” or “might”), which may become barriers to public participation [4]. On the other hand, loosening the constraints on rigor and accuracy to involve a wide audience inevitably brings misuse and misinterpretations of research outputs, such as exploiting science work for conspiracy theories and extremist ideology [58, 93]. To this end, **how to strategically propose research-sensemaking questions**, considering both accessibility and accuracy for public participation, is a significant yet challenging problem. Nonetheless, which strategies are naturally developed in these research-sensemaking questions to engage wide and high-quality

<sup>1</sup>In this study, the scope of research includes both hard science (e.g., biology and physics) and soft science (e.g., social science).

responses, and how they correlate with user participation and knowledge co-construction, are still less understood. Meanwhile, how collaborative question editing, an important feature of CQA platforms in enhancing question quality [12], influences the construction of research-sensemaking questions is also underinvestigated. Therefore, we propose the following research questions:

- **RQ1:** What are users' strategies in proposing research-sensemaking questions in CQA platforms?
- **RQ2:** How do users' strategies in proposing research-sensemaking questions correlate with public engagement and knowledge construction in CQA platforms?
- **RQ3:** How do users strategically apply collaborative editing to improve research-sensemaking questions in CQA platforms?

To answer these questions, we collected 837 science-related questions with 157,684 answers from the Zhihu platform, and conducted a mixed-methods study to investigate the questioning strategies and their potential effects. Through an open coding approach (RQ1), we captured users' linguistic and non-linguistic strategies in curating research-sensemaking question titles and descriptions, which reflected users' efforts for both rigorous (e.g., hedging and data use) and attractive (e.g., counter-intuitive and emotional statements) questioning. By quantifying knowledge construction in answers with text classifiers and applying regression analysis (RQ2), we systematically uncovered how these strategies correlated with public participation and knowledge construction, especially the divergence between engaging a wider audience and attracting epistemic and argumentative answers. Through inductive log analysis (RQ3), we revealed the value of collaborative work in constructing and refining research-sensemaking questions, such as co-establishing the topic scope and scientific reframing for rigorous presentations. Based on the findings, we discuss design implications for accessible, accurate, and effective research sensemaking in CQA platforms.

This work makes the following contributions to science communication in HCI and CSCW communities: (1) we deepened the understanding of questioning as an emerging pattern of user participation in science communication, and captured a comprehensive taxonomy of user-developed strategies in proposing engaging and rigorous research-sensemaking questions; (2) we revealed how different questioning strategies correlated with user engagement and knowledge construction, and unpacked the tension between wide participation and quality knowledge establishment; (3) we demonstrated the opportunities and challenges of collaborative editing in crafting research-sensemaking questions. With the trend of involving, engaging and empowering the public in research-related discussions, this work provides rich dynamics on attracting user participation and improving knowledge construction that may shed light on effective science communication.

## 2 RELATED WORK

This section contextualizes the current study within HCI and CSCW literature on science communication and CQA platforms. We first present an overview of science communication on social media in Section 2.1, outlining its existing practices and challenges on the participatory web. Under the challenge of engaging users with scientific knowledge, we survey the existing literature on communication strategies for science communication in Section 2.2. We finally situate science communication within CQA platforms in Section 2.3, navigating how the sociotechnical context of CQA platforms may shape the science communication practice.

### 2.1 Science Communication on Social Media

Science communication refers to *“the use of appropriate skills, media, activities, and dialogue to produce one or more of the following personal responses to science: Awareness, Enjoyment, Interest,*

*Opinion-forming, and Understanding*” [10]. With the development of social media, science communication has gradually evolved from one-way information dissemination to two-way participatory interactions [52, 90], which focus more on cultivating public awareness of science through dialogue [71] and public engagement [13]. Scholars in HCI and CSCW have begun to investigate how different social media platforms afford science communication, such as general social media sites (e.g., Twitter) [39, 42], online communities (e.g., Reddit) [4, 42, 52], and video sharing platforms (e.g., Youtube) [91, 94].

Characterized by two-way participatory interactions, diverse *practices* of science communication emerged on social media, informing and engaging the public with science knowledge. For example, *publicizing one’s own research* has been widely adopted by researchers on social media sites like Twitter [39, 54]. By sharing their recent or ongoing work, researchers exchange knowledge with peers and colleagues as well as establish academic connections, which benefit them in developing research [20, 90]. This process may also promote the impact of work by attracting users’ attention to the research progress with comprehensible explanations (e.g., well-crafted explanatory Twitter threads as “tweetorials”) [39]. *Science popularization* on social media, the summarization, simplification, and interpretation of specific scientific topics, has also become a common practice such as scientific blogs [38, 66] and videos [91, 94]. The participatory interactions, such as user-generated comments, make science popularization on social media more accessible when they help to facilitate others’ understanding and establish the feedback mechanism on the content quality [46, 94]. Apart from science communication with a central communicator, users also gather on social media to *collectively make sense of research*. For example, online science communities (e.g., r/science [4, 52] and r/AskHistorians [40, 42]) attract group members who share similar interests yet have diverse levels or focuses of expertise, which provide an effective channel of peer learning [52] and knowledge co-construction [47, 52].

On the other hand, science communication on social media also suffers from various *challenges*, including linguistic barriers that hinder reaching and engaging audiences with less domain knowledge [4, 7, 57], difficulties in explaining complicated topics to laypeople accurately [28, 81], the potential risks of misinformation and misinterpretation [15, 58, 93], and the danger of harassment to communicators [83].

Given the opportunities and challenges of science communication on social media, it is important to understand user participation and engagement in the specific sociotechnical context, and design corresponding interfaces to support them. Nonetheless, the existing literature lying in the intersection between science communication and HCI is still scarce. There is limited understanding of QA-based science communication, especially in a non-western context [90]. This work aims to contribute to this venue by unpacking strategies of proposing researching-sensemaking questions and their potential effects in a Chinese CQA platform.

## 2.2 Science Communication Strategies

Given the importance and difficulties in communicating engaging, understandable, and accurate scientific knowledge [81, 90], effective communication strategies are crucial components in science communication on social media. In this section, we review the literature on science communication strategies from the perspective of *information* and *interaction*.

How to craft good science-related *information* for the general public, whether textual or visual, has been widely investigated by scholars in education and communication [5, 9, 33, 39, 53]. For example, Dahlstrom argued that using narratives and storytelling to communicate science with nonexpert audiences may enhance comprehension, interest, and engagement [19]. Flemming et al. revealed the value of emotionalization in science communication such as promoting knowledge gain [35]. Researchers also concluded knowledge-crafting strategies in more specific settings. For

science writing, Gero et al. summarized some good practices including *an implicit structure, an attention-grabbing lede, specific stories as a driving force, explanatory strategies like analogy and metaphor, and return-to-question conclusions*, and further explored their use on scientific Twitter threads created by experts [39]. For video-based science communication, Finkler and León proposed a SUCCESS framework of the visual rhetoric that urged for producing videos that are Simple, Unexpected, Concrete, Credible, Emotional, and Science Storytelling [33].

A growing body of work in science communication has also turned to the *interactions* among communicators and audiences, beyond building high-quality scientific information, that may contribute to effective science communication. One line of work explored enhancing and deepening scientist-public dialogue to promote science communication [13, 30, 77, 97]. For example, Zorn et al. found that dialogue between laypersons and scientists on human biotechnology could make scientists' and laypeople's attitudes toward human biotechnology converge, and increase laypeople's communicative self-efficacy [97]. Another strand of work focused on involving the public to participate in and contribute to science communication, taking the public as not merely information receivers but also knowledge contributors [46, 52, 75, 84, 89]. This process could exploit collective wisdom in constructing and making sense of scientific knowledge through opinion sharing, knowledge exchange, discussion and debates [52]. For instance, users' comments on science news or videos not only reflect the public understanding of the transmitted knowledge [27], but also influence others' perceptions and consumption of the scientific information [89].

This work contributes to the understanding of science communication strategies from both *information* and *interaction* perspectives. For *information*, we unveiled users' linguistic and non-linguistic strategies in asking and describing researching-sensemaking questions, and how they correlated with public participation and knowledge construction. For *interaction*, we investigated how collaborative work was presented in CQA platforms and facilitated science communication, including collaborative editing in questions and knowledge co-construction in answers.

### 2.3 Community Question Answering (CQA) Platforms and Science Communication

Community Question Answering (CQA) platforms, such as Quora, Stack Exchange, and Zhihu, have been crucial online information hubs that millions of users turn to for information seeking and sharing [82]. CQA platforms attract users with different experiences and expertise to participate in the QA-based discussions, and largely benefit from the "wisdom of crowds" in knowledge construction and exchange [82, 87].

As a typical and important CSCW system [44], CQA platforms have gained research attention in HCI and CSCW communities from many different perspectives. Some scholars evaluated *user behaviors* such as answering [2, 74, 92] and lurking [55] on CQA platforms. Generally, users' preferences in choosing questions to answer were rather different [92], and only a few answerers contributed a large share of answers [72]. Different users had substantially distinct activity patterns (e.g., questioning, answering, and socializing) and linguistic features, making it easy to automatically identify experts and non-experts on CQA platforms [76]. Some prior work examined the factors that influenced *information quality* and the effects of information quality on user behaviors on CQA platforms, covering both questions [3, 80, 87] and answers [45, 61]. For example, Ravi et al. found that both the question topic and length were significant predictors of question quality in CQA sites [80]. Asaduzzaman et al. revealed a set of features from unanswered questions that potentially made them hard to get answers, such as those too specific or hard to follow [3].

Another line of work focused on the dynamics of *user identity* in CQA platforms [21, 25, 43, 86], which largely shapes the ecosystem of Q&A communities and influences user engagement under the sociotechnical context. For example, Das et al. revealed that the sociotechnical mechanisms of governance on Bengali Quora would privilege certain identities and marginalize others regarding

linguistic practices, nationalities, and religious affiliations [21]. In the same vein, Dubois et al. found that Q&A community cultures contributed to gender differences in contribution styles and user appreciation [26], and Gilbert noted that the default masculine whiteness of Reddit challenged the moderation of Q&A in a history-centered academic sub-community [40]. More recent work also explored the influences of *platform-specific features* on user participation and collaboration, such as the affordances of co-editing questions [12] and flagging mechanisms [59].

Science communication on CQA platforms shares some similarities with other community-based science communication, and also has its unique characteristics. First, the QA-initiated science discussion on CQA platforms is similar to science-centered QA communities such as r/AskHistorians [40], which necessitates both high-quality questions and answers for effective science communication. Such question-answering activities extend beyond one-off answers to a knowledge co-creation process, developing long-lasting value [2]. Also, the science knowledge co-construction in public-generated answers is a typical example of argumentative knowledge construction in computer-supported collaborative learning (CSCL) [88]. Therefore, this work followed the framework proposed by Weinberger and Fischer to analyze user participation and quality knowledge construction in answers, including dimensions of *participation*, *epistemic*, *argumentative*, and *social mode* [88].

On the other hand, compared to science-related subreddits with community members as the major participators [4, 40, 52], science communication on CQA platforms is more question-oriented and less constrained by the community size, focus, and norms, thus demanding careful curation of questions to make them attractive and accessible to wider users. That necessitates the understanding of well-crafted question-asking strategies to attract and engage the general public, which is the focus of this study. Relevant to it, Zhihu allows collaborative editing of questions to improve the question quality and make them applicable to a broader audience [12]. We also examined such collaborative work in enhancing the quality of research-sensemaking questions.

### 3 METHOD

This section describes the mixed-methods approach used to understand the questioning strategies for research-sensemaking on a CQA platform. We first briefly introduce the platform in Section 3.1, the data collection process in Section 3.2, and the dataset description in Section 3.3. Then, we present (1) how we adopted *open coding* to capture strategies in proposing research-sensemaking questions (RQ1, Section 3.4); (2) how we quantified quality knowledge construction with *text classification*, and applied *regression analysis* to figure out questioning strategies' correlations with user engagement and answer development (RQ2, Section 3.5); and (3) how we conducted *log analysis* to unpack the community's efforts to craft research-sensemaking questions through collaborative editing (RQ3, Section 3.6). The overall analytical flow is shown in Figure 1.

#### 3.1 Platform

We situated this study on Zhihu, one of the most popular CQA platforms in China. As of May 2022, Zhihu had more than 100 million average monthly active users [50] and accumulated more than 44 million questions [68]. The basic interfaces of Zhihu are similar to Quora, such as the affordances of question asking, question answering, question following, and question commenting. However, different from Quora which had disabled the “question details” feature [79], Zhihu allows users to provide question descriptions with formatted text (such as highlighted quotes, code blocks, and formulas), links, figures, and videos [43]. This feature is important for users to supplement research-sensemaking questions, typically complex in nature, with detailed explanations of research and specific questions. Besides, Zhihu also affords comprehensive features for collaborative question editing from modifying question titles and question descriptions to adding

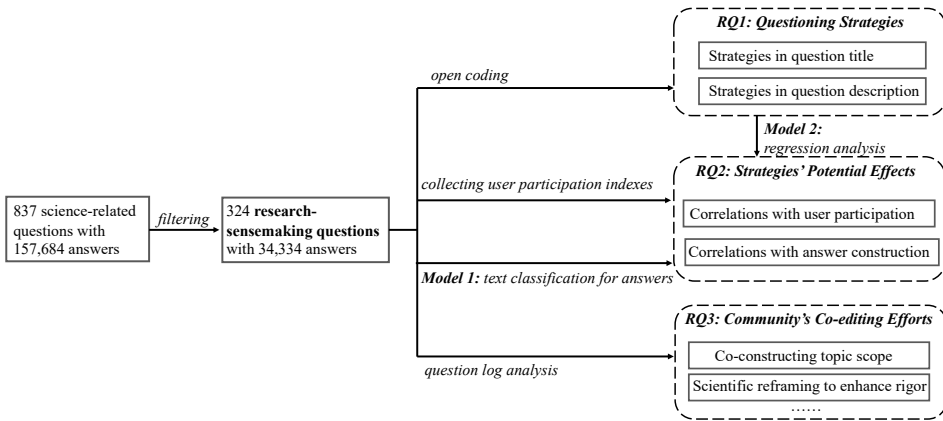


Fig. 1. The analytical flow to understand user-developed strategies in proposing research-sensemaking questions.



Fig. 2. An example of Zhihu question interface. The example research-sensemaking question is “*The latest research in The Lancet: With the basis of two doses of inactivated vaccine, the efficacy of a third dose of mRNA vaccine is 20-30% higher than that of inactivated vaccine. How to interpret it?*”

or removing question topics [12]. This interface supports collaborative work to enhance the quality of research-sensemaking questions. An example of question-related interfaces of Zhihu is shown in Figure 2.

### 3.2 Data Collection

We first conducted an exploratory scanning step to determine the data inclusion criteria for research-sensemaking questions. Our goal was to generate a filtering approach for research-sensemaking questions that were (1) unbiased in science branches, (2) inclusive to capture different ways of asking research-sensemaking questions, and (3) able to largely exclude irrelevant data. Therefore, we did not choose topic-based collection that utilized science-related topics as hashtags to collect data, as we found that many research-sensemaking questions did not have subject hashtags, and science-related topics had miscellaneous question types (e.g., the “physics” topic had many general questions similar to “how great is Einstein?” or “What is the longest formula?”). Instead, we chose keyword-based collection, and carefully designed the keyword set. To do so, we first randomly picked ten research-sensemaking questions that appeared on the “hot” page (top 50 popular questions in real-time) on the Zhihu platform in 2022. Taking these questions as the seeds, we

continuously went through “related questions” recommended by Zhihu to examine similar research-sensemaking questions, through which process we managed to summarize some general patterns that research-sensemaking questions were asked. Before adding one keyword to the keyword set, we manually checked the search results for impartiality, diversity, and purity (e.g., the keywords biased toward specific disciplines were excluded). After these steps, we reached a final keyword set to collect research-sensemaking questions, including: “*research finds*”, “*research claims*”, “*research shows*”, “*research proves*”, “*research indicates*”, and “*research reveals*”. These keywords characterized research-sensemaking questions yet were not skewed towards specific questioning strategies or science branches.

Taking the refined keywords as the search queries, we collected all questions that appeared in the search results through web scraping. Specifically, we searched every keyword on Zhihu without logging into an account to mitigate the influence of personalization [48]. We customized the search by limiting the search results to only questions (i.e., excluding articles and videos), adopting the default relevance ranking of Zhihu, and setting the period to “anytime”. The Zhihu search engine returned 100 to 200 most relevant questions for each search. Then, we used the question IDs crawled from the search page to find the question page so that we managed to collect the data of each question as well as its answers. Note that due to the finite size of returned search results on Zhihu, we could not establish a comprehensive dataset. However, the curation of the keyword set and the collection in an anonymous search setting ensured the representativeness and randomness of collected research-sensemaking questions, which constructed the basis for capturing questioning strategies.

We completed the data collection in October 2022, generating an initial dataset including 837 science-related questions with 157,684 answers. We then reviewed each question to filter out those non-research-sensemaking questions, following the criteria of (1) whether it focused on specific research papers or projects; (2) whether it raised specific questions about the scientific research. It narrowed down the dataset to 324 distinct questions with 34,334 answers, covering a 10-year range from January 2012 to September 2022. The collected metadata of questions included *question id*, *title*, *link*, *created time*, and the number of *answers*, *views*, *followers*, *comments* and *upvotes*. We further collected asker-related information including *anonymity*, *follower number* and *following number*. The metadata of answers included *answer id*, *answer content*, *created time*, *author id*, and the number of *thanks*, *comments* and *upvotes*.

### 3.3 Dataset Description

The preprocessed dataset contained 324 distinct research-sensemaking questions with 34,334 answers. The questions covered science branches including *health & medical science* (43.8%), *life science* (23.8%), *earth science & space science* (13.0%), *social science & arts* (12.7%), *math & physics* (4.3%), *computers & technology* (1.2%), and *chemistry & material science* (1.2%). A proportion of 30.9% questions aimed to make sense of research topics that were strongly *decision-related* (e.g., “*Whether to get COVID-19 vaccine boosters*”), and the remaining were less close to personal decision-making (e.g., “*Why the dinosaurs became extinct*”).

In addition to general sensemaking questions typically structured as “*How do you view/evaluate/think of/understand/interpret (the research)*”, we identified some extended *sensemaking goals*, including: (1) sensemaking and discussing implications (49.7%), e.g., “*Research found that ‘exercise actually has a very limited effect on weight loss’, so is it still meaningful to exercise to lose weight? How to lose weight scientifically?*”; (2) sensemaking and credibility assessment (9.9%), e.g., “*Research claims that the HPV vaccine may lead to increased female infertility. Is it true?*”; (3) sensemaking and reasoning (5.6%), e.g., “*How do you view the Chinese social mentality research report stating that ‘Chinese men have a stronger sense of fairness than women’? What might be the reasons?*”



The questions had an average *title length* of 37.8 Chinese characters ( $SD=8.3$ ) and an average *topic (tag) size* of 4.5 ( $SD=0.9$ ). A proportion of 68.2% questions were asked by non-anonymous users.

The aforementioned dimensions naturally characterized the questions and might intrinsically influence how users were engaged in viewing and answering (e.g., text length and hashtag numbers are potential factors that influence user engagement on social media [23, 41]). Therefore, we took these dimensions as control variables in regression analysis to be demonstrated in Section 4.2.

### 3.4 RQ1: Identifying Users' Strategies in Proposing Research-sensemaking Questions

We took an open coding approach [36] to inductively identify how users strategically propose research-sensemaking questions, letting the codes naturally emerge from the analysis. Specifically, two authors independently coded 50 initial samples of research-sensemaking questions, focusing on both linguistic and non-linguistic features that might enhance science communication and attract user engagement. In particular, the analysis examined two separate parts, including the *question title*, the brief question text that introduced the research and the focus of sensemaking; and the *question description*, which allowed detailed explanations of the research for sensemaking in formatted text, visuals, and hyperlinks. After generating initial codes, two coders took several rounds of meetings, comparisons and discussions to reach a consensus on the codebook. Example codes were “*data use*” and “*counter-intuitive statements*”. Finally, the two coders applied affinity diagramming [16] to organize the codes and group similar codes into high-level themes. Example themes were “*significance signs*” and “*eyecatching narratives*”.

To provide a quantitative description of strategies as well as prepare for regression analysis, the two coders took an annotation round on all 324 research-sensemaking questions based on the codebook, i.e., evaluating each question on whether it applied these strategies. They first re-coded the 50 initial samples independently and compared their labels on each dimension to test the inter-rater reliability. With the agreement ratio (accuracy) greater than 0.9 and Cohen's Kappa [69] greater than 0.8 for every dimension, substantial agreement was achieved between the two coders. Finally, they coded half of the remaining 274 questions separately, yielding 324 fully-labeled research-sensemaking questions. We took a similar approach to identify other basic features of questions (i.e., *science topic* and *sensemaking goal* in Section 3.3) as control variables for regression, which fundamentally characterized these questions but did not reflect strategic question construction.

### 3.5 RQ2: Investigating Questioning Strategies' Potential Effects

RQ1 uncovered a comprehensive taxonomy of user-developed strategies in asking and describing research-sensemaking questions on Zhihu. In this section, we further investigated how these research-questioning strategies correlated with user participation and knowledge co-construction through regression analysis. Specifically, we quantified *user participation* with public engagement indexes of questions (*views*, *upvotes*, *followers*, and *answers*), and used *Poisson regression* to predict these count dependent variables. We examined quality *knowledge construction* based on Weinberger and Fischer's framework of argumentative knowledge construction in CSCL [88] (i.e., quantifying proportions of *epistemic*, *argumentative*, and *social* answers for each question with text classification), and applied *Beta regression* to investigate questioning strategies' correlations with these proportion dependent variables.

**3.5.1 Dependent Variables: Measuring User Participation.** For each research-sensemaking question, we took public engagement indexes (*views*, *upvotes*, *followers*, and *answers*) as the reflection

of the general *user participation*. They measured the *quantity* of users engaged in knowledge construction [88] and research sensemaking, as described below.

- **Participation**

- *Views* (count variable): The number of users viewing the question. It reflected users' general attention to the research-sensemaking question.
- *Upvotes* (count variable): The number of users upvoting the question, which action would make the question be recommended to more users. It captured users' approval of the quality of the research-sensemaking question.
- *Followers* (count variable): The number of users following the question, which action would make the users get notified about new answers for the question. It represented users' long-term interest in the research-sensemaking question.
- *Answers* (count variable): The number of users answering the question. It showed users' contributions to the research-sensemaking question.

**3.5.2 Dependent Variables: Quantifying Knowledge Co-construction Based on Text Classification.** We built various types of binary text classifiers for answers regarding *epistemic*, *argumentative*, and *social* dimensions [88], which represented the *quality* of knowledge co-construction from different aspects, i.e., whether the answer was on-topic, well-justified, and taking others' thoughts into consideration. For each research-sensemaking question, we calculated the proportions of answers having these specific features to measure the knowledge co-construction dimensions.

- **Epistemic:** The epistemic dimension estimates how people work on the knowledge construction task they face [34], and the primary task is to examine whether users are engaging in activities to solve the task (*on-task discourse*) or rather concerned with off-task aspects [88]. Specific to the scenario of QA-based research sensemaking, we defined such *on-task discourse* as on-topic answers that contributed to the research-sensemaking question in knowledge, in contrast to off-topic digressions (e.g., just venting emotions to the research/researcher). Therefore, we applied the following dependent variable to measure the *epistemic* dimension for each question:

- *On-task discourse* (continuous variable): The proportion of on-topic answers contributing to research-sensemaking in knowledge.

To measure this variable, we trained a pairwise text classification model to identify the semantic relations for question-answer pairs (on-topic vs. off-topic). The specific process was in four steps: (1) *annotation*. we randomly sampled 1000 question-answer pairs from the whole dataset. Two authors independently coded the first 100 samples to determine whether the answer contributed to the research-sensemaking question in knowledge, and assigned the on-topic or off-topic label. The Cohen's Kappa reached 0.85 (accuracy=94%), indicating substantial agreements between coders. After several rounds of discussions to resolve the difference, the two authors further annotated 450 question-answer pairs each, generating 1000 label-assigned samples. (2) *building a pairwise text classifier for on-task discourse*. We performed the text classification using Bidirectional Encoder Representations from Transformers (BERT). We used the BERT model not only for its good performance across different NLP tasks, but also considering that the base training task of next sentence prediction in BERT supported the fine-tuning for sequence pair classification [24]. We adopted the Chinese pretrained model of BERT-wwm [18]. We used the basic structure of sequence pair classification [73], i.e., separating tokens from questions and answers with the [SEP] token, identifying the two types with a binary mask (*token\_type\_ids*), and jointly feeding them through the model. Taking 900 samples as the training set and 100 samples as the test set, the text classification achieved good performance with the F1-score = 87.7% on the

test set. (3) *predicting on-task discourse for question-answer pairs*. We applied the classifier to assign the label of *on-task discourse* for all 34,334 question-answer pairs. (4) *quantifying epistemic dimension for questions*. For each question, we calculated the percentage of answers with on-task discourse as the index of epistemic dimension (ranging from 0 to 1).

- **Argumentative:** The argumentative dimension measures the construction of arguments facing complex problems [88]. We took a simplified framework of argumentative claims to evaluate whether and how answers were argumentative [70], i.e., *evidence* and *reasoning*.
  - *Evidence* (continuous variable): The proportion of answers that provided explicit evidence that supported the claim, such as statistical data and theories.
  - *Reasoning* (continuous variable): The proportion of answers that logically justified why the claim was valid.

We also took a text classification approach to measure the two argumentative indexes. The process was similar to measuring the Epistemic dimension, except that the task became single-document classification for answers. Steps included (1) initial coding (N=100, with Cohen's Kappa = 0.86 for *evidence* and Cohen's Kappa = 0.81 for *reasoning*), discussing, and disagreement resolving; (2) two coders' annotation to generate a label-assigned dataset (N=1000); (3) building BERT-based classifiers based on 900 training samples and evaluating on a test dataset of 100 samples (F1 score = 0.86 for *evidence*, F1 score = 0.84 for *reasoning*); (4) prediction to scale up the labels to the whole answer dataset; and (5) calculating the proportion of answers with *evidence* and *reasoning* labels as the measurement of argumentative dimension for each research-sensemaking question.

- **Social** (continuous variable): Users might socialize with others to enhance knowledge co-construction, which is a crucial component in CSCL for knowledge acquisition and establishment [88]. Typical social modes in our dataset included (1) elicitation, e.g., "*According to my understanding, the logic seems to be correct? (some details) I hope some experts can explain the flaws of the article*"; (2) agreeing and supplementing, e.g., "*The main contributions of this research have already been explained by @[User]. For details, please refer to this answer: [LINK]. But I must make some elaborations on some places that are easily overlooked...*"; (3) disagreeing and rebutting, e.g., "*I object to @[User1] @[User2]'s claim that the population of the United States will continue to grow steadily...*" Considering all these potential social modes, we measured the social dimension as:
  - *Social* (continuous variable): The proportion of answers socializing with other users for knowledge exchange, discussion and establishment.

We followed the same text classification-based approach as the *evidence* and *reasoning* indexes to compute the *social* proportion for each question (Cohen's Kappa = 0.90 during the first round of coding). Due to the lower frequency compared to other dimensions, we coded a larger training set (N=1500), and evaluated 100 positive samples (recall=0.89) after prediction to ensure its practicality.

**3.5.3 Regression Analysis.** We took questioning strategies identified in Section 3.4 as the independent variables to explore how they correlated with the various dimensions of user engagement and knowledge co-construction. Considering the influence of basic question characteristics, we also included control variables covering: (1) the topic features, i.e., science branches and relevance to decision making; (2) sensemaking goals; (3) asker information, i.e., anonymity, follower and following; (4) title length and topic (tag) size, as shown in Section 3.3. We used only question title-related strategies to predict the count of *views* as users could not see the description before opening the question, and combined title and description strategies in predicting the remaining

indexes. We pre-processed the independent variables including one-hot encoding for categorical variables (e.g., science branches of the topic and goals of the question) and normalization.

We applied (1) Poisson regression for count dependent variables (i.e., *views*, *upvotes*, *followers* and *answers*), which could effectively estimate count data [17]; and (2) Beta regression for continuous dependent variables as proportions (i.e., *on-task discourse*, *evidence*, *reasoning*, and *social*), which is well-suited to dependent variable in the form of fractions or percentages [32]. As highly related features would lead to poor estimation in regression, we tested multicollinearity in features through Variance Inflation Factor (VIF) [1]. We found all features had  $VIF < 3.5$  (VIF of 5 and above generally suggests high multicollinearity and bad regression performance [1]), indicating low multicollinearity. We excluded questions with fewer than 5 answers to reduce the small-sample bias in regressions for continuous dependent variables that used proportion-based estimation. We excluded new questions asked within 1 month before the collecting date in regressions for count dependent variables as their count indexes might not reach stability.

### 3.6 RQ3: Understanding Community's Collaborative Work in Constructing Research-sensemaking Questions

RQ1 and RQ2 investigated users' strategies in proposing research-sensemaking questions on a CQA platform, and their correlations with user participation and knowledge construction. This section further examined the construction of research-sensemaking questions from the perspective of community's collaborative work, unpacking how community members collectively crafted research-sensemaking questions and improved their quality through collaborative editing.

To achieve this goal, we conducted inductive coding on question logs that recorded users' collaborative editing. Specifically, two authors independently analyzed logs of the questions that enabled collaborative editing, letting the codes naturally emerge from the analysis. For each question, they first read through all editing logs, including the specific edits and reasons for edits, from the created time to the time when the collaborative editing was locked<sup>2</sup>. They grouped edits into title edits, description edits and tag change [12], and particularly paid attention to *how the edits reflected specific question-asking strategies, and how they might improve the research-sensemaking questions*. Two authors coded till saturation when no new codes emerged, and reached a consensus through several rounds of comparisons and discussions. In total, they coded 66 questions with 536 edit logs.

## 4 FINDINGS

Through a mixed-methods approach, this work enriches the understanding of how users strategically proposed research-sensemaking questions, and the potential effects. In this section, we first present the taxonomy of user-developed strategies in crafting research-sensemaking question titles and descriptions in Section 4.1 (RQ1). We then demonstrate how these strategies correlated with user participation and quality answer construction in Section 4.2 (RQ2), which unpacked the divergence between engaging a wide audience and facilitating on-topic and argumentative discourse in answers. We finally reveal the unique values of collaborative editing from the community in constructing research-sensemaking questions in Section 4.3 (RQ3).

### 4.1 RQ1: Strategies in Proposing Research-sensemaking Questions

This section presents the taxonomy of users' strategies in asking research-sensemaking questions (Section 4.1.1) and providing detailed question descriptions (Section 4.1.2) on a Chinese CQA

<sup>2</sup>Collaborative editing of Zhihu questions would be locked when: (1) The question has received a specific number of high-quality answers; (2) The question has been collected into the "hot" page; (3) The question was identified with little room for improvement by moderators [96].

platform. The findings enlighten how users strategically crafted questions to set up for the research-sensemaking tasks and engage the audience for participation.

**4.1.1 Strategic Question Titles.** Table 1 describes the linguistic strategies of asking research-sensemaking questions along with their definitions, examples and proportions. Generally, users embed rich information in the short text of question titles to serve for both effectively capturing users' eye and accurately presenting the research information, covering themes including:

- **Significance signs.** We noticed that users commonly adopted significance signs to attract users' attention and help establish the credibility of work. The signs could be explicit as *impact indication* and *timeliness indication*, which directly used linguistic cues like "great contribution" or "the latest work" to point out the significance of research; or implicit in the way of pointing out the *publication venue* and *researcher background*, indicating the research was published in venues with a good reputation or conducted by famous institutes or scientists. Specific examples could be found in the first category of Table 1.
- **Rigorous descriptions.** We observed two prevalent practices in rigorously presenting the research findings, including adopting *hedging* to soften statements with words such as "possible" or "might", and applying *data use* to accurately describe the research. Specific examples could be found in the second category of Table 1.
- **Eyecatching narratives.** A common eyecatching practice was *quoting key findings* with quotation marks to highlight research statements for sensemaking. We also identified that users tended to craft research claims with *emotional arousal* (such as raising anxiety among the audience) or apply *counter-intuitive statements* to attract users' attention and participation. Specific examples could be found in the third category of Table 1.

Note that these strategies may not be equivalent to "good practice". For example, using emotional and counter-intuitive narratives may attract users but potentially introduce exaggerations or misrepresentations, which might bring off-topic discussions; using rigorous descriptions may unintentionally lift the linguistic barrier [4], which might exclude some users in discussion. Therefore, we investigated these strategies' associations with both participation and high-quality knowledge co-construction in answers.

**4.1.2 Strategic Question Descriptions.** Due to the complexity of research, question descriptions played an important role for askers to supplement and clarify important details in research-sensemaking questions. We found that question askers strategically crafted question descriptions to engage users to participate in answering or facilitate on-topic and high-quality knowledge construction in answers. Figure 3 presents a representative example of users' linguistic and non-linguistic strategies in constructing the description. We conclude their definitions below:

- **Supplementing additional resources.**
  - *Paper sources (38.6%)*: Add the publication source of the research to support original paper tracing. A proportion of 17.3% contained direct links to the e-print, and 21.3% provided the title of the paper in the original language to help target the publication.
  - *Supplementary materials (63.0%)*: Users frequently attached links to supplementary materials, mostly linking to second-hand explanatory articles in Chinese and sometimes providing other related work, to provide more relevant information that helped users have a basic knowledge of the research. Note that *supplementary materials* had a higher proportion than *paper sources*, indicating the more frequent use of second-hand articles translating and interpreting the paper in Chinese than the original paper. Therefore, media and some individual domain experts, as typical authors of such second-hand articles, played an important role as mediators in such cross-language knowledge communication.

Table 1. Strategic Question Titles: The codebook of user-developed strategies in asking research-sensemaking questions

| Category                      | Strategy                     | Definition  | Example   | Proportion |
|-------------------------------|------------------------------|---|---|------------|
| <b>Significance Signs</b>     | Impact indication            | Use linguistic cues to indicate the impact of the work, such as “great contribution” and “significant breakthrough”                                       | <i>Scientists have revealed the origin and evolution of jawed vertebrates, which is an important breakthrough in the “from fish to human” research. What does this imply?</i>   | 12.3%      |
|                               | Timeliness indication        | Use linguistic cues to indicate the timeliness of the work, such as “the latest work” and “a new study”   | <i>The latest research found that when humans saw food, the brain would have a short-term inflammatory response. How do you view this research?</i>   | 21.0%      |
|                               | Publication venue            | Include the (typically well-known) publication venue of the research paper to indicate the significance of work   | <i>New research in Nature finds potential evidence of the oldest animal fossil, dating back 890 million years. what are the implications?</i>   | 14.8%      |
|                               | Researcher background        | Include the background of researchers such as institutes, labs or leading scientists (that are typically well-known) to indicate the significance of work | <i>A study by the University of Copenhagen showed that global warming may cause each person to lose 58 hours of sleep per year. What is the specific situation? What are their connections?</i>                                 | 38.6%      |
| <b>Rigorous Descriptions</b>  | Hedging                      | Soften research statements, such as using words “possibly” and “sometimes”  | <i>American scientists have found that the air transmission rate of the coronavirus is <u>possibly</u> a thousand times higher than transmission through contact surfaces. How does this finding help the epidemic control?</i> | 17.9%      |
|                               | Data use                     | Use data to present conclusions or explanations of research   | <i>How to evaluate the latest research in Nature Medicine that uses large samples of real data to prove that the mortality risk of Omicron is reduced by 79%?</i>   | 27.2%      |
| <b>Eyecatching Narratives</b> | Quoting                      | Use quotation marks to highlight key statements   | <i>South Africa found a new strain of the coronavirus that “has a large number of mutations, and potentially evades body defenses”. What is the specific situation? How should humans protect themselves?</i>                   | 34.0%      |
|                               | Emotional arousal            | Trigger positive or negative emotional responses  | <i>What do you think of the research saying that “COVID-19 may cause diabetes in healthy people”? (anxiety)</i>   | 44.8%      |
|                               | Counter-intuitive statements | Propose counter-intuitive research findings (potentially exaggerated or misrepresented)   | <i>How to understand the study saying that “mental illness diagnoses are scientifically meaningless”?</i>   | 32.0%      |

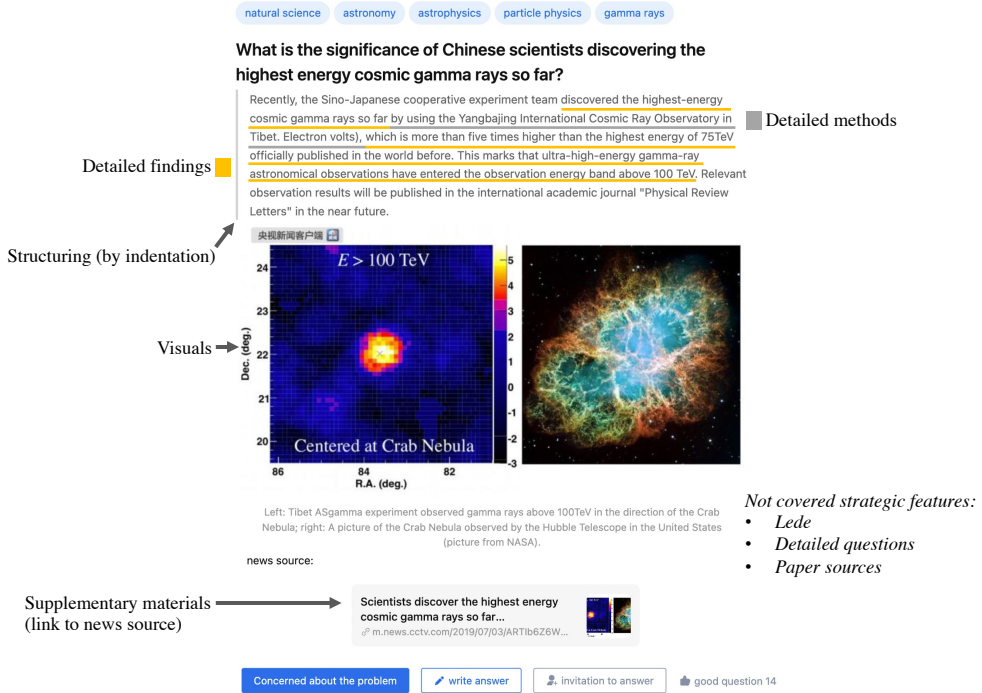


Fig. 3. An example of strategies in question descriptions. The screenshot of the research-sensemaking question was taken from Zhihu after google translation to keep the original presentations on the Zhihu platform.

- *Visuals* (9.9%): Add images or videos (adapted from the original paper, second-hand articles, or drafted by the asker) to facilitate users' comprehension of research. Images (8.0%) were more commonly used than videos (2.2%), and only one question in the sample adopted both.
- **Explaining comprehensive details.**
  - *Detailed methods* (37.3%): Introduce the major methods used in the study (example shown in Figure 3).
  - *Detailed findings* (69.4%): Detail the key findings of the study (example shown in Figure 3).
  - *Detailed questions* (9.9%): Specify research-related questions, typically more detailed than the question title. For example, an asker proposed a question titled "How do you understand the research claiming that long-term consumption of salt substitutes can reduce the incidence of cardiovascular diseases", and specified detailed questions in descriptions: "What are the health risks of a high-salt diet? What is the difference between salt substitutes and regular salt? What are the health effects of regular consumption of substitute salt containing potassium?"
- **Providing clear and engaging presentations.**
  - *Lede* (17.1%): Use several opening sentences to introduce the research and entice users to continue reading, rather than directly describe research detail, e.g., "How old is the Milky Way? How did it form and evolve? The latest research published in Nature pointed out that the Milky Way may have gone through different stages of evolution..."
  - *Structuring* (16.4%): Use quotation or indentation to show a clear description structure that eases reading and helps users get the key points, as shown in Figure 3.

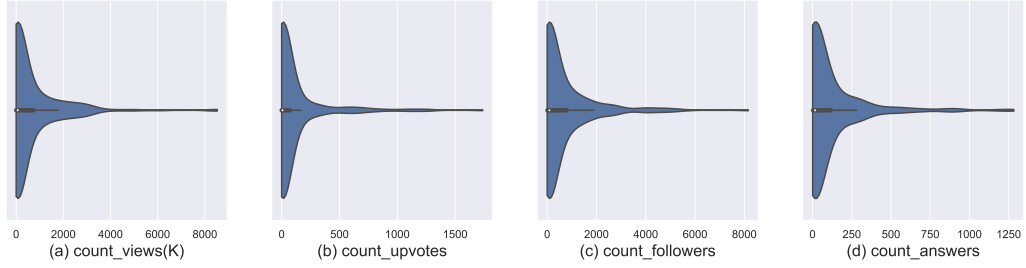


Fig. 4. The violin plots that illustrate the long-tailed distribution of user participation in research-sensemaking questions.

## 4.2 RQ2: Strategies' Correlations with Public Participation and Knowledge Construction

The findings of RQ1 presented a comprehensive taxonomy of users' strategies in asking and describing research-sensemaking questions in CQA platforms. In this section, we demonstrate how these strategies correlated with user participation and knowledge co-construction in answers. The regression analysis on *user participation* dimensions shed light on which strategies may attract more users' engagement (*views*, *upvotes*, *followers* and *answers*), and the regression analysis on *knowledge co-construction* dimensions elucidated which question-asking strategies may lead to high-quality answers for knowledge construction - the answers that were on-topic (*on-task discourse*), well-justified (*evidence* and *reasoning*), and based on socialized knowledge exchange and discussion (*social*).

**Interpreting Regression Analysis:** This section presents regression models that reveal the correlation of question-asking strategies with user participation and knowledge co-construction in Table 3 and Table 4. To ease the interpretation of effect size, we converted the regression coefficients to the ratio change of the dependent variable according to the specific link function of regression models. Particularly, for user participation dimensions, we report IRR (Incidence Rate Ratio) that denotes the rate ratio change of the dependent variable when increasing an independent variable by one unit ( $\frac{y_{x+}}{y}$ ); for knowledge co-construction dimensions, we present OR (Odds Ratio) that indicates the odds change of the dependent variable when increasing an independent variable by one unit ( $\frac{y_{x+}/(1-y_{x+})}{y/(1-y)}$ ). Therefore, both IRR and OR indicate positive correlations when their values exceed 1, with stronger positive correlations observed with larger values; conversely, IRR and OR below 1 indicate negative correlations, with stronger negative correlations observed with smaller values.

**4.2.1 Descriptive Statistics.** Table 2 presents the mean and standard deviation of dependent variables covering user participation and knowledge construction dimensions. Generally, research-sensemaking questions in the data sample attracted wide user engagement with 619,822 views and 109 answers on average. User participation dimensions also exhibited long-tailed distributions, which is demonstrated in Figure 4. Based on the text classification on fine-grained knowledge construction dimensions, we found that research-sensemaking questions typically had a high proportion of answers with on-task discourse for research-related discussion (86.4% on average). Nonetheless, only about half of the answers provided evidence such as data or theory to support their arguments, and only about 65% answers were well-justified with logical reasoning. Socialized knowledge exchange and discussion, such as supplementing or rebutting, was less commonly observed (7.3%) but was not negligible.



Table 2. Mean and standard deviation of user participation and knowledge construction dimensions for research-sensemaking questions

|         | User Participation (Count) |         |           |         | Knowledge Construction (Proportion) |          |           |        |
|---------|----------------------------|---------|-----------|---------|-------------------------------------|----------|-----------|--------|
|         | Views                      | Upvotes | Followers | Answers | On-task Discourse                   | Evidence | Reasoning | Social |
| Mean    | 619822.4                   | 100.1   | 634.8     | 108.9   | 86.4%                               | 50.7%    | 64.9%     | 7.3%   |
| Std Dev | 1151473.1                  | 234.8   | 1115.4    | 201.6   | 0.18                                | 0.28     | 0.25      | 0.13   |

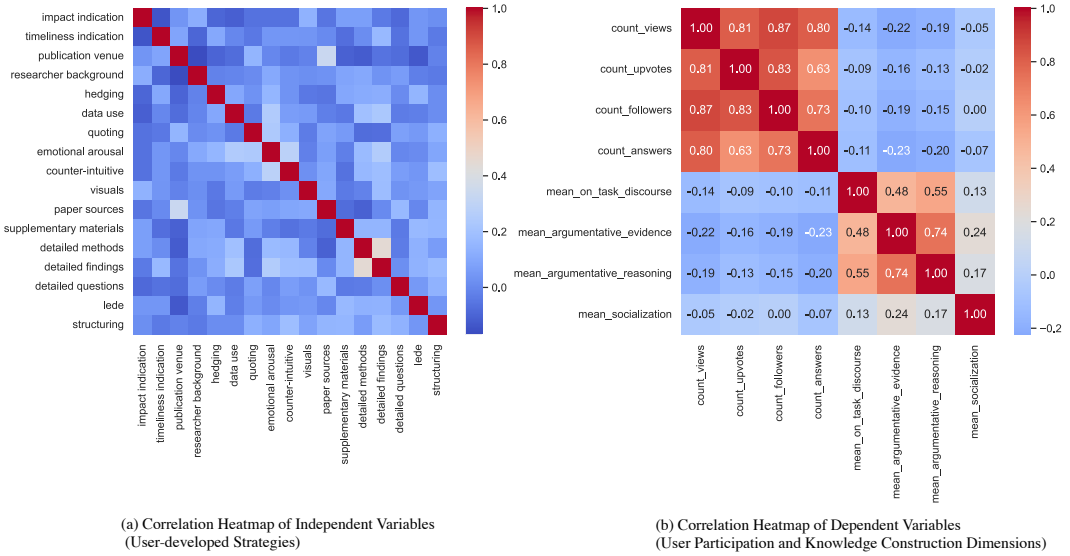


Fig. 5. The correlation heatmaps of (a) user-developed strategies (the independent variables of regression analysis). The low correlation scores suggested no exclusions of strategies in regression analysis; and (b) user participation and knowledge construction dimensions (the dependent variables of regression analysis). It demonstrated negative correlations between user participation indexes and knowledge construction dimensions.

Figure 5 demonstrates the correlation heatmaps of user-developed strategies (independent variables) and user participation and knowledge construction dimensions (dependent variables). No pair of strategies in proposing research-sensemaking questions had a correlation score greater than 0.5 as described in Figure 5 (a). Therefore, we did not exclude any strategies for regression analysis. Figure 5 (b) shows that two variables among user participation indexes or knowledge construction dimensions generally had a positive correlation, which was expected (e.g., questions with more views typically had more answers). However, it is important to note that **negative correlations were detected between user participation indexes and knowledge construction dimensions**. For instance, the correlation score between the number of answers and the proportion of evidence-supported answers was -0.23, which suggests that questions attracting more users to answer typically had a lower proportion of argumentative answers. This finding indicates the divergence between wide participation and epistemic and well-justified answer construction.

**4.2.2 Questioning Strategies' Associations with User Participation.** Table 3 shows the results of Poisson regression models predicting user participation in research-sensemaking questions across

views (*Model 1a*), upvotes (*Model 1b*), followers (*Model 1c*), and answers (*Model 1d*). We highlight the following important findings:

- Applying *emotional* and *counter-intuitive* narratives was correlated with wider user participation. Both the two narrative styles in question titles contributed to an increase of all user engagement indexes, especially for *emotional arousal* feature which was associated with 52% more views and 57% more answers. It indicated that users might be attracted by questions with claims of research findings that triggered their emotional responses (e.g., anxiety or excitement) and were contrary to common-sense expectations (e.g., some controversial statements).
- Detailing *researcher background* showed a positive correlation with wider user attention, resulting in over 20% more views and over 10% more upvotes, followers, and answers. Clarifying who conducted the research, such as the institutes of the research team and the leading scientists, might be more powerful to indicate the significance and establish the credibility compared to explicitly emphasizing the *impact* or *timeliness* of research, both of which did not show substantial promotion of user engagement.
- The factors gaining users' approval for the question quality (reflected from more "upvotes") may not align with the factors attracting more views and answers. For example, both question titles with *data narrative* and *quoting*, and question descriptions with *visual representations* and *supplementary materials*, received more upvotes from users, indicating users' appreciation of the question quality. Nonetheless, these features did not exhibit an elevation of views and answers.
- The efforts to present a *clear and engaging question description*, including adding opening sentences as *lede* and *structuring* the description with indentation, positively correlated with user engagement in the research-sensemaking question with more upvotes, followers and answers.
- Demonstrating comprehensive details of research in question descriptions may not always work to augment user engagement. For instance, questions with *detailed research methods* actually received 16% fewer answers and *detailed questions* also reduced answers by 9%, potentially due to the higher linguistic barriers that hindered user participation.
- In addition to linguistic strategies in presenting the question titles and descriptions, *research topics* also substantially correlated with user engagement (e.g., users' interest in social science-related research). Besides, questions asked by *non-anonymous* accounts with *more followers* received wider participation. Higher user engagement also manifested in questions with longer *question titles* and more attached *topics*, both of which were correlated with a 25% increase in answer numbers.

**4.2.3 Questioning Strategies' Associations with High-quality Knowledge Co-construction.** Table 4 shows the results of Beta regression models predicting knowledge co-construction dimensions in research-sensemaking questions across (1) *on-task discourse* (Model 2, epistemic dimension); (2) *evidence* and *reasoning* (Model 3a and 3b, argumentative dimension); and (3) *social* (Model 4, social dimension). The regression analysis generated the following primary findings:

- Rigorously describing scientific research may potentially attract higher proportions of quality answers. In particular, questions with *hedging* were significantly correlated with reasoning ( $OR = 1.17^{**}$ ) and social answers ( $OR = 1.15^*$ ). Providing *original paper links* of the research for sensemaking, as a valuable approach for users to find and interpret the original work, also received more high-quality answers in knowledge construction, which was positively associated with all epistemic ( $OR = 1.12$ ), argumentative ( $OR = 1.17^{**}$  for evidence and

Table 3. Poisson regression models predicting user participation in research-sensemaking questions. IRR (Incidence Rate Ratio) indicates the ratio change of the dependent variable when increasing an independent variable by one unit. We color-coded features having significantly positive correlations with all dependent variables in **green**, and those having significantly negative correlations with all dependent variables in **red**.

\*\*\*p<0.001; \*\*p<0.01; \*p<0.05.

|  |                             | M1a: Views |           | M1b: Upvotes |           | M1c: Followers |           | M1d: Answers |           |
|--|-----------------------------|------------|-----------|--------------|-----------|----------------|-----------|--------------|-----------|
|  |                             | IRR        | Std. Err. | IRR          | Std. Err. | IRR            | Std. Err. | IRR          | Std. Err. |
| <b>Strategies in asking research-sensemaking question (question title)</b>           |                             |            |           |              |           |                |           |              |           |
| Significance Signs   | impact indication           | 0.94***    | 0.002     | 0.95***      | 0.006     | 1.10***        | 0.002     | 1.01*        | 0.006     |
|  | timeliness indication       | 0.88***    | 0.003     | 0.92***      | 0.006     | 0.91***        | 0.003     | 0.84***      | 0.007     |
|  | publication venue           | 0.98***    | 0.003     | 1.01         | 0.007     | 1.05***        | 0.003     | 0.89***      | 0.007     |
|  | researcher background       | 1.23***    | 0.002     | 1.10***      | 0.006     | 1.12***        | 0.002     | 1.13***      | 0.006     |
| Rigorous Descriptions  | hedging                     | 1.01***    | 0.002     | 1.02**       | 0.006     | 1.03***        | 0.002     | 1.00         | 0.006     |
|  | data use                    | 0.89***    | 0.003     | 1.02***      | 0.006     | 1.02***        | 0.003     | 0.97***      | 0.006     |
| Eyecatching Narratives   | quoting                     | 0.96***    | 0.003     | 1.08***      | 0.006     | 1.01***        | 0.003     | 0.88***      | 0.006     |
|  | emotional arousal           | 1.52***    | 0.003     | 1.38***      | 0.007     | 1.25***        | 0.003     | 1.57***      | 0.007     |
|  | counter-intuitive           | 1.05***    | 0.002     | 1.08***      | 0.006     | 1.07***        | 0.002     | 1.08***      | 0.006     |
| <b>Strategies in describing research-sensemaking question (question description)</b> |                             |            |           |              |           |                |           |              |           |
| Additional Resources   | visuals                     | -          | -         | 1.06***      | 0.006     | 0.97***        | 0.002     | 0.92***      | 0.007     |
|  | paper sources               | -          | -         | 1.01         | 0.007     | 1.10***        | 0.003     | 1.03***      | 0.007     |
|  | supplementary materials     | -          | -         | 1.02***      | 0.006     | 0.94***        | 0.002     | 0.96***      | 0.006     |
| Comprehensive details  | detailed methods            | -          | -         | 0.84***      | 0.006     | 0.91***        | 0.003     | 0.84***      | 0.006     |
|  | detailed results            | -          | -         | 1.07***      | 0.008     | 0.92***        | 0.003     | 1.02**       | 0.008     |
|  | detailed questions          | -          | -         | 1.07***      | 0.005     | 1.06***        | 0.002     | 0.91***      | 0.006     |
| Clear presentations  | lede                        | -          | -         | 1.13***      | 0.006     | 1.09***        | 0.002     | 1.10***      | 0.006     |
|  | structuring                 | -          | -         | 1.09***      | 0.006     | 1.16***        | 0.002     | 1.18***      | 0.006     |
| <b>Control Variables</b>   |                             |            |           |              |           |                |           |              |           |
| Topic - science branch (ref: life science)   | chemistry&material          | 0.97***    | 0.003     | 0.87***      | 0.017     | 1.03***        | 0.003     | 0.84***      | 0.017     |
|  | earth&space science         | 1.06***    | 0.003     | 1.21***      | 0.008     | 1.16***        | 0.003     | 1.25***      | 0.008     |
|  | social science              | 1.12***    | 0.003     | 1.11***      | 0.008     | 1.20***        | 0.003     | 1.41***      | 0.008     |
|  | computers&technology        | 0.78***    | 0.010     | 0.90***      | 0.015     | 0.87***        | 0.006     | 0.80***      | 0.025     |
|  | health                      | 0.96***    | 0.004     | 1.18***      | 0.010     | 1.22***        | 0.004     | 1.12***      | 0.011     |
|  | math&physics                | 1.09***    | 0.002     | 1.36***      | 0.005     | 1.23***        | 0.002     | 1.18***      | 0.007     |
| Topic - decision relevance   | decision-related            | 0.96***    | 0.003     | 0.92***      | 0.007     | 0.91***        | 0.003     | 1.10***      | 0.006     |
| Goal (ref: only sensemaking)   | with assessing credibility  | 1.04***    | 0.002     | 1.04***      | 0.005     | 1.04***        | 0.002     | 0.96***      | 0.006     |
|  | with discussing implication | 0.98***    | 0.003     | 0.88***      | 0.007     | 0.96***        | 0.003     | 0.88***      | 0.007     |
|  | with reasoning              | 0.98***    | 0.002     | 0.97***      | 0.006     | 0.96***        | 0.002     | 0.99*        | 0.005     |
| Asker Info   | non-anonymous               | 1.30***    | 0.003     | 1.49***      | 0.009     | 1.32***        | 0.003     | 1.10***      | 0.007     |
|  | follower                    | 1.09***    | 0.002     | 0.93***      | 0.006     | 1.07***        | 0.002     | 1.08***      | 0.005     |
|  | following                   | 0.75***    | 0.004     | 0.61***      | 0.012     | 0.78***        | 0.003     | 0.71***      | 0.010     |
| Other Meta Info  | question length             | 1.20***    | 0.003     | 1.11***      | 0.007     | 1.22***        | 0.003     | 1.25***      | 0.007     |
|  | topic size                  | 1.31***    | 0.004     | 1.50***      | 0.011     | 1.32***        | 0.004     | 1.25***      | 0.008     |

$OR = 1.23^{**}$  for reasoning), and social dimensions ( $OR = 1.20^{*}$ ). Besides, *clear introduction (lede)* might also contribute to more on-task ( $OR = 1.20^{*}$ ) and argumentative discourse ( $OR = 1.03$  for evidence and  $OR = 1.18^{**}$  for reasoning).

- Surprisingly, adding *supplementary materials* in question descriptions correlated with more off-topic discussions ( $OR = 0.89$ ) as well as non-argumentative answers ( $OR = 0.91^{*}$  for

evidence and  $OR = 0.89^*$  for reasoning). Through manual coding of a set of supplementary materials, we noticed that some supplementary materials appeared to be unprofessional news and articles created by media with extensive use of sensational expressions, which might suffer from exaggeration and misrepresentation and thus distract public attention.

- The *researcher background* also negatively correlated with argumentative answers ( $OR = 0.89^*$  for evidence and  $OR = 0.88^*$  for reasoning). Note that this feature largely promoted public participation as revealed in Section 4.2.2. We observed that as many questions tended to create contrast between famous researchers (or institutes) and controversial results for public attention, some answers were distracted to criticize the researcher rather than focus on the research-sensemaking task itself.
- More *following* users of the asker positively predicted on-task ( $OR = 1.32^{***}$ ) and argumentative answers ( $OR = 1.13^*$  for evidence and  $OR = 1.26^{**}$  for reasoning), which was potentially due to the askers' behavior of inviting relevant and high-quality contributors. Questions asked for *discussing implication* had lower proportions of on-task answers ( $OR = 0.75^{***}$ ) compared to other question-asking goals such as *only sensemaking* (as the reference group) and *credibility assessment* ( $OR = 1.14^*$ ). Different *science branches of topics* significantly associated with the social dimension in answers, indicating the varied collaboration levels of knowledge co-construction in different-discipline research-sensemaking questions.
- Different from their positive correlations with user participation, both *question length* and *topic size* negatively correlated knowledge co-construction dimensions. In particular, longer questions received lower proportions of on-task discourse ( $OR = 0.83^*$ ). Also, even though adding more topics might attract broader users, *topic size* had significantly negative correlations with the percentage of evidence-supported ( $OR = 0.90^*$ ) or collaboratively-constructed ( $OR = 0.89^*$ ) answers, possibly due to the contributions of non-expert users outside of the subject-related community.

### 4.3 RQ3: Users' Collaborative Work in Co-constructing Research-sensemaking Questions

RQ1 and RQ2 revealed the strategies in proposing research-sensemaking questions and their correlations with user engagement and knowledge construction. In this section, we took a further step to unpack the *construction process* of research-sensemaking questions through collaborative editing. We illustrated the dynamics of users' collaborative work in co-constructing high-quality research-sensemaking questions, including *co-establishing the topic scope*, *scientific reframing and correction to enhance rigor*, *co-contributing reliable and understandable external references*, and *refinement for clear and engaging narratives*.

**4.3.1 Co-establishing the Topic Scope.** Adding or deleting topics, attached as tags of questions, has been identified as the most common co-editing behavior on the Zhihu platform [12]. Through the log analysis, we found that such topic co-editing was a crucial process for users to co-establish the topic scope for research-sensemaking questions. It was a common practice of topic co-editing to exclude less relevant or redundant topics to refine the topic scope. For instance, in the question titled "A survey shows that only 35% of Chinese people sleep for more than 8 hours a day. Experts recommend going to sleep between 10pm and 11pm. How is your sleep quality?", the original asker used tags of "health" and "medical treatment". In the co-editing process, one user removed "medical treatment", and another user supplemented "sleep cycle" to better describe the topic. More importantly, users managed to balance the scientific topic scope for these questions based on topic co-editing, keeping them from being over-general without a focus, or getting limited into over-narrow and professional spaces that could not engage the public. For example, in the question titled "Research has proved

Table 4. Beta regression models predicting knowledge co-construction dimensions in research-sensemaking questions, including (1) on-task discourse (Model 2, *epistemic dimension*); (2) evidence and reasoning (Model 3a and 3b, *argumentative dimension*); and (3) social (Model 4, *social mode dimension*). OR (Odds Ratio) indicates the odds change of the dependent variable when increasing an independent variable by one unit. We color-coded features having significantly positive correlations with at least two dependent variables in **green**, and those having significantly negative correlations with at least two dependent variables in **red**. \*\*\*p<0.001; \*\*p<0.01; \*p<0.05.

|   |                             | M2: On-task |           | M3a: Evidence |           | M3b: Reasoning |           | M4: Social |           |
|---|-----------------------------|-------------|-----------|---------------|-----------|----------------|-----------|------------|-----------|
|   |                             | OR          | Std. Err. | OR            | Std. Err. | OR             | Std. Err. | OR         | Std. Err. |
| Strategies in asking research-sensemaking question (question title)           |                             |             |           |               |           |                |           |            |           |
| Significance Signs  | impact indication           | 1.32***     | 0.086     | 1.02          | 0.052     | 1.02           | 0.065     | 1.03       | 0.076     |
|   | timeliness indication       | 0.85*       | 0.073     | 1.00          | 0.050     | 1.02           | 0.066     | 0.98       | 0.072     |
|   | publication venue           | 1.11        | 0.075     | 0.96          | 0.053     | 1.02           | 0.065     | 1.02       | 0.074     |
|   | researcher back-ground      | 0.96        | 0.075     | 0.89*         | 0.053     | 0.88*          | 0.064     | 1.04       | 0.070     |
| Rigorous Descriptions   | hedging                     | 1.13        | 0.075     | 1.05          | 0.049     | 1.17**         | 0.064     | 1.15*      | 0.071     |
|   | data use                    | 1.00        | 0.077     | 0.93          | 0.052     | 0.90           | 0.066     | 1.10       | 0.072     |
| Eyecatching Narratives  | quoting                     | 0.80**      | 0.077     | 1.07          | 0.051     | 1.01           | 0.065     | 1.09       | 0.075     |
|   | emotional arousal           | 0.96        | 0.082     | 0.91          | 0.055     | 0.97           | 0.070     | 1.01       | 0.081     |
|   | counter-intuitive           | 1.05        | 0.073     | 1.08          | 0.050     | 1.17**         | 0.063     | 1.00       | 0.075     |
| Strategies in describing research-sensemaking question (question description) |                             |             |           |               |           |                |           |            |           |
| Additional Resources  | visuals                     | 1.00        | 0.072     | 0.97          | 0.049     | 0.99           | 0.063     | 0.98       | 0.071     |
|   | paper sources               | 1.12        | 0.082     | 1.17**        | 0.055     | 1.23**         | 0.070     | 1.20*      | 0.077     |
|   | supplementary materials     | 0.89        | 0.079     | 0.91*         | 0.050     | 0.89*          | 0.065     | 0.93       | 0.072     |
| Comprehensive details   | detailed methods            | 0.92        | 0.078     | 1.13          | 0.054     | 0.94           | 0.069     | 1.06       | 0.077     |
|   | detailed results            | 0.92        | 0.073     | 1.01          | 0.058     | 0.99           | 0.075     | 1.02       | 0.085     |
|   | detailed questions          | 1.12        | 0.075     | 1.08          | 0.049     | 1.13*          | 0.064     | 1.07       | 0.068     |
| Clear presentations   | lede                        | 1.20*       | 0.078     | 1.03          | 0.051     | 1.18*          | 0.067     | 0.93       | 0.074     |
|   | structuring                 | 0.94        | 0.079     | 0.90*         | 0.052     | 0.89           | 0.069     | 0.97       | 0.079     |
| Control Variables   |                             |             |           |               |           |                |           |            |           |
| Topic - science branch (ref: life science)                                    | chemistry&material          | 0.92        | 0.066     | 1.12*         | 0.051     | 1.09           | 0.063     | 1.28***    | 0.054     |
|   | earth&space science         | 0.94        | 0.088     | 0.96          | 0.058     | 1.29***        | 0.078     | 1.20*      | 0.084     |
|   | social science              | 0.98        | 0.100     | 0.94          | 0.066     | 0.93           | 0.085     | 1.37**     | 0.100     |
|   | computers&technology        | 1.03        | 0.080     | 1.07          | 0.054     | 1.00           | 0.070     | 1.23***    | 0.058     |
|   | health                      | 1.03        | 0.123     | 1.02          | 0.080     | 1.14           | 0.102     | 1.45**     | 0.121     |
|   | math&physics                | 0.90        | 0.081     | 0.93          | 0.054     | 0.93           | 0.070     | 1.21*      | 0.074     |
| Topic - decision relevance  | decision-related            | 1.28**      | 0.087     | 1.09          | 0.059     | 1.06           | 0.076     | 0.89       | 0.086     |
| Goal (ref: only sensemaking)  | with assessing credibility  | 1.14*       | 0.076     | 1.00          | 0.051     | 1.04           | 0.067     | 1.03       | 0.070     |
|   | with discussing implication | 0.75***     | 0.087     | 0.99          | 0.057     | 0.91           | 0.075     | 0.98       | 0.084     |
|   | with reasoning              | 0.97        | 0.072     | 0.98          | 0.050     | 1.03           | 0.064     | 0.86*      | 0.075     |
| Asker Info  | non-anonymous               | 1.03        | 0.077     | 1.02          | 0.051     | 1.09           | 0.065     | 0.96       | 0.072     |
|   | follower                    | 0.96        | 0.069     | 1.00          | 0.051     | 1.00           | 0.063     | 1.08       | 0.071     |
|   | following                   | 1.32***     | 0.083     | 1.13*         | 0.053     | 1.26***        | 0.071     | 1.00       | 0.076     |
| Other Meta Info   | question length             | 0.83**      | 0.077     | 0.97          | 0.052     | 0.91           | 0.067     | 0.92       | 0.073     |
|   | topic size                  | 0.97        | 0.077     | 0.90*         | 0.050     | 0.99           | 0.065     | 0.89*      | 0.068     |

that a large number of white-tailed deer in the eastern United States have been infected with COVID-19. Will the COVID-19 infection of wild animals have an impact on the prevention and control of the

*pandemic?*”, the original asker attached topics of “research” and “health”. Another user removed them and added a topic of “SARS-CoV-2” to narrow down the topic to a more specific scope.

**4.3.2 Scientific Reframing and Correction to Enhance Rigor.** We observed that many original questions intentionally or unintentionally mistranslated, exaggerated, or misrepresented research findings such as omitting essential conditions. In this scenario, users’ collaborative editing afforded scientific framing and correction to enhance rigor and rectify misinformation for some questions, which helped prevent public sensemaking in a misguided direction. For instance, an asker posted an original research-sensemaking question titled “*The Lancet published the latest inactivated vaccine data in Chile: Inactivated booster injections are 20-30% less effective than mRNA vaccines. How to interpret it?*”. Another user supplemented an important condition of the research, “*with the basis of two doses of inactivated vaccine*”, to make the statement more accurate and rigorous. Such scientific reframing was not limited to the quoted research findings, but also applied to the meta information about the research. For example, when the original asker posted “*What do you think of Nature’s comment: In praise of replication studies and null results?*”, another user changed the “comment” to “editorial” to correctly describe the article type. In some other examples, scientific reframing was also reflected in ways such as adding the specific time the research was conducted and published for time-sensitive work, highlighting the studied populations, and changing approximated numbers to specific numbers.

**4.3.3 Co-contributing Reliable and Understandable External References.** The original paper and its supplementary interpretations intuitively played a significant role in research sensemaking, which was also validated in RQ1 and RQ2. The log analysis further revealed the prevalence of collaborative contribution on the external references for the question, including links to both the original paper and supplementary materials. Some research-sensemaking questions initially lacked links to the original paper, and other domain experts added them based on the research description. It was surprising to note that even when the original asker had provided external links to the original paper or relevant materials, other users might supplement references that were more understandable or reliable. For instance, in a question where the original asker attached the link to the paper in English, another user supplemented a Chinese article from domain experts translating and interpreting the paper in a more comprehensible way, aiming to engage a wider audience. In another example, the original question provided a post on social media Weibo discussing the research, and another user later modified it to a more authoritative Chinese article summarizing relevant research findings. These examples indicated users’ collaborative efforts to establish external references that could elucidate the research to a broad audience.

**4.3.4 Refinement for Clear and Engaging Narratives.** We found that some narrative strategies aiming for clear and engaging presentations appeared to be the results of collaborative refinement. For example, the strategy of *quoting key findings* in question titles was frequently implemented in collaborative editing by voluntary contributors, who added quotation marks for the key findings to highlight them. Some users also volunteered to add one or several opening sentences as the *lede* for the question descriptions that initially began with direct research explanations, some just simply as “*a recent study published in <publication venue> has attracted widespread attention. It is a collaborative research project conducted by <institutes>*”. These efforts of collaborative refinement manifested unwritten norms that were commonly accepted by users as “good practice” for asking research-sensemaking questions.

## 5 DISCUSSION

This work identifies strategies used in composing research-sensemaking questions that are associated with engaging with the audience and eliciting quality answers from the community. Generally, this work unveils a new pattern of participatory science communication with great potential to facilitate knowledge dissemination and sensemaking, as well as points out its existing challenges especially the gap between wide public participation and quality answer construction. This section situates the findings within the literature, reflects on the opportunities and challenges for CQA-based research-sensemaking, and proposes design implications to facilitate accurate, engaging, and effective science communication.

### 5.1 From Passive Receivers to Active Askers: Rethinking Users' Efforts in Initiating Research-sensemaking

The participatory knowledge exchange, with a focus on dialogue and public engagement, has characterized contemporary science communication on social media [90]. It goes beyond the traditional model of linear transmission from scientists to media and then public [6, 13], which necessitates rethinking the participatory practice of stakeholders. This study deepened the understanding of an emerging role of the public in science communication, i.e., the question-askers who initiated research-sensemaking on CQA platforms. In this section, we discussed how users were engaged in this process and how it afforded new opportunities for science communication.

Effectively engaging the general public with science-related information has been a challenging task in science communication even for mature science communicators like scientists [39]. This work uncovered a comprehensive taxonomy of users' strategies in proposing research-sensemaking questions in CQA platforms, shedding light on **public wisdom in initiating research-sensemaking** when their roles became askers. Most user-developed strategies naturally reflected practical principles established in prior work. For example, the SUCCESS framework proposed by Finkler and León emphasized the significance of being Simple, Unexpected, Concrete, Credible, Emotional, and Science Storytelling for science-related rhetoric [33]. Resonating with it, users crafted strategic research-sensemaking questions that noted researcher background and publication venue to establish *credibility*, used data narratives to enhance *concreteness*, and applied counter-intuitive and emotional statements for *unexpectedness* and *emotionalization*, some of them managing to attract millions of views. As shown in Section 4.3, some strategies even developed as unwritten norms collaboratively implemented by voluntary community members. These findings contributed new empirical evidence to strategic science communication in real-world settings [19, 33, 35]. More importantly, they added new nuances on how users naturally integrated the art of questioning and narratives for knowledge exchange in participatory science communication [62]. On this note, we suggest future work broadly investigate how to empower the general public as active initiators in knowledge exchange, and exploit the opportunities of "speaking the language of the general public" for science communication.

Nonetheless, limitations of users' spontaneous questioning for research sensemaking also emerged. Some more professional science communication strategies for rigor and accuracy, such as the use of visual representations to describe the research, were less observed in our dataset. Meanwhile, we noticed that part of user-developed strategies such as counter-intuitive and emotional expressions potentially introduced distortions, exaggerations or misrepresentations of the research outputs, leading to misguided research sensemaking. It could contribute to polarized discourse, a feature of "post-normal" science communication on digital media [6]. Such misinformation in research-sensemaking questions might also be unintentionally brought by mistranslation and missed essential conditions when non-experts lacked specific domain knowledge. Therefore,

we suggest the involvement of scientists and experts to assist with question crafting to improve its quality and upskill question askers, e.g., affording a mentorship program for scientific questioning [37]. The deployment of flagging mechanisms [59] for research-sensemaking questions, leveraging AI-supported or crowdsourced credibility checking to mark potentially unreliable or misguided questions, is also a promising direction.

## 5.2 Questioning Strategies Failed to Ensure Both: Understanding the Tension Between Wide Participation and Quality Knowledge Construction

The existing literature in science communication on social media has indicated the dilemma of engaging the general public and ensuring the rigor of knowledge exchange [4, 39, 90]. Epistemic science discussion intrinsically correlates with linguistic barriers (e.g., scientific terminology and hedge words) that gatekeep public participation [4], while public science discussion without the constraints of rigor and objectivity may exploit research outputs for conspiracy theories and extremist ideology [58, 93]. By analyzing different knowledge co-construction dimensions on research-sensemaking questions, this work unearthed more nuanced dynamics of *the tension between wide participation and quality knowledge construction* in science communication, and proposed design implications to cope.

**5.2.1 The Dilemma between Wide Participation and Quality Knowledge Construction.** The content analysis of the answers to research-sensemaking questions provided empirical evidence of the divergence between broad public participation and high-quality science discourse. As revealed in Section 4.2.1, all public participation indexes of questions (views, votes, followers and answers) had varied degrees of negative correlations with the proportions of high-quality knowledge construction dimensions (on-task discourse, argumentative claims with evidence and reasoning, and socialized knowledge discussion). This finding validated the difficulty of balancing the quantity of public participation and quality of knowledge construction [4, 94]. In fact, such tension reflects a potential challenge under the blurring boundaries of science and journalism in the era of social media [6]. Communicating science to the public no longer necessarily relies on professional intermediaries as gatekeepers; instead, the participation of “scientist citizens” and contextual interpretation of science characterize such post-normal science communication [78]. On this note, being “newsworthy” may become a more salient feature for public research sensemaking rather than being “scientific”, along with which misguidance and misinterpretation may prevail [58, 93]. Therefore, more investigations are warranted to afford more specialized guidance interfaces for the public to contribute high-quality knowledge, and develop more thorough moderation mechanisms to mitigate off-topic, polarized, and misleading discourse in addition to affording accessibility [10].

This work further demonstrated how question-asking strategies that attracted more public participation may not align with strategies facilitating epistemic and argumentative knowledge establishment. For example, Section 4.2 revealed that the strategy of highlighting *researcher background* correlated with more views and answers, but diverted the focus of some answers away from sensemaking to commenting institutes or researchers. In contrast, *detailed questions*, typically calling for more specific and professional answers, correlated with high-quality knowledge construction but not more answers. These findings emphasized the significance of curating research-sensemaking questions that considered its influence on both engaging a wider audience and stimulating responses with higher quality. How questions were proposed largely influenced the public engagement and sensemaking directions in CQA platforms, which was rather different than tweeterials that disseminated knowledge themselves [39] or more targeted questions with specific receivers in a community [42]. Also, the discrepancy between wide participation and rigorous sensemaking necessitates nuanced investigations focusing on fine-grained dimensions measuring



public engagement on science-related topics rather than only numerical indexes, which echoes prior work [90, 91].

**5.2.2 Design Implications.** Notably, we highlight the crucial role of the sociotechnical context in shaping the trade-off. A widely applicable and long-lasting value in knowledge co-creation, rather than one-off and personal questions, has gradually become the focus of CQA platforms [2, 63, 95]. Based on it, the platform design such as voting mechanisms and recommendation algorithms may encourage widely applicable questions and answers acknowledged by more users rather than very specific ones [2, 3] (also consider the removal of question details in Quora [79]). This incentive also intersects with the monetization model of CQA platforms [31], which usually motivates askers/answerers to prioritize wide attention for potential profits. However, it may inadvertently conflict with certain research-sensemaking questions that require specificity to achieve rigor. For example, askers may omit specific descriptions or conditions that are actually crucial for particular research to attract wider attention. Moreover, with the aim of gaining wide engagement, users may intentionally pose misguided research-sensemaking questions, such as misinterpreting research with counter-intuitive conclusions, just setting a target for public emotional venting; or selectively reporting controversial findings only to initiate off-topic debates. Therefore, we suggest a knowledge-related feedback mechanism, in addition to upvotes, as an alternative feedback and evaluation system for research-sensemaking Q&A. For example, verified domain experts or established users on related topics may endorse questions or answers with knowledge-related tags (e.g., “experimental evidence” and “reasoning with data”), and other users could show their approval of them. The reputation and recommendation systems may take these dimensions into consideration to spur quality knowledge construction.

Besides, when broad participation on science topics inevitably introduces off-task discourse, we suggest CQA platforms afford options to ease the seeking for quality knowledge-related answers, so as to facilitate those aiming to figure out the research (in contrast to users with an entertaining or conversational goal). The aforementioned knowledge-related tags provide one potential way to assist with quality answer filtering. Another possible approach is a specialized AI-supported sorting and filtering interface, through which users could sort answers by AI-evaluated relevance, argumentative claims, or socialized discussions. The good performance of automatic classifications on these knowledge construction dimensions in this work suggests the feasibility of this interface. Nonetheless, though the AI-supported sorting and filtering interface might ease research sensemaking and spur quality answers, further investigations are warranted to understand potential limitations such as algorithmic biases and possible risks of crowd gaming algorithm [29].

Finally, it is a promising direction to utilize the power of social connections for constructive research sensemaking, such as enhancing the reward mechanism for high-quality answer inviting. Our findings also noted the influence of social networks on answer quality, when questions proposed by askers following more users (and thus potentially inviting more relevant and quality contributors) had more epistemic and argumentative discussions.

### 5.3 Proposing Research-Sensemaking as a Collaborative Effort: Unpacking Opportunities and Challenges of Co-editing Research-Sensemaking Questions

The affordance of collaborative question editing, similar to Quora [67] and Stack Overflow [60], reflected the attempts of leveraging collective wisdom to construct high-quality questions on Zhihu [12]. This section discusses how collaborative editing brings opportunities and challenges to the specific setting of research-sensemaking.

When presenting accurate and engaging science-related information naturally puts a high demand on questioning techniques, we note that collaborative editing plays an irreplaceable role in

constructing and refining research-sensemaking questions. Section 4.3 demonstrates the value of co-editing ranging from setting up a suitable topic scope, establishing reliable and comprehensible external information, to enhancing the rigor and clarity of presentations. These findings enrich the understanding of the potential of collaborative editing in scientific questioning [60, 67]. When constructing high-quality research-sensemaking might be challenging for non-experts, the involvement of collective wisdom may not only be beneficial to generate more engaging and rigorous research-sensemaking questions, but also work as a descriptive influence of norms [4, 14] that guide newcomers to ask in an acceptable way. It could also alleviate unique challenges of science communication, such as moderating misuse and polarization [58, 93] and establishing credibility [8] with collaborative work, especially when only “credible” users were qualified to co-edit questions on Zhihu [12]. To this end, it is warranted to investigate how the power of collaborative editing could be adapted to other community-based or social network-based science communication settings.

On the other hand, co-editing research-sensemaking questions inherit and develop some issues of collaborative editing. Though the relatively high barrier of scientific language made “editing war” [56] less frequently observed in our sample, it also complicated the justification of editing [12]. Many users used the default reasons such as “punctuation and formatting error” that did not point out the essential rigor-related problems and might not convince the original askers. On this note, we suggest some lightweight design improvements such as adopting specialized editing terminologies for science-related topics, or automatically examining edits to provide suggestions on justification reasons. Besides, some reframing on subjective narratives (e.g., detailed personal questions in the description) faces the tension between the subjective nature of question-asking and the objective nature of science communication. Though one reason for introducing collaborative editing on Zhihu is to generalize some questions [96], it may harm specific users’ interest in research-sensemaking and limit more narrow directions of knowledge exchange. Therefore, we call for future investigations on question redirection and branching systems to make science communication more accessible to both general and personal research-sensemaking questions.

## 5.4 Limitations

This work has the following limitations: (1) We used a keyword-based data collection approach, which could promote the purity and representativeness of the dataset but inevitably limited the dataset size and diversity (e.g., questions implicitly proposing research-sensemaking were largely excluded in our dataset); (2) We adopted regression analysis to capture how question-asking strategies were correlated with user-participation or knowledge construction dimensions. Even though the results were valuable in helping understand the potential influence of question-asking strategies, the correlations are not sufficient to claim causal relationships as a general limitation of regression analysis; (3) We measure the potential effects of questioning strategies only from a quantitative perspective (i.e., how they correlated with the *observed* user participation and knowledge construction). A qualitative view focusing on how users perceived and responded to these questioning strategies (e.g., through interviews) was also significant to understanding this science communication setting; (4) Due to the scope of this study, we did not pay close attention to some more nuanced research questions that emerged from the analysis, such as a comprehensive understanding of how the tension between wide participation and quality knowledge construction appeared, how askers perceived and balanced this trade-off, and how misrepresented or exaggerated questions misguided the knowledge construction in answers. We suggest future work to systematically investigate these specific challenges of CQA-based science communication to facilitate accurate and effective knowledge construction and dissemination online. Besides, though this work contributes to the HCI and CSCW literature by expanding the understanding of science communication in a non-western context, a cross-platform or cross-cultural comparative study would be beneficial to unpack the

sociocultural factors in science communication and provide valuable insights into CQA-based research sensemaking.

## 6 CONCLUSION

Users are increasingly gathering in research-sensemaking questions to discuss science-related topics in CQA platforms, and good question-asking strategies are crucial to attract public participation and facilitate knowledge construction. This work makes the first attempt to investigate strategies in crafting research-sensemaking questions, their correlations with user engagement and knowledge construction, and co-editing efforts from the community to implement them. To achieve this, we collected 837 science-related questions with 157,684 answers and conducted a mixed-methods study on the Zhihu platform. Through an open coding approach, we captured a comprehensive taxonomy of user-developed strategies in question titles and descriptions to enhance rigor (e.g., hedging and data use) and engage the audience (e.g., emotional and counter-intuitive statements). Through regression analysis, we identified these strategies' correlations with public participation and knowledge construction, such as the increased views and answers in emotional questions. It helped to unpack the divergence between wide participation and argumentative knowledge construction when few strategies could promote both. Finally, the inductive log analysis suggested the unique values of collaborative editing in promoting question quality, such as collaborative reframing and correction to enhance rigor. We discuss design implications for accessible, accurate, and effective research-sensemaking in CQA platforms.

## ACKNOWLEDGMENTS

The research was supported in part by a RGC RIF grant under the contract R6021-20, RGC CRF grants under the contracts C7004-22G and C1029-22G, and RGC GRF grants under the contracts 16209120, 16200221 and 16207922.

## REFERENCES

- [1] Michael Olusegun Akinwande, Hussaini Garba Dikko, Agboola Samson, et al. 2015. Variance inflation factor: as a condition for the inclusion of suppressor variable (s) in regression analysis. *Open journal of statistics* 5, 07 (2015), 754.
- [2] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 850–858.
- [3] Muhammad Asaduzzaman, Ahmed Shah Mashiyat, Chanchal K Roy, and Kevin A Schneider. 2013. Answering questions about unanswered questions of stack overflow. In *2013 10th Working Conference on Mining Software Repositories (MSR)*. IEEE, 97–100.
- [4] Tal August, Dallas Card, Gary Hsieh, Noah A Smith, and Katharina Reinecke. 2020. Explain like I am a Scientist: The Linguistic Barriers of Entry to r/science. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [5] Deborah Blum, Mary Knudson, and Robin Marantz Henig. 2005. *A field guide for science writers: The official guide of the National Association of Science Writers*. Oxford University Press.
- [6] Michael Brüggemann, Ines Lörcher, and Stefanie Walter. 2020. Post-normal science communication: Exploring the blurring boundaries of science and journalism. *Journal of Science Communication* 19, 3 (2020), A02.
- [7] Wändi Bruine de Bruin and Ann Bostrom. 2013. Assessing what to address in science communication. *Proceedings of the National Academy of Sciences* 110, supplement\_3 (2013), 14062–14068.
- [8] Massimiano Bucchi. 2017. Credibility, expertise and the challenges of science communication 2.0. , 890–893 pages.
- [9] Katie Burke. 2015. 12 Tips for Scientists Writing for the General Public. <https://www.americanscientist.org/blog/from-the-staff/12-tips-for-scientists-writing-for-the-general-public>. Accessed: 2022-12-27.
- [10] Terry W Burns, D John O'Connor, and Susan M Stocklmayer. 2003. Science communication: a contemporary definition. *Public understanding of science* 12, 2 (2003), 183–202.
- [11] China Internet Information Center. 2022. Zhihu's "Ask for Answers from Science" event ends with professional discussions reaching 5.1 billion views (in Chinese). <http://tech.china.com.cn/roll/20220715/389258.shtml>. Accessed: 2023-01-13.

- [12] Si Chen, Haocong Cheng, and Yun Huang. 2020. Who Is Changing Your Question on a Social Q&A Website?. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [13] Jason Chilvers. 2010. Sustainable participation? Mapping out and reflecting on the field of public dialogue on science and technology. (2010).
- [14] Robert B Cialdini, Raymond R Reno, and Carl A Kallgren. 1990. A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology* 58, 6 (1990), 1015.
- [15] Chad E Cook, Neil E O'connell, Toby Hall, Steven Z George, Gwendolen Jull, Alexis A Wright, Enrique Lluch Gírbés, Jeremy Lewis, and Mark Hancock. 2018. Benefits and threats to using social media for presenting and implementing evidence. *journal of orthopaedic & sports physical therapy* 48, 1 (2018), 3–7.
- [16] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- [17] Stefany Coxe, Stephen G West, and Leona S Aiken. 2009. The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of personality assessment* 91, 2 (2009), 121–136.
- [18] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 3504–3514.
- [19] Michael F Dahlstrom. 2014. Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the national academy of sciences* 111, supplement\_4 (2014), 13614–13620.
- [20] Emily S Darling, David Shiffman, Isabelle M Côté, and Joshua A Drew. 2013. The role of Twitter in the life cycle of a scientific publication. *arXiv preprint arXiv:1305.0435* (2013).
- [21] Dipto Das, Carsten Østerlund, and Bryan Semaan. 2021. "Jol" or "Pani"? How Does Governance Shape a Platform's Identity? *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
- [22] Sarah R Davies and Maja Horst. 2016. *Science communication: Culture, identity and citizenship*. Springer.
- [23] Lucas Machado de Oliveira and Olga Goussevskaia. 2020. Sponsored content and user engagement dynamics on Instagram. In *Proceedings of the 35th annual ACM symposium on applied computing*. 1835–1842.
- [24] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [25] Patrick Marcel Joseph Dubois, Mahya Maftouni, and Andrea Bunt. 2022. Towards More Gender-Inclusive Q&As: Investigating Perceptions of Additional Community Presence Information. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–23.
- [26] Patrick Marcel Joseph Dubois, Mahya Maftouni, Parmit K Chilana, Joanna McGrenere, and Andrea Bunt. 2020. Gender differences in graphic design q&as: How community and site characteristics contribute to gender gaps in answering questions. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–26.
- [27] Ilana Dubovi and Iris Tabak. 2021. Interactions between emotional and cognitive engagement with science on YouTube. *Public Understanding of Science* 30, 6 (2021), 759–776.
- [28] Elaine Howard Ecklund, Sarah A James, and Anne E Lincoln. 2012. How academic biologists and physicists view science outreach. *PloS one* 7, 5 (2012), e36240.
- [29] Ziv Epstein, Gordon Pennycook, and David Rand. 2020. Will the crowd game the algorithm? Using layperson judgments to combat misinformation on social media by downranking distrusted sources. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–11.
- [30] Elisabetta Falchetti, Silvia Caravita, and Alessandra Sperduti. 2007. What do laypersons want to know from scientists? An analysis of a dialogue between scientists and laypersons on the web site Scienzaonline. *Public Understanding of Science* 16, 4 (2007), 489–506.
- [31] Bin Fang, Xin Fu, Shaoxia Liu, and Shun Cai. 2021. Post-purchase warranty and knowledge monetization: Evidence from a paid-knowledge platform. *Information & Management* 58, 3 (2021), 103446.
- [32] Silvia Ferrari and Francisco Cribari-Neto. 2004. Beta regression for modelling rates and proportions. *Journal of applied statistics* 31, 7 (2004), 799–815.
- [33] Wiebke Finkler and Bienvenido León-Anguiano. 2019. The power of storytelling and video: a visual rhetoric for science communication. (2019).
- [34] Frank Fischer, Johannes Bruhn, Cornelia Gräsel, and Heinz Mandl. 2002. Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction* 12, 2 (2002), 213–232.
- [35] Danny Flemming, Ulrike Cress, Sophia Kimmig, Miriam Brandt, and Joachim Kimmerle. 2018. Emotionalization in science communication: The impact of narratives and visual representations on knowledge gain and risk perception. *Frontiers in Communication* 3 (2018), 3.
- [36] Uwe Flick. 2022. *An introduction to qualitative research*. sage.
- [37] Denae Ford, Kristina Lustig, Jeremy Banks, and Chris Parnin. 2018. "We Don't Do That Here" How Collaborative Editing with Mentors Improves Engagement in Social Q&A Communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

- [38] Anna Gardiner, Miriam Sullivan, and Ann Grand. 2018. Who are you writing for? Differences in response to blog design between scientists and nonscientists. *Science Communication* 40, 1 (2018), 109–123.
- [39] Katy Ilonka Gero, Vivian Liu, Sarah Huang, Jennifer Lee, and Lydia B Chilton. 2021. What Makes Tweetorials Tick: How Experts Communicate Complex Topics on Twitter. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–26.
- [40] Sarah A Gilbert. 2020. "I run the world's largest historical outreach project and it's on a cesspool of a website." Moderating a Public Scholarship Site on Reddit: A Case Study of r/AskHistorians. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.
- [41] Dimitris C Gkikas, Katerina Tzafilkou, Prokopis K Theodoridis, Aristogiannis Garmpis, and Marios C Gkikas. 2022. How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook. *International Journal of Information Management Data Insights* 2, 1 (2022), 100067.
- [42] Anatoliy Gruz, Priya Kumar, Deena Abul-Fottouh, and Caroline Haythornthwaite. 2020. Coding and classifying knowledge exchange on social media: A comparative analysis of the# Twitterstorians and AskHistorians communities. *Computer Supported Cooperative Work (CSCW)* 29, 6 (2020), 629–656.
- [43] Cheng Guo and Kelly Caine. 2021. Anonymity, user engagement, quality, and trolling on Q&A sites. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–27.
- [44] Reza Hadi Mogavi, Yuanhao Zhang, Ehsan-Ul Haq, Yongjin Wu, Pan Hui, and Xiaojuan Ma. 2022. What Do Users Think of Promotional Gamification Schemes? A Qualitative Case Study in a Question Answering Website. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–34.
- [45] F Maxwell Harper, Daphne Raban, Sheizaf Rafaeli, and Joseph A Konstan. 2008. Predictors of answer quality in online Q&A sites. In *Proceedings of the SIGCHI Conference on human factors in computing systems*. 865–874.
- [46] Changyang He, Lu He, Tun Lu, and Bo Li. 2021. Beyond Entertainment: Unpacking Danmaku and Comments' Role of Information Sharing and Sentiment Expression in Online Crisis Videos. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–27.
- [47] Lu He and Changyang He. 2022. Help Me# DebunkThis: Unpacking Individual and Community's Collaborative Work in Information Credibility Assessment. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–31.
- [48] Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on YouTube. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–27.
- [49] Zhihu Inc. 2022. Zhihu. <https://www.zhihu.com/>. Accessed: 2022-10-20.
- [50] Zhihu Inc. 2022. Zhihu Inc. First Quarter 2023 Financial Results. <https://ir.zhihu.com/2022-05-25-Zhihu-Inc-Reports-First-Quarter-2022-Unaudited-Financial-Results>. Accessed: 2022-10-20.
- [51] Xiao-Ling Jin, Zhongyun Zhou, Matthew KO Lee, and Christy MK Cheung. 2013. Why users keep answering questions in online question answering communities: A theoretical and empirical investigation. *International Journal of Information Management* 33, 1 (2013), 93–104.
- [52] Ridley Jones, Lucas Colusso, Katharina Reinecke, and Gary Hsieh. 2019. r/science: Challenges and opportunities in online science communication. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.
- [53] Kathleen Kendall-Tackett. 2007. *How to write for a general audience: A guide for academics who want to share their knowledge with the world and have fun doing it*. American Psychological Association.
- [54] Teck-Ee Keng and Ming-Yu Cheng. 2023. How do Researchers Use Social Media for Science Communication? *Bulletin of Science, Technology & Society* (2023), 02704676231165654.
- [55] Lara Khansa, Xiao Ma, Divakaran Liginlal, and Sung S Kim. 2015. Understanding members' active participation in online question-and-answer communities: A theory and empirical analysis. *Journal of Management Information Systems* 32, 2 (2015), 162–203.
- [56] Aniket Kittur, Bongwon Suh, Bryan A Pendleton, and Ed H Chi. 2007. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 453–462.
- [57] Kathrin Kopke, Jeffrey Black, and Amy Dozier. 2019. Stepping out of the ivory tower for ocean literacy. *Frontiers in Marine Science* (2019), 60.
- [58] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M Jones, and Arvind Satyanarayan. 2021. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–18.
- [59] Chen Li. 2022. Flagging Controversies: The effect of flagging mechanisms on Zhihu platform. In *European Conference on Social Media*, Vol. 9. 237–246.
- [60] Guo Li, Haiyi Zhu, Tun Lu, Xianghua Ding, and Ning Gu. 2015. Is it good to be like Wikipedia? Exploring the trade-offs of introducing collaborative editing model to Q&A sites. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1080–1091.

- [61] Lei Li, Daqing He, Wei Jeng, Spencer Goodwin, and Chengzhi Zhang. 2015. Answer quality characteristics and prediction on an academic Q&A Site: A case study on ResearchGate. In *Proceedings of the 24th international conference on world wide web*. 1453–1458.
- [62] Jingjing Liang, Ximing Liu, and Weiyu Zhang. 2019. Scientists vs laypeople: How genetically modified food is discussed on a Chinese Q&A website. *Public Understanding of Science* 28, 8 (2019), 991–1004.
- [63] Chengzhong Liu, Shixu Zhou, Dingdong Liu, Junze Li, Zeyu Huang, and Xiaojuan Ma. 2023. CoArgue: Fostering Lurkers' Contribution to Collective Arguments in Community-based QA Platforms. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [64] P Liu and RP Lin. 2015. Research on knowledge sharing and dissemination behavior of Online Q & A Services: Taking Zhihu as an example. *Documentation, Information & Knowledge* 6 (2015), 109–119.
- [65] Ignacio López-Goñi and Manuel Sánchez-Angulo. 2018. Social networks as a tool for science communication and public engagement: focus on Twitter. *FEMS Microbiology letters* 365, 2 (2018), fnx246.
- [66] María José Luzón. 2013. Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience. *Written Communication* 30, 4 (2013), 428–457.
- [67] Suman Kalyan Maity, Aman Kharb, and Animesh Mukherjee. 2018. Analyzing the linguistic structure of question texts to characterize answerability in quora. *IEEE Transactions on Computational Social Systems* 5, 3 (2018), 816–828.
- [68] SODA Global Marketing. 2022. Brand marketing on Zhihu - largest Chinese Q&A platform 2021: Reach your educated professional users. <https://www.soda-global.com/post/zhihu-brand-marketing-guide-2021>. Accessed: 2022-10-20.
- [69] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282.
- [70] Katherine L McNeill and Dean M Martin. 2011. Claims, evidence, and reasoning. *Science and Children* 48, 8 (2011), 52.
- [71] Steve Miller. 2001. Public understanding of science at the crossroads. *Public understanding of science* 10, 1 (2001), 115–120.
- [72] Kevin Kyung Nam, Mark S Ackerman, and Lada A Adamic. 2009. Questions in, knowledge in? A study of Naver's question answering community. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 779–788.
- [73] Malte Ostendorff, Terry Ruas, Moritz Schubotz, Georg Rehm, and Bela Gipp. 2020. Pairwise multi-class document classification for semantic relations between wikipedia articles. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. 127–136.
- [74] Aditya Pal, Shuo Chang, and Joseph Konstan. 2012. Evolution of experts in question answering communities. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 6. 274–281.
- [75] Vineet Pandey, Amnon Amir, Justine Debilius, Embriette R Hyde, Tomasz Kosciolk, Rob Knight, and Scott Klemmer. 2017. Gut instinct: Creating scientific theories with online learners. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6825–6836.
- [76] Sumanth Patil and Kyumin Lee. 2016. Detecting experts on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors. *Social network analysis and mining* 6 (2016), 1–11.
- [77] Hans Peter Peters. 2013. Gap between science and media revisited: Scientists as public communicators. *Proceedings of the National Academy of Sciences* 110, supplement\_3 (2013), 14102–14109.
- [78] Pamela Pietrucci and Leah Ceccarelli. 2019. Scientist citizens: Rhetoric and responsibility in L'Aquila. *Rhetoric and Public Affairs* 22, 1 (2019), 95–128.
- [79] Quora. 2017. Why were question details removed from Quora on August 3, 2017? <https://www.quora.com/Why-were-question-details-removed-from-Quora-on-August-3-2017/answer/Quora>. Accessed: 2022-10-20.
- [80] Sujith Ravi, Bo Pang, Vibhor Rastogi, and Ravi Kumar. 2014. Great question! question quality in community q&a. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- [81] Ronald E Rice and Howard Giles. 2017. The contexts and dynamics of science communication and language. *Journal of Language and Social Psychology* 36, 1 (2017), 127–139.
- [82] Ivan Srba and Maria Bielikova. 2016. A comprehensive survey and classification of approaches for community question answering. *ACM Transactions on the Web (TWEB)* 10, 3 (2016), 1–63.
- [83] Bonnie Stewart. 2016. Collapsed publics: Orality, literacy, and vulnerability in academic Twitter. *Journal of Applied Social Theory* 1, 1 (2016), 61–86.
- [84] Ramine Tinati, Max Van Kleek, Elena Simperl, Markus Luczak-Rösch, Robert Simpson, and Nigel Shadbolt. 2015. Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 4069–4078.
- [85] Debbie Treise and Michael F Weigold. 2002. Advancing science communication: A survey of science communicators. *Science Communication* 23, 3 (2002), 310–322.
- [86] Andrew W Vargo and Shigeo Matsubara. 2018. Identity and performance in technical Q&A. *Behaviour & Information Technology* 37, 7 (2018), 658–674.

- [87] Gang Wang, Konark Gill, Manish Mohanlal, Haitao Zheng, and Ben Y Zhao. 2013. Wisdom in the social crowd: an analysis of quora. In *Proceedings of the 22nd international conference on World Wide Web*. 1341–1352.
- [88] Armin Weinberger and Frank Fischer. 2006. A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & education* 46, 1 (2006), 71–95.
- [89] Spencer Williams and Gary Hsieh. 2021. The Effects of User Comments on Science News Engagement. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.
- [90] Spencer Williams, Ridley Jones, Katharina Reinecke, and Gary Hsieh. 2022. An HCI Research Agenda for Online Science Communication. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–22.
- [91] Haijun Xia, Hui Xin Ng, Zhutian Chen, and James Hollan. 2022. Millions and Billions of Views: Understanding Popular Science and Knowledge Communication on Video-Sharing Platforms. In *Proceedings of the Ninth ACM Conference on Learning@ Scale*. 163–174.
- [92] Jie Yang, Ke Tao, Alessandro Bozzon, and Geert-Jan Houben. 2014. Sparrows and owls: Characterisation of expert behaviour in stackoverflow. In *International conference on user modeling, adaptation, and personalization*. Springer, 266–277.
- [93] Satrio Baskoro Yudhoatmojo, Emiliano De Cristofaro, and Jeremy Blackburn. 2021. "We won't even challenge their lefty academic definition of racist:" Understanding the Use of e-Prints on Reddit and 4chan. *arXiv preprint arXiv:2111.02455* (2021).
- [94] Yu Zhang, Changyang He, Huanchen Wang, and Zhicong Lu. 2023. Understanding Communication Strategies and Viewer Engagement with Science Knowledge Videos on Bilibili. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [95] Yan Zhang, Mingli Zhang, Nuan Luo, Yu Wang, and Tao Niu. 2019. Understanding the formation mechanism of high-quality knowledge in social question and answer communities: A knowledge co-creation perspective. *International Journal of Information Management* 48 (2019), 72–84.
- [96] Zhihu. 2020. How to participate in the public editing of Zhihu questions? (in Chinese). <https://www.zhihu.com/question/24501563>. Accessed: 2022-10-20.
- [97] Theodore E Zorn, Juliet Roper, C Kay Weaver, and Colleen Rigby. 2012. Influence in science dialogue: Individual attitude changes as a result of dialogue between laypersons and scientists. *Public Understanding of Science* 21, 7 (2012), 848–864.

Received January 2023; revised July 2023; accepted November 2023